VII Workshop de Visão Computacional

Anais do Evento

www.wvc2011.ufpr.br

22 a 25 de maio de 2011 - UFPR - Curitiba

ISSN - 2175-6120

WVC 2011

Anais do VII Workshop de Visão Computacional

Universidade Federal do Paraná

ISSN 2175-6120

Editores:
Luiz Antônio Pereira Neves
Lucas Ferrari de Oliveira
Hugo Vieira Neto

Curitiba, 22 a 25 de Maio de 2011

Bem-vindos ao WVC 2011

A Universidade Federal do Paraná, representada pelo seu Setor de Educação Profissional e Tecnológica (SEPT) no campus Jardim Botânico, sente-se honrada e orgulhosa em sediar a sétima edição do Workshop de Visão Computacional. Este evento tem atraído trabalhos de pesquisadores de diferentes partes do país e, desse modo, está se consolidando como um evento de referência em âmbito nacional, sendo o único a tratar especificamente de pesquisas envolvendo todos os aspectos de estudos em Visão Computacional e Processamento de Imagens. Vale ressaltar que é a primeira vez que o WVC está sendo organizado fora do Estado de São Paulo, representado pelos pesquisadores da área de Visão Computacional de três instituições paranaenses (UFPR, UTFPR e PUCPR), que possuem grupos de pesquisas consolidados, integrando-os nesta edição.

Quando fizemos a candidatura da nossa Universidade para sediar o WVC, em 2010 em Presidente Prudente, sabíamos das dificuldades e responsabilidades para realizar este evento. Por outro lado, o apoio da UFPR e de colegas do Comitê Organizador foi significativo para esta organização. Esperamos que nosso trabalho possa atender às expectativas daqueles que nos confiaram esta responsabilidade.

Gostaria muito de agradecer à CAPES, ao CNPq, à OPTO e ao LACTEC pelo suporte financeiro que nos foi concedido. Sem eles seria muito difícil realizar um evento desta natureza. Além disso, parabenizo todos os autores e co-autores, de diversas áreas, que submeteram seus trabalhos científicos ao WVC 2011 com alta qualidade técnica e metodológica, enobrecendo a imagem do evento na difusão da pesquisa. Quero registrar os meus sinceros agradecimentos aos colegas que compõem o Comitê Organizador, que representam as três Instituições de Pesquisa do Paraná (UFPR, UTFPR e PUCPR). Por fim, aos alunos do Setor de Educação Profissional e Tecnológica, dos cursos superiores de Tecnologia em Análise e Desenvolvimento de Sistemas, em Secretariado Executivo e em Construção de Instrumentos Musicais (Luteria), expresso minha gratidão por terem operacionalizado este evento. Sem a participação de cada um de vocês, seria absolutamente impossível realizar o WVC 2011.

Esperamos que o VII Workshop de Visão Computacional seja um ambiente propício de compartilhamento de ideias, de divulgação de pesquisas e de novas parcerias acadêmicas. Aproveitem ao máximo este evento!

Sinceras saudações Universitárias,

Luiz Antônio Pereira Neves – Presidente do Comitê Organizador do WVC 2011

Comitê Permanente

Adilson Gonzaga, Universidade de São Paulo
Evandro Luis Linhari Rodrigues, Universidade de São Paulo
Maurílio Boaventura, Universidade Estadual Paulista
Inês Boaventura, Universidade Estadual Paulista
Aparecido Nilceu Marana, Universidade Estadual Paulista
Maurício Marengoni, Universidade Presbiteriana Mackenzie
Marco Antonio Piteri, Universidade Estadual Paulista

Comitê Organizador

Luiz Antônio Pereira Neves, Universidade Federal do Paraná
Lucas Ferrari de Oliveira, Universidade Federal do Paraná
Dieval Guizelini, Universidade Federal do Paraná
Hugo Vieira Neto, Universidade Tecnológica Federal do Paraná
Humberto Remígio Gamba, Universidade Tecnológica Federal do Paraná
Edson José Rodrigues Justino, Pontifícia Universidade Católica do Paraná
Chidambaram Chidambaram, Universidade do Estado de Santa Catarina

Comitê Científico

Adilson Gonzaga, Universidade de São Paulo
Alejandro César Frery Orgambide, Universidade Federal de Alagoas
Antonio Maria Garcia Tomaseli, Universidade Estadual Paulista
Aparecido Nilceu Marana, Universidade Estadual Paulista
Carlos Eduardo Thomaz, Centro Universitario da FEI
Chidambaram Chidambaram, Universidade do Estado de Santa Catarina
Diego Gonçalves, Universidade Federal do Paraná
Dieval Guizelini, Universidade Federal do Paraná
Edson José Rodrigues Justino, Pontifícia Universidade Católica do Paraná
Evandro Luis Linhari Rodrigues, Universidade de São Paulo
Fabiana Rodrigues Leta, Universidade Federal Fluminense
Fátima de Lourdes dos Santos Nunes, Universidade de São Paulo
Flávio Bortolozzi, Centro Universitário De Maringá
Gilson Antonio Giraldi, Laboratório Nacional de Computação Científica
Gustavo Benvenutti Borba, Universidade Tecnológica Federal do Paraná
Hélio Pedrini, Universidade Estadual de Campinas
Hemerson Pistori, Universidade Católica Dom Bosco
Homero Schiabel, Universidade de São Paulo
Hugo Vieira Neto, Universidade Tecnológica Federal do Paraná
Humberto Remígio Gamba, Universidade Tecnológica Federal do Paraná
Inês Boaventura, Universidade Estadual Paulista
Ivan Nunes da Silva, Universidade de São Paulo
Jacques Facon, Pontifícia Universidade Católica do Parana
Jander Moreira, Universidade Federal de São Carlos
Jefferson Martins, Universidade Tecnológica Federal do Paraná
João Paulo Papa, Universidade Estadual de Campinas
João do Espírito Santo Batista Neto, Universidade de São Paulo
João Manuel Ribeiro da Silva Tavares, Universidade do Porto
Joaquim de Mira Júnior, Universidade Tecnológica Federal do Paraná
José Hiroki Saito, Universidade Federal de São Carlos
José Roberto Nogueira, Universidade Estadual Paulista
Júlio César Nievola, Pontifícia Universidade Católica do Parana
Karl Apaza Agüero, Universidade Federal do Paraná
Leonardo Vidal Batista, Universidade Federal da Paraiba
Leonardo Nogueira Matos, Universidade Federal de Sergipe
Lucas Ferrari de Oliveira, Universidade Federal do Paraná
Luciano Volcan Agostini, Universidade Federal de Pelotas
Luiz Antônio Pereira Neves, Universidade Federal do Paraná
Luiz Eduardo Soares de Oliveira, Universidade Federal do Paraná
Marcelo Barreiro, Instituto Federal do Triângulo Mineiro
Marcelo Nepomoceno Kapp, Universidade Federal da Integração Latino-Americana
Marcelo Mansano, Universidade Federal do Paraná
Marcelo Ponciano da Silva, Universidade de São Paulo
Marcelo Vieira Contin, Universidade de São Paulo
Marco Antônio Piteri, Universidade Estadual Paulista
Maria Paiva, Universidade de São Paulo
Marilton Sanchotene de Aguiar, Universidade Federal de Pelotas
Mário Fernando Montenegro de Campos, Universidade Federal de Minas Gerais
Maurício Segundo, Universidade Federal do Paraná
Maurício Galo, Universidade Estadual Paulista
Maurício Marengoni, Universidade Presbiteriana Mackenzie
Maurilio Boaventura, Universidade Estadual Paulista
Paula Nakamoto, Instituto Federal do Triângulo Mineiro
Paulo de Azevedo Marques, Universidade de São Paulo
Paulo Eduardo Ambrósio, Universidade Estadual de Santa Cruz
Paulo Vinicius Wolski Radtke, Universidade Federal do Paraná
Rafael Santos, Instituto Nacional de Pesquisas Espaciais
Ricardo Torres, Universidade Estadual de Campinas
Roberto Tadeu Raittz, Universidade Federal do Paraná
Ronaldo Martins da Costa, Universidade Federal de Goiás
Valentin Obac Roda, Universidade Federal do Rio Grande do Norte
William Schwartz, Universidade Estadual de Campinas
Yandre Costa, Universidade Estadual de Maringá

Conteúdo

Human Iris Characterization Through Optimum-Path Forest and Dynamic Features 1
Ronaldo Costa, Adilson Gonzaga, João Paulo Papa, Aparecido Marana

Desenvolvimento de Técnicas Computacionais para Auxiliar Testes de Qualidade em Imagens Mamográficas que Utilizam o *Phantom* CDMAM 3.4 5
Maria Sousa, Bruno Barufaldi, Homero Schiabel, Leonardo Batista

Uso de Morfologia Matemática para Remoção de Ruídos da Subtração de Fundo 10
Wagner Amaral, Clésio Luis Tozzi

Sistema para Controle de Qualidade de Mamógrafos Através da Análise Computacional de Imagens de *Phantom* Utilizando a Ferramenta WEKA 16
Bruno Barufaldi, Eduardo Santana, Leonardo Batista, Homero Schiabel, José Carvalho

Mapeamento Estatístico Paramétrico das Alterações Estruturais Cerebrais Devido à Doença de Alzheimer e ao Transtorno Cognitivo Leve 22
Michel Fernandes, João Sato, Geraldo Busatto, Carlos Thomaz

Classificação Automática de Grupos de Espécies Florestais 28
Jefferson Martins, Luiz Oliveira, Pedro Luiz Paula Filho, Silvana Nisgoski

Segmentation Techniques Based on Background Subtraction and Supervised Learning: A Comparative Study for Images of Mice and Human Skins 34
Bruno Brandoli Machado, Wesley Nunes Gonçalves, Jonathan Andrade Silva, Vinícius Saueia, Kleber Padovani de Souza, Bruno Toledo, Hemerson Pistori

Computer Vision Applied to Dengue's Larvae Death Rate Calculation: Preliminary Results 40
Kleber Padovani de Souza, Jaime Queiroz, Hemerson Pistori

Extração de Características de Imagens Mamográficas Utilizando Redes Complexas para Detecção de Câncer de Mama 45
Jonathan Ramos, Carolina Watanabe

Segmentação e Restauração Digital para Eliminação de Artefatos em Imagens Frontais de Face 51
André Sobiecki, Luiz Neves, Gilson Giraldi, Gilka Gattás, Carlos Thomaz

Hand Edge - Detecção da Borda da Mão pela Diferença de Luminosidade 57
André Luis Borsato, Lucas Godoi, Luis Degrandi, Osvaldo Severino Junior

EVEREVIS: Sistema de Detecção e Indexação Automática de Objetos em Vídeos Digitais 62
Bruno do Nascimento Teixeira, Júlia Engrácia, Fillipe Souza, Tiago Oliveira Cunha, Lucas Gonçalves, Christiane Okamoto, Igor Oliveira, Arnaldo Araújo

Análise Discriminante e Classificação de Imagens 2D de Ultrassonografia Mamária 67
Albert Xavier, João Sato, Gilson Giraldi, Carlos Thomaz

Segmentação de Texturas Dinâmicas: Um Novo Método Baseado em Caminhadas Determinísticas 73
Wesley Nunes Gonçalves, Bruno Brandoli Machado, Odemir Bruno

Uma Abordagem para a Localização e o Reconhecimento de Placas de Licenciamento Veicular por meio de Operadores Morfológicos e Busca por *Template* 79
Danilo Sousa Abreu, Wonder Alexandre Luz Alves, Sidnei Alves de Araújo, André Librantz

Uma Ferramenta Visual para Analisar as Imagens de Microscopia de Força Atômica AFM 85
Carolina Guillén, Marcelo da Silva, Roberto Faria, Odemir Bruno

Efficient Retina-like Resampling from Cartesian Images 91
Hugo Vieira Neto, Diogo Kuiaski, Gustavo Borba

Reconhecimento Multibiométrico de Pessoas em Imagens de Vídeo para Monitoramento Visual de Ambientes 97
Gustavo Botelho de Souza, Ícaro Ribeiro, Aparecido Marana

Cálculo de Histograma em FPGA para Processamento de Imagens em Tempo Real 103
Jovander Silva Freitas, Maximiliam Luppe

Localização e Reconhecimento de Caracteres em Placas de Automóveis 107
Leonardo de Oliveira, Adilson Gonzaga

Detecção de Tomadas em Vídeos de Futebol Utilizando Teoria da Informação Não-Extensiva 113
Guilherme Lopes, Werner Fukuma, Paulo Rodrigues

Influência do Uso de Sensores Coloridos em Retinógrafos Digitais 118
Flávio Vieira, Evandro Rodrigues

Detecção de Expressões Faciais: Uma Abordagem Baseada em Análise do Fluxo Óptico 124
Leonardo Panta Leão, Leonardo Matos, Maria Augusta Nunes

Detecção Automática da Área em Imagens de Impressão Digital 131
Marcos William Oliveira, Maurílio Boaventura, Inês Boaventura

Fractal Image Encoding Using a Constant Size Domain Pool 137
Vitor de Lima, William Schwartz, Hélio Pedrini

Feature Space Visualization to Image Analysis: Preliminary Results 143
Bruno Brandoli Machado, Danilo Eler, João do Espírito Santo Batista Neto

Um Algoritmo Rápido para o Reconhecimento de Caracteres 149
Francisco Silva, Almir Artero, Maria Stela de Paiva, Ricardo Luís Barbosa

Image Decomposition with Anisotropic Diffusion Applied to Leaf-texture Analysis 155
Bruno Brandoli Machado, Wesley Nunes Gonçalves, Odemir Bruno

Segmentação de Íris Utilizando *Bag of Keypoints* 161
Fabio Brandão, Maurício Marengoni

Influência da Padronização do Sistema de Cor RGB nos Métodos de Visão Computacional 167
Dalcimar Casanova, Maurício Falvo, Odemir Bruno

Uma Metodologia Híbrida para Segmentação de Lesões de Pele 173
Alex Araujo, João Manuel Tavares, Roberta Oliveira, Norian Marranghello, Aledir Silveira Pereira, Ricardo Rossetti

Geração de um Classificador para Detecção do HLB em Citros Utilizando Técnica *Box-Counting Multifractal* 179
Diego Silva, Lúcio Jorge, Adolfo Posadas, Maria Stela de Paiva

Spatiotemporal Gabor Filters: A New Method for Dynamic Texture Recognition 184
Wesley Nunes Gonçalves, Bruno Brandoli Machado, Odemir Bruno

Segmentação do Músculo Peitoral em Imagens Mamográficas Através do Recorte Automático de ROI e da Binarização de Otsu 190
José Rogério Filho, Leonardo Batista, Homero Schiabel, Luciana Verçosa

Segmentação Não-Supervisionada de Imagens em Duas Regiões Baseada no Método Competição entre Regiões *Fuzzy* Utilizando Distribuições de Probabilidade 196
Vinicius Pereira Borges, Celia Barcelos, Marcos Batista, Karla Freitas

Sistema de Visão Computacional para Classificação de Grãos de Café por Cor e Forma 202
Pedro Ivo Oyama, Lúcio Jorge, Carlos Cesar Gomes

Análise Óptica de Sistemas Eletro-ópticos por Meio do Cálculo da MTF 208
Elisa Barbarini, Evandro Rodrigues

Comparação de Métodos de Classificação da Cobertura do Solo Urbano Utilizando Imagens IKONOS II com e sem o Auxílio de Dados LiDAR 213
Leonardo Rodrigues de Deus, Rafael Santos, Leila Fonseca

Algoritmo Baseado no Operador de Energia de Teager (TEO) para Segmentação de Linhas e Palavras em Textos Manuscritos e Impressos 219
Josimeire do Amaral Tavares, Igor Peretta, Gerson Lima, Keiji Yamanaka, Mônica Pais

Diferenciação da *Greening* do Citros de Outras Doenças Foliares a Partir de Descritores de Cor e Forma 225
Patricia Estevam Ribeiro, Lúcio Jorge, Maria Stela de Paiva

Uma Abordagem para Verificação Facial em Dispositivos Móveis 231
Tiago Pereira, Marcus Angeloni

Autenticação Automática de Assinaturas *Online*: Investigando Alternativas de Representação 237
Cassia Silva, Guilherme Mota, Aruquia Peixoto, Alexandre Sztajnberg

Reconstrução Digital de Documentos Mutilados Usando Programação Dinâmica e Algoritmo de Prim Modificado 243
Andre Miranda Pimenta, Edson Justino, Luiz Oliveira

Abordagem Morfológica de Segmentação do Melanoma 249
Arlete Teresinha Beuren, Rodrigo Gomes Pinheiro, Jacques Facon

Mosaico Automático de Imagens Agrícolas Aéreas 255
André Tarallo, Maria Stela de Paiva, Lúcio Jorge

Segmentação de Imagens de Escoamentos Bifásicos Gás-líquido Baseada em Métodos *Level Set* e Morfologia Matemática 260
Leyza Dorini, Mauren Coelho de Andrade, Lúcia Valéria Arruda

Software Para Auxílio no Diagnóstico do Glaucoma 266
Virginia Andersson, Lucas de Oliveira

Focus of Expansion Estimation for Motion Segmentation from a Single Camera 272
José Kuiaski, André Lazzaretti, Hugo Vieira Neto

Sistema Automático de Diagnóstico de Parásitos Helmintos 278
Gabriel Humpire, Juan Chuctaya, Juan Carlos Gutiérrez Cáceres, Odemir Bruno, Cesar Beltrán

Seleção Automática de Parâmetros de Detectores de Bordas de Imagens Digitais Baseado em Algoritmos Genéticos 284
Cristiane da Costa, Inês Boaventura

Detecção Automática de Manchas Solares Utilizando Técnicas de Visão Computacional 290
Wander Costa, Israel dos Santos, Eduardo Spagiari, Adriana Valio, Maurício Marengoni

The Power of Kulback-Leibler Divergence Extendable for Tsallis Entropy in Comparing Cone-Beam and Multi-Slice Tomography Images 296
André Sobiecki, C. Gallão, D. Cosme, Paulo Rodrigues

Estudo Comparativo da Utilização de Filtros de Suavização para Reconstrução 3D de Superfícies 302
Sérgio Fujii, Lauro Wichert-Ana, Paulo de Azevedo Marques, Lucas de Oliveira

Algoritmo Genético Aplicado ao Estudo de Imagens de Gel de Eletroforese Bidimensional 308
Juliana Tibães, Vanely de Souza, Aline Schultz, Angélica Juliani, Letícia de Jesus, Michelly Gehlen, Roberto Tadeu Raittz

Impact of Classifier Prototyping on Classification Systems Optimization 312
Paulo Vinicius Wolski Radtke, Robert Sabourin, Tony Wong

Modelagem Contextual de Padrões Aplicada ao Monitoramento de Eventos em Vídeo 318
Fabio Borges dos Santos, Edson Justino, Alceu Britto Junior

Recuperação de Imagens por Conteúdo Utilizando Lógica *Fuzzy* - Um Estudo de Caso Sobre Imagens Faciais 324
Leila Bergamasco, Fátima Nunes

Traçado e Análise Cefalométrica: Uma Solução Computacional 330
Lucas Lima Batista, Roberto Silva Nunes Lago, Claudio Goes, Mauricio Cunha Escarpinati, Michele Angelo

Retrieval of Labeled Shape Classes in Binary Images Using Standardized Granulometric Intervals 335
Alexandre Harano, Ronaldo Hashimoto

NURBS Interpolation with FPGA Acceleration 341
Luiz Marcelo Silva, Maria Stela Paiva

HUMAN IRIS CHARACTERIZATION THROUGH OPTIMUM-PATH FOREST AND DYNAMIC FEATURES

R. M. da Costa *A. Gonzaga* *J. P. Papa, A. N. Marana*

Institute of Informatics
Federal University of Goiás
ronaldocosta@inf.ufg.br

School of Engineering at São Carlos
University of São Paulo
adilson@sel.eesc.usp.br

Department of Computing
Univ Estadual Paulista (UNESP)
{papa,nilceu}@fc.unesp.br

ABSTRACT

Liveness detection methods have been extensively addressed to avoid frauds in biometric systems. In this paper, we introduce a recently method for the characterization of iris dynamics in the context of image classification, by assessing its performance with a novel pattern recognition technique called Optimum-Path Forest, which demonstrated to be similar to Self Organizing Maps, but much faster, and superior to Support Vector Machines and Artificial Neural Networks using Multilayer Perceptrons.

Index Terms— Iris recognition, Biometrics, Optimum-Path Forest, Pattern Recognition

1. INTRODUCTION

The seminal work about human iris was initially proposed by Frank Burch in 1936. However, it was only with John Daugman that computerized applications were developed based on human identification assessing their iris properties [1]. Although several methods have been developed in the last decade for recognition based on iris features, most of the them are based on the algorithm created by Daugman [1]. The identification techniques are extremely precise, completing the process in a split second. However, all literature-known methods are applied to static images. In such a way, by observing the current methods, it is verified that they do not guarantee that the individual being evaluated is actually present or if the captured image is a photograph, prosthesis or some type of digital image, which may serve as entry data to the recognition system for frauds.

The art of distinguishing real people from photos or non real subjects receives the name of liveness detection, which has been extensively used to avoid frauds in biometric systems. Jee et al. [2], for instance, applied a methodology to detect real faces from non real ones. Kanematsu et al. [3] proposed a liveness detection method to detect fake iris from real ones using a variation in the brightness of an iris pattern induced by a pupillary reflex. Recently, Costa and Gonzaga [4] proposed an innovating method composed by an electronic device and a software to capture the dynamic properties of the iris, such as pupil contraction and dilation time, which is very robust to frauds regarding fake and computerized created iris. However, this work did not evaluate the dynamic features of the iris in the context of classification, only for precision and recall.

Support Vector Machines (SVM) and Artificial Neural Networks (ANN) have been also used for iris recognition. The former was applied by Roy and Bhattacharya [5], and the latter was addressed by [6, 7], among others. Despite the use of these artificial intelligence techniques have been increasing, some flaws need to be revisited. An ANN with multi-layer perceptrons (ANN-MLP), for example, can address linearly and non-linearly separable problems, but not non-separable situations with maximum effectiveness [8]. As an unstable classifier, collections of ANN-MLP can improve its performance up to some unknown limit of classifiers. Although Self Organizing Maps (SOM) [8] have been also extensively used in several applications, the selection of its parameters, i.e., the number of neurons in the squared lattice and the number of iterations of the algorithm, is a hard task and extremely dependent of the application.

Support Vector Machines (SVM) have been proposed to overcome the problem, by assuming linearly separable classes in a higher-dimensional feature space [9]. Its computational cost rapidly increases with the training set size and the number of support vectors. As a binary classifier, multiple SVM are required to solve a multi-class problem. Another important question is that the assumption of separability may also not be valid in any space of finite dimension.

Recently, a novel framework for graph-based classifiers that reduce the pattern recognition problem as an optimum path forest computation (OPF) in the feature space induced by a graph were presented [10]. These kind of classifiers do not interpret the classification task as a hyperplanes optimization problem, but as a combinatorial optimum-path computation from some key samples (prototypes) to the remaining nodes. Each prototype becomes a root from its optimum-path tree and each node is classified according to its strongly connected prototype, that defines a discrete optimal partition (influence region) of the feature space. The OPF-based classifiers have some advantages with respect to the aforementioned classifiers: (i) one of them is free of parameters, (ii) they do not assume any shape/separability of the feature space and (iii) run training phase faster, which allows the development of real time applications for fraud detection in electricity systems.

This work has as the main goal to characterize human iris by means of an OPF classifier, and has two main contributions: (i) to be the first into applying the OPF classifier for iris recognition and to be pioneer into using the dynamic iris' method proposed by Costa and Gonzaga [4] in the context of supervised classification. The remainder of this paper is organized as follows. Sections 2 and 3 present the OPF and the dynamic features of the iris method, respectively. Section 4 shows the experimental results, in which we compared OPF among two kind of Support Vector Machines, ANN-MLP and SOM networks. Section 5 states conclusions.

2. OPTIMUM-PATH FOREST

Let Z_1 and Z_2 be the training and test sets with $|Z_1|$ and $|Z_2|$ samples such as points or image elements. Let $\lambda(s)$ be the function that assigns the correct label i, $i = 1, 2, \ldots, c$, from class i to any sample

$s \in Z_1 \cup Z_2$. Z_1 is a labeled set used to the design of the classifier and Z_2 is used to assess the performance of classifier and it is kept unseen during the project.

Let $S \subset Z_1$ be a set of prototypes of all classes (i.e., key samples that best represent the classes). Let v be an algorithm which extracts n attributes (color, shape or texture properties) from any sample $s \in Z_1 \cup Z_2$ and returns a vector $\vec{v}(s) \in \Re^n$. The distance $d(s,t)$ between two samples, s and t, is the one between their feature vectors $\vec{v}(s)$ and $\vec{v}(t)$. One can use any valid metric (e.g., Euclidean) or a more elaborated distance algorithm. Our problem consists of using S, (v,d) and Z_1 to project an optimal classifier which can predict the correct label $\lambda(s)$ of any sample $s \in Z_2$. The OPF classifier creates a discrete optimal partition of the feature space such that any sample $s \in Z_2$ can be classified according to this partition [10]. This partition is an optimum path forest (OPF) computed in $\Re^n$ by the image foresting transform (IFT) algorithm [11].

Let (Z_1, A) be a complete graph whose the nodes are the samples in Z_1 and any pair of samples defines an arc in $A = Z_1 \times Z_1$. The arcs do not need to be stored and so the graph does not need to be explicitly represented. A path is a sequence of distinct samples $\pi = \langle s_1, s_2, \ldots, s_k \rangle$, where $(s_i, s_{i+1}) \in A$ for $1 \leq i \leq k-1$. A path is said *trivial* if $\pi = \langle s_1 \rangle$. We assign to each path π a cost $f(\pi)$ given by a path-cost function f. A path π is said optimum if $f(\pi) \leq f(\pi')$ for any other path π', where π and π' end at a same sample s_k. We also denote by $\pi \cdot \langle s,t \rangle$ the concatenation of a path π with terminus at s and an arc (s,t).

The OPF algorithm may be used with any *smooth* path-cost function which can group samples with similar properties [11]. We are interested in prototypes that fall in the region between classes, which are generally overlapped regions. So, we will address the path-cost function f_{max}, because of its theoretical properties for estimating prototypes that have this behavior (Section 2.1 gives the details about this procedure):

$$\begin{aligned} f_{max}(\langle s \rangle) &= \begin{cases} 0 & \text{if } s \in S, \\ +\infty & \text{otherwise} \end{cases} \\ f_{max}(\pi \cdot \langle s,t \rangle) &= \max\{f_{max}(\pi), d(s,t)\}, \end{aligned} \quad (1)$$

such that $f_{max}(\pi)$ computes the maximum distance between adjacent samples in π, when π is not a trivial path.

The OPF algorithm assigns one optimum path $P^*(s)$ from S to every sample $s \in Z_1$, forming an optimum path forest P (a function with no cycles which assigns to each $s \in Z_1 \backslash S$ its predecessor $P(s)$ in $P^*(s)$ or a marker nil when $s \in S$. Let $R(s) \in S$ be the root of $P^*(s)$ which can be reached from $P(s)$. The OPF algorithm computes for each $s \in Z_1$, the cost $C(s)$ of $P^*(s)$, the label $L(s) = \lambda(R(s))$, and the predecessor $P(s)$, as follows.

Algorithm 1 – OPF ALGORITHM

INPUT:	A λ-labeled training set Z_1, prototypes $S \subset Z_1$ and the pair (v,d) for feature vector and distance computations.
OUTPUT:	Optimum-path forest P, cost map C and label map L.
AUXILIARY:	Priority queue Q and cost variable cst.

1. *For each* $s \in Z_1 \backslash S$, *set* $C(s) \leftarrow +\infty$.
2. *For each* $s \in S$, *do*
3. $\quad$ $C(s) \leftarrow 0$, $P(s) \leftarrow nil$, $L(s) \leftarrow \lambda(s)$, *and insert* s *in* Q.
4. *While* Q *is not empty, do*
5. $\quad$ *Remove from* Q *a sample* s *such that* $C(s)$ *is minimum.*
6. $\quad$ *For each* $t \in Z_1$ *such that* $t \neq s$ *and* $C(t) > C(s)$, *do*
7. $\quad\quad$ *Compute* $cst \leftarrow \max\{C(s), d(s,t)\}$.
8. $\quad\quad$ *If* $cst < C(t)$, *then*
9. $\quad\quad\quad$ *If* $C(t) \neq +\infty$, *then remove* t *from* Q.
10. $\quad\quad\quad$ $P(t) \leftarrow s$, $L(t) \leftarrow L(s)$, $C(t) \leftarrow cst$
11. $\quad\quad\quad$ *Insert* t *in* Q.

Lines $1-3$ initialize maps and insert prototypes in Q. The main loop computes an optimum path from S to every sample s in a non-decreasing order of cost (Lines $4-10$). At each iteration, a path of minimum cost $C(s)$ is obtained in P when we remove its last node s from Q (Line 5). Ties are broken in Q using first-in-first-out policy. That is, when two optimum paths reach an ambiguous sample s with the same minimum cost, s is assigned to the first path that reached it. Note that $C(t) > C(s)$ in Line 6 is false when t has been removed from Q and, therefore, $C(t) \neq +\infty$ in Line 9 is true only when $t \in Q$. Lines $8-11$ evaluate if the path that reaches an adjacent node t through s is cheaper than the current path with terminus t and update the position of t in Q, $C(t)$, $L(t)$ and $P(t)$ accordingly.

2.1. Training

We say that S^* is an optimum set of prototypes when Algorithm 1 minimizes the classification errors for every $s \in Z_1$. S^* can be found by exploiting the theoretical relation between minimum-spanning tree (MST) and optimum-path tree for f_{max}. The training essentially consists of finding S^* and an OPF classifier rooted at S^*.

By computing an MST in the complete graph (Z_1, A), we obtain a connected acyclic graph whose nodes are all samples of Z_1 and the arcs are undirected and weighted by the distances d between adjacent samples. The spanning tree is optimum in the sense that the sum of its arc weights is minimum as compared to any other spanning tree in the complete graph. In the MST, every pair of samples is connected by a single path which is optimum according to f_{max}. That is, the minimum-spanning tree contains one optimum-path tree for any selected root node.

The optimum prototypes are the closest elements of the MST with different labels in Z_1 (i.e., elements that fall in the frontier of the classes). By removing the arcs between different classes, their adjacent samples become prototypes in S^* and Algorithm 1 can compute an optimum-path forest with minimum classification errors in Z_1. Note that, a given class may be represented by multiple prototypes (i.e., optimum-path trees) and there must exist at least one prototype per class.

2.2. Classification

For any sample $t \in Z_2$, we consider all arcs connecting t with samples $s \in Z_1$, as though t were part of the training graph. Considering all possible paths from S^* to t, we find the optimum path $P^*(t)$ from S^* and label t with the class $\lambda(R(t))$ of its most strongly connected prototype $R(t) \in S^*$. This path can be identified incrementally, by evaluating the optimum cost $C(t)$ as

$$C(t) = \min\{\max\{C(s), d(s,t)\}\}, \ \forall s \in Z_1. \quad (2)$$

Let the node $s^* \in Z_1$ be the one that satisfies Equation 2 (i.e., the predecessor $P(t)$ in the optimum path $P^*(t)$). Given that $L(s^*) = \lambda(R(t))$, the classification simply assigns $L(s^*)$ as the class of t. An error occurs when $L(s^*) \neq \lambda(t)$.

3. DYNAMIC FEATURES FOR IRIS CHARACTERIZATION

The human eye is very sensitive to visible light. That is, the light within the violet-red range causes some type of reaction to the eye,

from the cones and rods up to the sclera (the white outer part). For example: the pupil contracts and dilates under the effect of the visible light, and the iris and the sclera exceptionally reflect within this range. In order to capture an image of the human iris by using visible light, there is a problem: how to keep the natural reflexes on the globe of the eye, iris and sclera surfaces from affecting the quality of the digitalized image? Several techniques are employed by professional photographers in order to deviate the light beam, by appropriately positioning the camera. Still, to acquire iris images at a good resolution thus allowing for the extraction of features aiming at the biometric recognition the photograph techniques cannot be utilized because, in general, the camera must be placed frontally to the iris and at a short distance. These images do not provide enough quality for a dependable biometric recognition.

The near infrared (NIR) illumination generates good resolution and definition images. However, due to the fact that they are not "visible" to the human eye, they do not allow for the necessary stimuli so that the pupil can perform the contraction and the dilation movements. The visible light offers the necessary stimulus. Nevertheless, the image quality is compromised, thus making the extraction of features difficult. In short: how to capture images with NIR illumination using visible light to contract and dilate the pupil, without causing reflexes on the iris, thus extracting the dynamic features?

Costa and Gonzaga [4] addressed this problem by taking into account the human optic system anatomy [1]. The eye captures through the cones and rods the light stimuli taken to the brain by the optic nerve, so that vision can be processed. In the transmission of the stimuli by the optic nerve, they pass through an area named optic chiasm. In this area, the mixture of the medium fibers of the optic nerve takes place. Fibers of the right optic nerve mix with the left ones and vice-versa. This causes the eyes to be connected, that is, the reflexes to stimuli applied to one of the eyes are presented in both. Such physiological function is denominated Consensual Reflex. This reflex is responsible for the synchronism of the movements for both eyes.

In such a way, Costa and Gonzaga [4] developed a device which performs different and independent tasks in each one of the eyes. The left eye receives visible light stimuli (white) in specific periods of time, controlled by the software developed, whereas the right eye image is digitalized in a video sequence, under NIR illumination. The acquisition device, connected to a computer, allows for the software to control the illumination on an eye, applied to in specific periods of time, whereas the image of the other eye is digitalized, thus forming a video sequence. The next section describes the method of Costa and Gonzaga [4] to extract and to select the dynamic features.

3.1. Feature extraction

In order to enable the evaluation of the behavior of each one of the features due to the illumination alteration conditions, five analysis periods are established. The first one is composed of the entire video (1,000 frames). The other periods are defined to record the precise period of illumination transition. This way, the periods comprise the following frames: (i) 1st period average of the 1,000 frames, (ii) 2nd period average between the frames 210 and 220, (iii) 3rd period average between the frames 420 and 430, (iv) 4th period average between the frames 630 and 640 and (v) 5th period average between the frames 840 and 850.

Based on that, twelve different dynamic features are extracted: 1) Pupil circularity; 2) Pupil diameter; 3) Pupil contraction/dilation time; 4) Pupil contraction/dilation rate; 5) Average of the gray levels of the segmented iris; 6) Standard deviation of the gray levels of the segmented iris; 7) Variation coefficient of the gray levels of the segmented iris; 8) Correlation; 9) Angular Second Moment (ASM); 10) Entropy; 11) Contrast; 12) Inverse Difference Moment (IDM);

These twelve features above form the base to generate a feature vector, and for each one of them the average is calculated within each period. Costa and Gonzaga [4] argued that some performed tests with the features confirmed the feasibility of the dynamic features. However, a selection of the most discriminative and dynamic features was performed through the Weka (Waikato Environment for Knowledge Analysis) software [12], on order to reduce the dimensionality of the feature vector, which was originally composed by 248 features.

4. EXPERIMENTAL RESULTS

We performed two series of experiments: in the former (Section 4.1) we used 50% of the whole dataset for training and the remaining 50% for testing classifiers, and in the last one (Section 4.2) we executed the experiments with different training and test set size percentages to allow a comparison about the robustness of the classifiers with respect to variations on the sets size. For both experiments, we executed OPF, SVM-RBF (SVM with RBF as kernel function), SVM-LINEAR (SVM without kernel mapping), ANN-MLP (ANN-MLP trained by backpropagation algorithm) and Kohonen Self Organizing Maps (SOM) 10 times with randomly generated training and test sets, to compute the mean accuracy and its standard deviation, and the mean training and test execution times (in seconds).

For SVM-RBF, we used the latest version of the LibSVM package [13] with Radial Basis Function (RBF) kernel, parameter optimization and the one-versus-one strategy for the multi-class problem. With respect to SVM-LINEAR, we used the LibLINEAR package [14] with C parameter optimized by cross-validation. For OPF we used the LibOPF [15], which is a library for the design of optimum-path forest-based classifiers, and for ANN-MLP we used the Fast Artificial Neural Network Library (FANN) [16]. The network configuration is i:h_1:h_2:o, where $i = 22$ (number of features), $h_1 = h_2 = 8$ and $o = 2$ (number of classes) are the number of neurons in the input, hidden and output layers, respectively. The ANN-MLP was trained with a backpropagation algorithm, and its architecture was empirically chosen. For SOM network we used a lattice with 100×100 neurons with 10 iterations for algorithm convergence.

4.1. Classifiers evaluation

We evaluate here the OPF, SVM-RBF, SVM-LINEAR, ANN-MLP and SOM for dynamic iris recognition using 50% for training and 50% for testing. Table 1 shows the mean accuracies and mean training and classification times (in seconds) after 10 runnings with randomly generated training and test sets.

Table 1. Mean accuracy and mean training and classification times OPF, SVM-RBF, SVM-LINEAR, ANN-MLP and SOM.

Classifier	Accuracy	Training time	Classif. Time
OPF	**83.68±1.13**	**0.01**	**0.017**
SVM-RBF	76.95±0.81	259.59	0.295
SVM-LINEAR	79.42±0.75	12.24	0.017
ANN-MLP	51.86±0.81	435.49	0.002
SOM	82.86±1.38	103.75	0.253

One can see that OPF and SOM obtained similar results if we take into account the standard deviation, and outperformed all remaining classifiers. However, the OPF classifier was much faster than SOM: 6900 times faster for training and 14.88 times faster for classification. This results confirmed that OPF is similar to the state of the art classification methodologies, but much faster [10].

4.2. Classifiers robustness

We evaluate the robustness of the classifiers with respect to variations on the training set size. We repeated the experiments shown in previous section with different training and test set sizes for OPF, SVM-RBF, SVM-LINEAR, ANN-MLP and SOM. Figure 1 displays the mean accuracies over the test set after 10 rounds of experiments for each training and test set size percentages configuration.

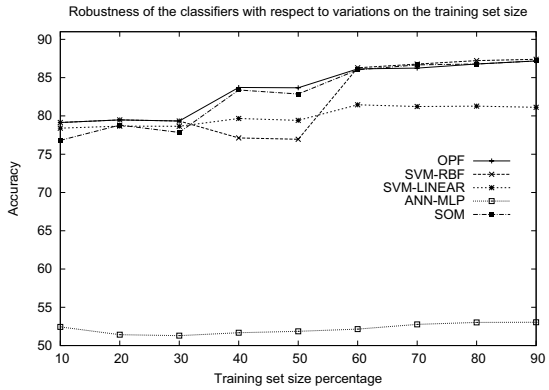


Fig. 1. Robustness of the classifiers with respect to different training and test size percentages.

One can see that OPF, SVM-RBF and SOM performed similar curves, except for SVM-RBF with 40%-50% of the training set size. We can see that ANN-MLP presented the worst results. One probably solution for that should be to use a collection of neural networks, although this is not guaranteed to solve the problem as a whole.

5. CONCLUSIONS

In this paper we introduced a recently developed dynamic iris characterization in the context of iris classification. Another main contribution concerns with the OPF first usage for automatic iris recognition. Experimental results comparing OPF among SVM-RBF, SVM-LINEAR, ANN-MLP and SOM have demonstrated the similarity between OPF and SOM in terms of effectiveness for this purpose, but the former was much faster both for training and classification phases.

In order to assess the robustness of the classifiers regarding variations on the training set size, we also performed an extra experiment with different percentages of the training set size. In this case, the OPF, SVM-RBF and SOM demonstrated to be similar, except for SVM-RBF with 40%-50% of training set size. One can see from this experiment that all techniques increased their effectiveness over the test set when we increase the training set size, as expected. However, a loss of SVM-LINEAR effectiveness with 30%-50% of training set size was observed.

Acknowledgments

The third author are grateful to FAPESP grant #2009/16206-1.

6. REFERENCES

[1] J. Daugman, "How iris recognition works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, 2004.

[2] H.-K. Jee, S.-U. Jung, and J.-H. Yoo, "Liveness detection for embedded face recognition system," *World Academy of Science, Engineering and Technology*, vol. 16, pp. 29–32, 2006.

[3] M. Kanematsu, H. Takano, and K. Nakamura, "Highly reliable liveness detection method for iris recognition," in *Proceedings of the Annual Conference SICE*, 2007, pp. 361–364.

[4] R. M. da Costa and A. Gonzaga, "Extraction and selection of dynamic features of the human iris," in *Proceedings of the XXII Brazilian Symposium on Computer Graphics and Image Processing*. 2009, pp. 1–7, IEEE Press.

[5] K. Roy and P. Bhattacharya, "Iris recognition with support vector machines," in *Advances in Biometrics*. 2005, vol. 3832, Lecture Notes in Computer Science, Springer Berlin / Heidelberg.

[6] A. M. Sarhan, "Iris recognition using discrete cosine transform and artificial neural networks," *Journal of Computer Science*, vol. 5, no. 5, pp. 369–373, 2009.

[7] M. Sarfraz, M. Deriche, M. Moinuddin, and S. S. A. Ali, "Intelligent iris recognition using neural networks," *Computer-Aided Intelligent Recognition Techniques and Applications*, pp. 145–167, 2005.

[8] S. Haykin, *Neural networks: a comprehensive foundation*, Prentice Hall, 1994.

[9] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[10] J. P. Papa, A. X. Falcão, and C. T. N. Suzuki, "Supervised pattern classification based on optimum-path forest," *International Journal of Imaging Systems and Technology*, vol. 19, no. 2, pp. 120–131, 2009.

[11] A. X. Falcão, J. Stolfi, and R. A. Lotufo, "The image foresting transform theory, algorithms, and applications," *IEEE Transactions on PAMI*, vol. 26, no. 1, pp. 19–29, 2004.

[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[13] C. C. Chang and C. J. Lin, *LIBSVM: A Library for Support Vector Machines*, 2001, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[14] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *JMLR*, vol. 9, pp. 1871–1874, 2008.

[15] J. P. Papa, C. T. N. Suzuki, and A. X. Falcão, *LibOPF: A library for the design of optimum-path forest classifiers*, 2009, Software version 1.0 available at `http://www.ic.unicamp.br/~afalcao/LibOPF`.

[16] S. Nissen, *Implementation of a Fast Artificial Neural Network Library (FANN)*, 2003, Department of Computer Science University of Copenhagen (DIKU). Software available at *http://leenissen.dk/fann/*.

Desenvolvimento de técnicas computacionais para auxiliar testes de qualidade em imagens mamográficas que utilizam o *phantom* CDMAM 3.4

Maria Angélica Zucareli Sousa[1], Bruno Barufaldi[2], Homero Schiabel[1], Leonardo V. Batista[2]
[1]*Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas – Departamento de Engenharia Elétrica - Universidade de São Paulo - São Carlos, SP, Brasil*
[2]*Departamento de Informática e Programa de Pós-graduação em Informática – Universidade Federal da Paraíba – João Pessoa, PB, Brasil*
angelica.zucareli@gmail.com

Abstract

The image quality in mammography can be achieved by conducting periodic tests. It is recommended that some quality parameters are measured with phantom images. Whereas this task is hard-working and time consuming, this study developed and compared two computational methods in order to assist the professional in performing the tests, reducing the subjectivity due to the observers. We used 27 phantom images obtained from six different digital mammographic systems. The first method is based on correlation matching technique to create circular filters whose external region matches the image background and the internal region matches the structure desired to locate. In the second method, structures are located from their contrast values calculated with circular filters. Both methods proved to be effective in detecting structures. However, the first one allowed the total image processing and not just a region of interest, which makes it quite advantageous.

1. Introdução

A qualidade da imagem em mamografia pode ser assegurada através da realização de testes periódicos. Os requisitos técnicos da mamografia estabelecidos na Portaria nº 453/98 do Ministério da Saúde [1] incluem parâmetros de qualidade da imagem como detalhes de alto contraste e o limiar de baixo contraste. Estes parâmetros podem ser medidos com imagens de *phantom*, possibilitando delimitar até que ponto um objeto se distingue do fundo da imagem.

O *phantom* CDMAM 3.4 [2] foi confeccionado especificamente para a realização deste teste e constitui-se de uma placa de alumínio envolvida por material acrílico, contendo uma matriz. Quatro outras placas acrílicas simulam a espessura da mama. Em cada célula da matriz estão distribuídos randomicamente dois discos de ouro idênticos. Os discos têm espessuras entre 0,03µm e 2,0µm e diâmetros entre 0,06mm e 2,0mm, variando numa escala logarítmica. A matriz é rotacionada 45° de forma a eliminar as estruturas que são facilmente detectáveis e as que certamente não serão detectadas. A Figura 1 mostra uma fotografia do *phantom* CDMAM 3.4.

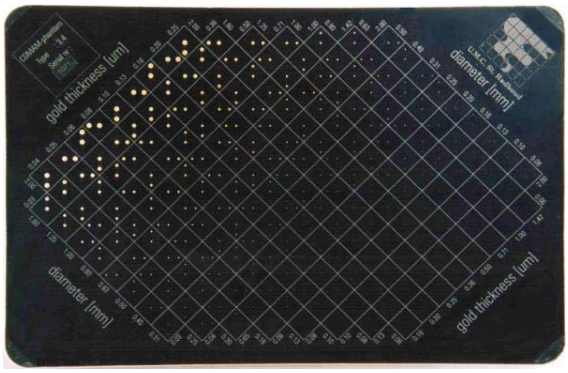


Figura 1. Vista frontal do simulador CDMAM 3.4 – Artinis Contrast-Detail Phantom.

Conforme recomendado pelo Protocolo Europeu [3] o teste deve ser efetuado anualmente por três observadores que fazem a leitura de duas imagens conferindo a posição dos discos com um gabarito, presente no manual do *phantom*, a fim de determinar um limiar de contraste. A leitura da imagem do *phantom* é considerada um tanto cansativa e consome muito tempo para executá-la. Além disso, tem significativa dependência da subjetividade do leitor, o que pode proporcionar erros inter-observador e entre observadores diferentes. A fim de minimizar esta subjetividade, um *software*, CDCOM, é proposto por Karssemeijer [4]. O CDCOM procura localizar os discos corretamente em imagens no formato DICOM. No entanto, um método ótimo para interpretar os dados não é definido, impossibilitando a busca pelo limiar de contraste, ou mesmo traçar os parâmetros do *phantom* que possam indicar a qualidade da imagem.

Nesse sentido, o propósito deste trabalho é o desenvolvimento de técnicas computacionais com a finalidade de detectar os discos na imagem do *phantom*, funcionando como um auxílio ao profissional que executa

o controle de qualidade dos mamógrafos digitais, possibilitando reduzir a subjetividade devida aos observadores.

Duas metodologias foram desenvolvidas e comparadas em relação à eficiência na detecção e ao custo computacional. A primeira baseia-se na técnica de casamento por correlação e constitui-se na elaboração de filtros circulares cuja região externa casa-se com o fundo da imagem e a região interna casa-se com a estrutura (disco) que se deseja localizar. Na segunda, os discos foram localizados a partir dos seus contrastes, calculados por meio de filtros circulares que executam a média dos valores de pixel tanto do fundo da imagem quanto da estrutura em questão.

2. Metodologia

Foram obtidas 27 imagens do *phantom* CDMAM 3.4 em cinco mamógrafos CR e um modelo digital (DR), gentilmente cedidas pelo Laboratório de Qualificação de Imagens Médicas (QualIM) – UNIFESP. As técnicas de kVp e mAs foram utilizadas conforme previsto no Protocolo Europeu. Doze delas foram adquiridas por meio do mamógrafo digital Selenia® Hologic/Lorad e as demais foram obtidas em cada um dos modelos de CR (três imagens por mamógrafo): Agfa 75, Agfa 85, Fuji 50, Fuji 100 e Kodak 975.

A primeira metodologia foi desenvolvida utilizando linguagem Java e com o emprego de *plugins* na interface do programa ImageJ®. Isso evitou a necessidade de se criar uma nova interface, possibilitando, ainda, a facilidade de uso dos comandos para o operador. A Figura 2, a seguir, mostra a interface do programa ImageJ® com o *plugin* de detecção das estruturas já acoplado.

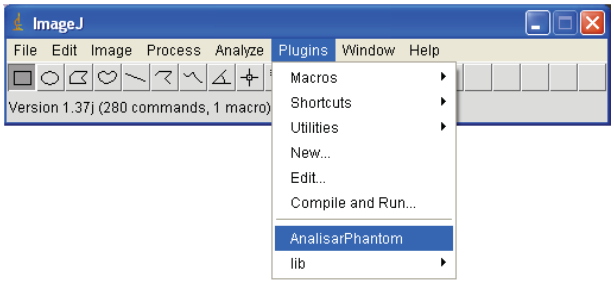


Figura 2. Interface do programa ImageJ® com o *plugin* de detecção das estruturas já acoplado.

O primeiro passo consistiu em localizar o ponto inicial de varredura do programa. Para isso, foi utilizado o detector de bordas Sobel que permitiu a aplicação da Transformada de Hough, a partir da qual foram encontradas as coordenadas dos vértices das células da matriz no plano (x,y).

Para a detecção dos discos foi utilizada a técnica de casamento por correlação [5] que, neste trabalho, foi empregada a partir da confecção de filtros circulares concêntricos compostos por uma região interna e uma externa. A região interna consiste em um circulo que abrange a estrutura e procura casar-se com o seu interior, enquanto a região externa constitui-se de um anel que procura casar-se com o plano de fundo. Os filtros foram confeccionados variando-se os seus raios de acordo com o raio dos discos presentes em uma imagem do *phantom* que serviu de gabarito. Na Figura 3 é mostrado um exemplo de filtro utilizado.

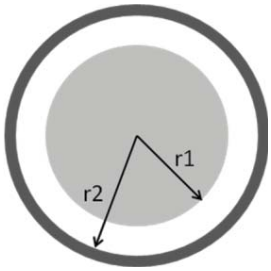


Figura 3. Modelo de filtro utilizado na detecção dos discos. Os raios r1 e r2 variam de acordo com o raio da estrutura em estudo, segundo uma imagem de referência (gabarito).

A etapa de localização das estruturas recebe uma imagem do *phantom* como entrada e devolve uma lista de estruturas marcadas para cada estrutura detectada. A imagem é girada em 45°, possibilitando que a varredura seja feita linha a linha da imagem, facilitando a busca.

O mecanismo que decide se uma estrutura é ou não detectada baseia-se no contraste entre a estrutura e o plano de fundo e no nível de cinza do plano de fundo.

Na segunda metodologia desenvolvida, também são utilizados filtros circulares concêntricos, conforme a Figura 3. O filtro interno abrange a estrutura e calcula a média dos valores de pixel contidos dentro do círculo. O filtro externo corresponde a um anel que calcula a média dos valores de pixel de uma região que abrange o fundo da imagem. O número de pixels do círculo e do anel é o mesmo. O contraste é calculado pela diferença entre as médias dos valores de pixel do fundo da imagem e da estrutura de interesse.

Utilizou-se uma rotina em MATLAB R2010a, em que o usuário seleciona uma região da imagem dando um *zoom* nas estruturas de interesse. Em seguida, ele seleciona o centro da estrutura a ser identificada e insere o valor do seu raio, de acordo com uma imagem de referência. Esta etapa é visualizada na Figura 4. Caso o raio não seja definido, o programa está preparado para executar o procedimento considerando um raio de 15 pixels. Neste momento, é feito o cálculo do contraste, cujo valor é utilizado para estabelecer um intervalo de

contraste para o qual é possível considerar a estrutura detectável.

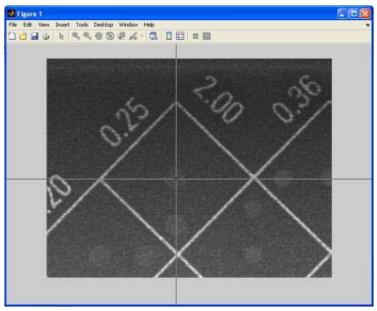
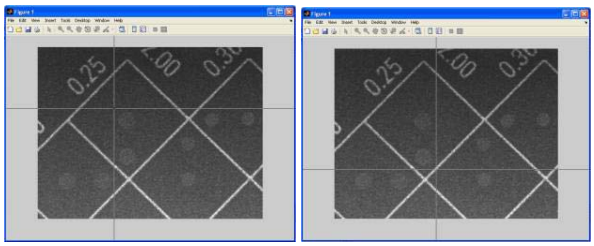


Figura 4. Etapa de seleção do centro da estrutura que se deseja localizar.

Em seguida, o usuário define os vértices da região de busca, conforme a Figura 5. O processo de busca pode ser realizado em intervalos definidos como passo. Caso o usuário não o defina, o programa efetua a busca com passo 1, ou seja, pixel a pixel. As estruturas encontradas são demarcadas com um círculo.



(a) (b)

Figura 5. Definição do primeiro (a) e segundo (b) vértices da região de busca.

Apesar dessa segunda metodologia não empregar a interface do ImageJ®, os comandos executados pelo usuário são muito simples, não exigindo muito esforço na localização das estruturas.

3. Resultados

No método de casamento por correlação, a varredura da imagem permitiu que a busca fosse feita abrangendo a imagem toda, não somente uma região de interesse. Ao detectar uma estrutura, o programa realiza uma demarcação no local do disco encontrado, como pode ser visualizado na Figura 6. No detalhe, um exemplo de duas estruturas localizadas em uma das células da matriz.

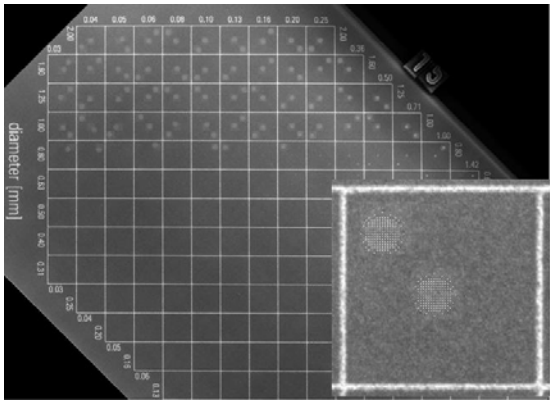


Figura 6. Resultado da detecção dos discos com o método de casamento por correlação em execução.

A ideia inicial não era a de detectar a posição exata dos discos, bastando somente detectá-los no centro e em um dos vértices de cada célula da matriz. Assim, foi possível detectar todos os discos da imagem do *phantom*, inclusive as não visíveis.

Na segunda metodologia empregada, a que utiliza a média dos valores de pixel dos filtros, as medidas foram efetuadas abrangendo somente os diâmetros considerados visíveis segundo o laudo técnico, a fim de se reduzir o tempo de execução e o custo computacional, já que a busca é realizada célula por célula.

Conforme já mencionado, o usuário tem a possibilidade de delimitar intervalos de busca definindo passos. Após a realização de alguns testes, optou-se por utilizar apenas o passo 1 e os raios foram definidos conforme uma imagem de referência. Verificou-se que, na região de interesse que abrangia uma célula da matriz, ambos os discos eram localizados e demarcados. O centro de cada demarcação foi armazenado como variável. No entanto, cada disco foi demarcado mais de uma vez, como pode ser verificado na Figura 7.

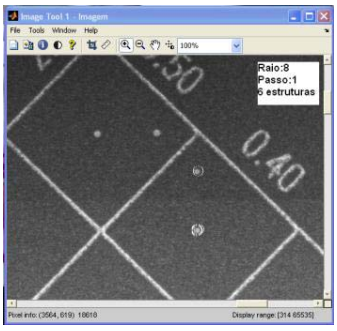


Figura 7. Resultado da detecção dos discos para de espessura 2,0µm e diâmetro 0,4mm, com o método das médias dos filtros em execução.

Em alguns casos, o programa detectou erroneamente uma das bordas da matriz como sendo uma estrutura, conforme mostrado na Figura 8.

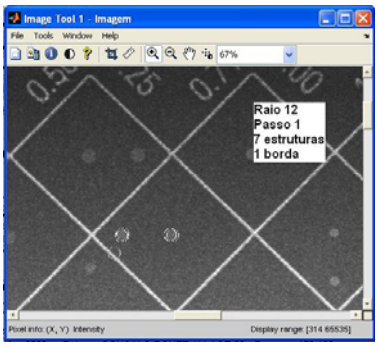


Figura 8. Resultado da detecção dos discos para de espessura 0,5µm e diâmetro 1,0mm, verificando-se a detecção de uma das bordas da matriz.

Isso se deve ao fato de os níveis de cinza da borda ser muito próximos aos do disco nesta região do *phantom*, dificultando a busca, nos casos em que a região selecionada pelo usuário abrangia parte das bordas da matriz da imagem. Dessa forma, verificou-se que a detecção ainda pode ser influenciada pelo usuário, considerando que é ele quem seleciona a região de busca. Assim, pode-se dizer que tal método é semi-automatizado.

Conforme a busca foi sendo executada em regiões com espessuras menores dos discos e raios maiores, a dificuldade de detecção foi aumentando, considerando que o disco já ocupa boa parte da região de busca. Isso limita a região que pertence ao fundo da imagem, assumindo ainda, que, nesta região do *phantom,* o contraste é menor. Assim, o número de vezes em que os discos foram detectados aumentou significativamente. Este erro pode ser amenizado aumentando-se o passo da busca, conforme visualizado na Figura 9.

(a) (b)

Figura 9. Resultado da detecção dos discos para de espessura 0,13µm e diâmetro 2,0mm utilizando (a) passo de 1 pixel e (b) passo de 3 pixels.

No entanto, essa melhora ocorre somente para os discos de diâmetros maiores. Ao se aplicar o passo 3 para os discos com diâmetros inferiores a 0,5 mm, o erro do número de detecções volta a aparecer. Dessa forma, os resultados foram obtidos utilizando-se apenas o passo 1 como intervalo de busca, pois, assim como no método de casamento por correlação, não foi necessário obter-se a posição exata dos discos. Portanto, consideraram-se detectadas corretamente as estruturas ainda que quando indicadas mais de uma vez.

Para a maioria dos casos com estruturas limiares selecionadas pelo laudo, foram detectados apenas um dos discos dentro da célula. A taxa de acertos considerada envolveu todas as detecções corretas que indicavam ambos os discos dentro de uma célula em cada imagem. Foram consideradas erradas as detecções em que apenas um dos discos da célula foi detectado e os casos em que a borda da matriz também era detectada. A Tabela 1, a seguir, mostra a taxa de detecção dos discos para a segunda metodologia empregada.

Tabela 1. Taxa de acertos e erros de detecção do método da média dos valores de pixel.

Taxa de acertos	83%
Detecção bordas da matriz	6%
Detecção de apenas um disco na célula	11%
Total de erros	17%

Tais erros, por vezes, podem ser corrigidos alterando-se o passo da busca ou diminuindo a área de interesse, o que, no entanto, torna a técnica um tanto dependente do usuário.

4. Conclusões

A primeira metodologia utilizada mostrou-se muito eficaz na detecção dos discos e com a vantagem de executar a varredura na imagem inteira, sem consumir muito tempo com baixo custo computacional. Além disso, a interface por meio do programa ImageJ® tornou o método completamente automatizado e simples de ser executado. Todos os discos puderam ser detectados.

O segundo método, apesar de também apresentar uma interface simples, mostrou-se um tanto dependente do usuário, causando variações nos resultados. O fato de haver detectado as estruturas mais de uma vez não implicou problemas, já que a intenção era apenas verificar a presença dos discos no centro e em um dos cantos das células.

A varredura por regiões de interesse e não da imagem completa implica maior tempo despendido na execução e torna-se um pouco cansativa para o usuário. No entanto, a técnica também se mostrou eficiente, pois detectou os

discos em 83% dos casos, podendo ser muito útil, se aprimorada.

Assim, considerou-se a técnica de casamento por correlação mais adequada ao procedimento, considerando que esta poderá tornar-se uma importante ferramenta de auxílio aos testes de qualidade ao ser correlacionada com a visão humana, possibilitando detectar somente as estruturas visíveis. É importante mencionar ainda, que, novas tarefas poderão ser incorporadas ao programa, sendo este um ponto de partida para a otimização do processo de obtenção parâmetro de qualidade da imagem descrito pela curva de contraste-detalhe.

5. Agradecimentos

Os autores agradecem a CAPES pelo apoio financeiro e ao Laboratório de Qualificação de Imagens Médicas (QualIM) – UNIFESP por obter e disponibilizar as imagens utilizadas na execução deste trabalho.

6. Referências

[1] Secretaria de Vigilância Sanitária de Ministério da Saúde. *Diretrizes de proteção radiológica em radiodiagnóstico médico e odontológico*. Portaria N°453, Brasília, 01/Jun/1998.

[2] K.R. Bijker, M.A.O Thijssen and Th.J.M. Arnoldussen. *Manual of CDMAM-phantom type 3.4*. University Medical Centre Nijmegen, St. Radboud, The Netherlands, 2002.

[3] N. Perry, M. Broeders, C. Wolf, S. Törnberg, R. Holland, L. Von Karsa. *European Guidelines for Quality Assurance in Mammography Screening and Diagnosis.* Fourth Edition. European Communities, Luxembourg, 2006.

[4] N. Karssemeijer, M.A.O. Thijssen. "Determination of contrast-detail curves of mammography systems by automated image analysis." *In: Proceedings of the 3rd International Workshop on Digital Mammography*, Elsevier, Amsterdam, 1996, pp.155-160.

[5] R.C. Gonzalez, R.E. Woods. *Digital Image Processing*, 2nd ed., Prentice Hall, New Jersey, USA, 2002.

Uso de morfologia matemática para remoção de ruídos da subtração de fundo

Wagner Machado do Amaral, Clésio Luis Tozzi (Orientador)
Departamento de Engenharia de Computação e Automação Industrial (DCA)
Faculdade de Engenharia Elétrica e de Computação (FEEC)
Universidade Estadual de Campinas (Unicamp)
Caixa Postal 6101, CEP 13083-970 – Campinas, SP, Brasil
{machado, clesio}@dca.fee.unicamp.br

Abstract

Background subtraction is a set of techniques used to extract information of moving objects in the scene. The main idea of these techniques is to build a model representing the background. This model is subtracted from each frame of the video and those regions of the frame that differs from the background models are considered to be the moving objects. Environment situations and camera quality may cause noise that difficult the moving objects identification. The goal of this work is to analyze the usage of binary image processing by mathematical morphology to deal with background subtraction noise.

1. Introdução

O desenvolvimento tecnológico tem tornado as câmeras de vídeo um equipamento bastante utilizado em nossa sociedade. Devido ao barateamento e à facilidade de manuseio desse tipo de equipamento, hoje é comum encontrar câmeras de vídeo em bancos, condomínios residenciais, aeroportos, estradas, estádios, entre outros lugares.

Assim como as câmeras de vídeo, os computadores têm evoluído com o avanço da tecnologia. O grande poder de processamento dos computadores atuais possibilita a criação de programas capazes de manipular as imagens adquiridas pelas câmeras de vídeo e extrair informações relacionadas ao ambiente filmado. Dentre as possíveis aplicações de programas que extraem informações de imagens, podemos destacar: rastreamento de pessoas e identificação de atividades de interesse, como caminhar ou correr [5]; segurança e monitoramento de ambientes [2]; detecção de iterações de pessoas com objetos, como carregar uma mochila ou abandonar um objeto na cena [2]; obtenção da trajetória de objetos deslocando-se na cena [1][4][6].

Sistemas que utilizam imagens obtidas por câmeras fixas podem utilizar uma importante técnica para detectar pessoas e objetos em movimento na cena, a subtração de fundo. Técnicas de subtração de fundo são baseadas na construção de um modelo que representa o fundo da cena, chamado de *background*. Este modelo é comparado à cada quadro do vídeo. As regiões do quadro que diferem consideravelmente do modelo do fundo são consideradas pertencentes aos objetos detectados na cena. Essas regiões são denominadas *foreground*.

Fatores relacionados ao ambiente filmado e à câmera utilizada podem fazer com que o resultado da subtração de fundo apresente falhas. *Pixels* pertencentes ao fundo podem ser erroneamente identificados como *foreground* e *pixels* pertencentes a objetos em movimento podem ser identificados como *background*. Esses *pixels* são considerados ruídos da subtração de fundo.

Alguns sistemas presentes na bibliografia eliminam ruídos resultantes da subtração de fundo através do uso de técnicas de processamento de imagens binárias por morfologia matemática [2][5].

O objetivo desse trabalho é analisar a aplicabilidade das técnicas de processamento de imagens binárias por morfologia matemática para eliminar ruídos resultantes da subtração de fundo.

O presente trabalho está dividido em cinco seções. A seção 1 apresenta a motivação e o objetivo do trabalho de pesquisa. A seção 2 descreve a origem dos ruídos e apresenta os resultados da subtração de fundo implementada nesse trabalho. Conceitos fundamentais relacionados ao processamento de imagens binárias por morfologia matemática são apresentados na seção 3. Maneiras de eliminar ruídos da subtração de fundo através do uso de morfologia matemática são abordadas na seção 4. A seção 5 apresenta conclusões.

2. Subtração de Fundo

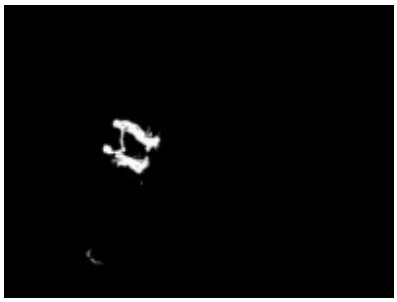
Conforme citado anteriormente, técnicas de subtração de fundo são baseadas na construção de um modelo que representa o fundo da cena, chamado de *background*. Este modelo é comparado à cada quadro do vídeo. As regiões do quadro que diferem consideravelmente do modelo do fundo são consideradas pertencentes aos objetos detectados na cena. Essas regiões são denominadas *foreground*.

No presente trabalho, a subtração de fundo utilizada no sistema de rastreamento W4 [2] foi implementada. Essa técnica foi desenvolvida para detectar objetos em movimento em tempo real, e não exige alto recurso computacional. Sua implementação é de fácil

compreensão, e os resultados são satisfatórios para o objetivo desse trabalho.

A Figura 1 (a) e a Figura 1 (b) apresentam respectivamente um quadro de um vídeo da base CAVIAR [8] e o resultado da subtração de fundo desse quadro. Nessa imagem é possível identificar dois homens cumprimentando-se.

(a) Quadro

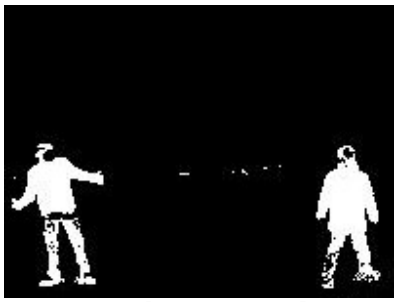


(b) *Foreground*

Figura 1. Subtração de fundo aplicada a um vídeo da base CAVIAR.

A Figura 2 (a) e a Figura 2 (b) apresentam respectivamente um quadro de um vídeo da base PETS [7] e o resultado da subtração de fundo desse quadro.

(a) Quadro



(b) *Foreground*

Figura 2. Subtração de fundo aplicada a um vídeo da base PETS.

Nos resultados da subtração de fundo implementada nesse trabalho, apresentados nas Figuras 1 e 2, é possível observar falhas na identificação do *foreground*.

Técnicas de subtração de fundo mais complexas, como a Mistura de Gaussianas [9], são capazes de incorporar movimentos oscilatórios, como galhos de árvores movimentando-se ao vento, ao modelo do fundo. Na Figura 2 (a) é possível observar uma fita que movimenta-se ao vento. Como a técnica de subtração de fundo implementada nesse trabalho não é capaz de incorporar ao modelo essa movimentação, na Figura 2 (b) pequenas regiões foram erroneamente identificadas como *foreground*.

Na figura 2 (b) é possível identificar ainda *pixels* nas regiões onde as pessoas estão posicionadas que foram erroneamente identificados como fundo. Considerando que a subtração de fundo implementada nesse trabalho processa imagens monocromáticas (conforme [2]), essa falha ocorreu porque a intensidade dos *pixels* pertencentes às pessoas é semelhante a intensidade dos *pixels* pertencentes ao fundo nessas posições. Sendo assim, esses *pixels* não foram identificados como *foreground*.

Essas falhas prejudicam a contagem de objetos na cena, e a identificação da forma de cada objeto detectado. Alguns sistemas presentes na bibliografia [2][5] corrigem essas falhas através do uso de técnicas de processamento de imagens por morfologia matemática. Nas seções a seguir, essas técnicas serão apresentadas e a usabilidade das mesmas para esse propósito será avaliada.

3. Morfologia Matemática

Conforme [3], o ramo do processamento e análise de imagens por morfologia matemática foi inicialmente desenvolvido por Georges Matheron e Jean Serra. O escopo dos métodos morfológicos é tão extenso quanto o escopo do processamento de imagens. Isso inclui segmentação, análise de partículas, restauração, compressão, análise de formas, esqueletonização, detecção de bordas, análise de textura, etc. Há muitas

áreas de aplicação da morfologia matemática, entre elas a visão robótica, processamento de imagens médicas, biologia e análise de documento digitais.

A ideia básica do processamento e análise de imagens por morfologia matemática é a de provar uma imagem com um elemento estruturante e avaliar quanto o elemento estruturante se "encaixa" (ou não) na imagem. O elemento estruturante geralmente é uma matriz com valores iguais a zero ou um.

Erosão

A erosão é uma das mais básicas operações da morfologia matemática. Uma possível forma de erodir uma imagem binária é mover a origem de um elemento estruturante por toda a imagem e marcar todas as posições da origem nas quais o elemento estruturante estiver totalmente contido na imagem. Mover um elemento estruturante é uma operação de translação definida como

$$B_x = \{b + x : b \in B\} \quad (1)$$

A erosão da imagem binária A pelo elemento estruturante B, denotada por $A \ominus B$, é definida como

$$A \ominus B = \{x : B_x \subset A\} \quad (2)$$

e consiste em todos os pontos x, tal que B transladado por x está contido em A.

Dilatação

A segunda operação básica da morfologia matemática é a dilatação. Uma possível forma de dilatar uma imagem binária por um elemento estruturante é mover a origem do elemento estruturante por toda a imagem e marcar todas as posições de cada ponto do elemento estruturante como resultado da dilatação. Conforme Dougherty e Lotufo [3], esse procedimento pode ser comparado ao processo de pintura, onde o elemento estruturante é um pincel e os pixels da imagem a ser dilatada apontam os lugares onde o centro do pincel deve ser passado.

A dilatação da imagem binária A pelo elemento estruturante B, denotada por $A \oplus B$, é definida como

$$A \oplus B = \bigcup_{a \in A} B_a \quad (3)$$

e consiste na união das translações de B por a, sendo $a \in A$.

Uma sequencia de n dilatações da imagem A pelo elemento estruturante B é definida por

$$(A \oplus B)^n = (((A \oplus B) \oplus B) \oplus ... \oplus B) \quad (4)$$

Uma forma de limitar o crescimento da imagem resultante de uma dilatação é realizar a intersecção da mesma com uma imagem que define os limites desejados para o crescimento. Seja A uma imagem contida em C, e B um elemento estruturante, a dilatação de A por B condicionada a C é definida como

$$A \oplus_c B = (A \oplus B) \cap C \quad (5)$$

Abertura

A abertura de uma imagem binária A pelo elemento estruturante B é definida como uma erosão seguida de uma dilatação, conforme abaixo

$$A \circ B = (A \ominus B) \oplus B \quad (6)$$

Fechamento

O fechamento de uma imagem binária A pelo elemento estruturante B é definida como uma dilatação seguida de uma erosão, conforme abaixo

$$A \bullet B = (A \oplus B) \ominus B \quad (7)$$

Reconstrução

Uma importante operação da morfologia matemática é a reconstrução por marcadores. Essa operação considera a imagem A como a união dos n componentes conexos,

$$A = \bigcup_{k=1}^{n} C_k \quad (8)$$

onde C_k corresponde a um componente conexo de A. A imagem M é um marcador, tal que $M \subset A$. A reconstrução morfológica de A pelo marcador M é a união dos componentes conexos de A que intersetam M, conforme abaixo

$$A \Delta_B M = \bigcup \{C_k : C_k \cap M \neq 0\} \quad (9)$$

onde B é o elemento estruturante que define a conectividade.

A reconstrução por marcadores pode ser formulada como n - dilatações condicionais (4). A reconstrução da imagem binária A pelo marcador M com conectividade definida pelo elemento estruturante B pode ser definida como repetidas dilatações (até estabilizar) de M pelo elemento estruturante B condicionadas a A,

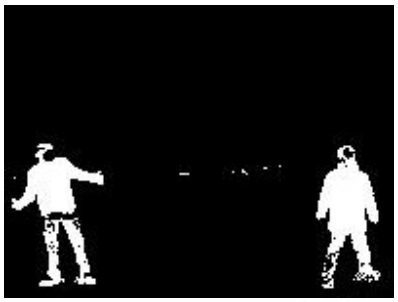
$$A \Delta_B M = (M \oplus_A B)^{\infty} \quad (10)$$

4. Remoção de Ruídos

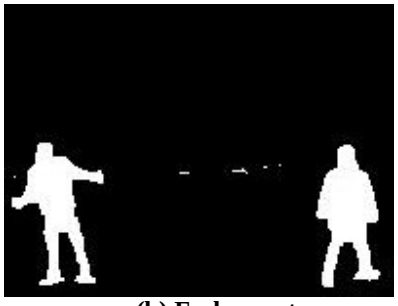
Os *pixels* erroneamente identificados como fundo são chamados de falsos negativos. Na Figura 2 (b) é possível observar alguns falsos negativos que causam alguns "buracos" nos componentes conexos que representam cada pessoa. Uma possível maneira de eliminar esses "buracos" é aplicando o operador morfológico de fechamento na imagem, conforme equação (7).

A dilatação da imagem expandira a área dos componentes conexos. Dessa forma, pequenos "buracos" nos componentes conexos serão eliminados. A erosão da imagem dilatada fará com que os componentes conexos reduzam para um tamanho próximo ao tamanho anterior.

Esse procedimento pode causar algumas distorções na forma dos componentes conexos, porém é bastante útil por eliminar falsos negativos na detecção do *foreground*. Conforme demostrado na Figura 3.



(a) *Foreground*



(b) Fechamento

Figura 3. Remoção de ruídos por Fechamento morfológico.

Os *pixels* erroneamente identificados como *foreground* são chamados de falsos positivos. Na Figura 2 (b) é possível observar alguns falsos positivos causados por uma fita movimentando-se ao vento. Uma possível maneira de eliminar essas regiões é aplicando o operador morfológico de abertura na imagem, conforme equação (6).

A Figura 4 apresenta o resultado da aplicação do operador morfológico de abertura à imagem exibida na Figura 3 (b).

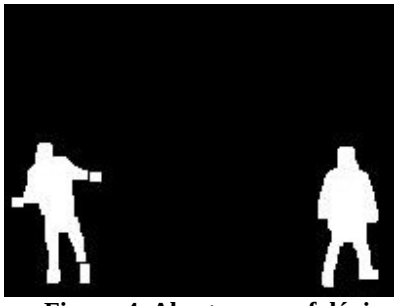


Figura 4. Abertura morfológica.

A erosão faz com que os componentes conexos menores que o elemento estruturante utilizado sejam eliminados. Esse procedimento pode ser utilizado para remover pequenas regiões de falso positivo. A dilatação faz com que os componentes conexos restantes retornem a um tamanho próximo ao anterior.

Observe na Figura 4 que o componente conexo que representa a pessoa da esquerda foi segmentado em três partes, devido à distorções causadas no formato do componente conexo pela erosão e dilatação. Para evitar que essas distorções ocorram, a técnica denominada Reconstrução por Marcadores pode ser utilizada.

Conforme equação da reconstrução por marcadores (10), esse procedimento pode ser interpretado como a reconstrução da subtração de fundo A pelo marcador $A \ominus B$.

$$A \Delta_B (A \ominus B) = [(A \ominus B) \oplus_A B]^{\infty} \qquad (11)$$

Esse procedimento elimina pequenas regiões de falso positivo e não distorce a forma dos demais componentes conexos.

A Figura 5 apresenta o resultado da aplicação da reconstrução à imagem exibida na Figura 3 (b).

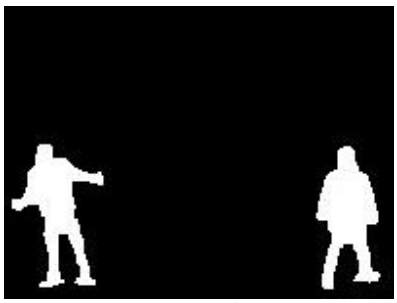
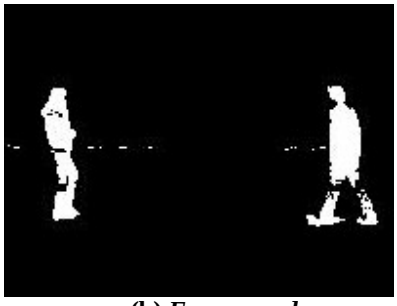


Figura 5. Reconstrução por marcadores.

Observe que as distorções observadas na Figura 4 não ocorreram na Figura 5.

A seguir nas Figuras 6 e 7, mais testes são realizados. Para cada teste, é apresentada a imagem original do quadro, o resultado da subtração de fundo, e a remoção de ruídos através do uso de fechamento e reconstrução.

(a) Quadro

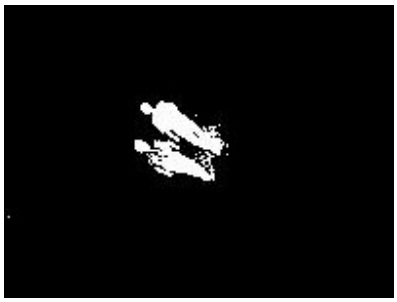


(b) *Foreground*

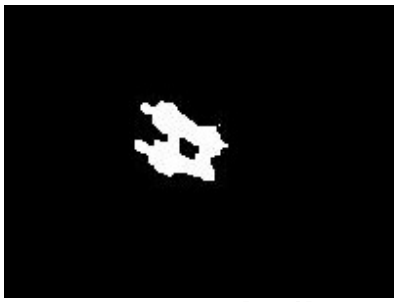
(c) Remoção de ruídos

Figura 6. Teste de remoção de ruídos 1

(a) Quadro



(b) *Foreground*



(c) Remoção de ruídos

Figura 7. Teste de remoção de ruídos 2

5. Conclusão

Nesse trabalho, foi analisada a aplicabilidade das técnicas de processamento de imagens binárias por morfologia matemática para eliminar ruídos resultantes da subtração de fundo.

A subtração de fundo utilizada no sistema de rastreamento W4 [2] foi implementada e testada. Os resultados obtidos foram satisfatórios para o objetivo desse trabalho.

Operadores morfológicos foram utilizados para eliminar ruídos da subtração de fundo implementada nesse trabalho.

O uso do operador morfológico de fechamento mostrou-se eficaz para eliminar falsos negativos. O uso dos operadores morfológicos de abertura e de reconstrução por marcadores mostraram-se eficazes para eliminar falsos positivos. Porém, o segundo apresentou menor distorção na forma dos componentes conexos.

Para os testes realizados, boa parte dos ruídos foram eliminados. Com isso, houve uma melhora na qualidade do resultado da subtração de fundo após a aplicação de operadores morfológicos. Essa constatação valida o uso de tais técnicas para o propósito de remoção de ruídos da subtração de fundo.

Referências

[1] G. M. Freitas; C. L. Tozzi. Rastreamento de objetos em vídeos e separação em classes. 2010. 102 p. Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

[2] I. Haritaoglu, D. Harwood, and L. S. Davis. W4: real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8):809-830, August 2000.

[3] DOUGHERTY, Edward R.; LOTUFO, Roberto de Alencar. “Hands-on morphological image processing”. Bellingham: SPIE, 2003. 272 p., il. (Tutorial texts in optical engineering; v. v. TT59). ISBN 081944720X.

[4] R. R. Pinho, J. M. R. S. Tavares. Tracking Features in Image Sequences with Kalman Filtering, Global Optimization, Mahalanobis Distance and a Management Model. Computer Modeling in Engineering & Sciences, Tech Science Press, ISSN: 1526-1492 (print) - 1526-1506 (online), vol. 46, no. 1, pp. 51-75, 2009.

[5] T. Zhao and R. Nevatia. Tracking multiple humans in complex situations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26:1208–1221, 2004.

[6] B. Lei and L.-Q. Xu. Real-time outdoor video surveillance with robust foreground extraction and object tracking via multi-state transition management. Pattern Recognition Letters, 27(6):1816 – 1825, 2006. Vision for Crime Detection and Prevention.

[7] James Ferryman. Pets dataset. http://www.cvg.rdg.ac.uk/VS/, 2009.

[8] CAVIAR. Ec funded caviar project/ist 2001 37540. http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/, 2001.

[9] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.,
volume 2, pages –252 Vol. 2, 1999.

Sistema Para Controle de Qualidade de Mamógrafos Através da Análise Computacional de Imagens de *Phantom* Utilizando a Ferramenta WEKA

Bruno Barufaldi,
Eduardo F. Santana,
Leonardo V. Batista
Programa de Pós-graduação em Informática
Universidade Federal da Paraíba
barufaldi@di.ufpb.br

Homero Schiabel
Escola de Engenharia de São Carlos
Universidade de São Paulo

José F. G. Carvalho
Agência Estadual de Vigilância Sanitária da Paraíba
Governo da Paraíba

Abstract

Breast cancer control represents one of the greatest challenges that public health faces nowadays. The use of breast phantoms allows the application of tests to measure the mammography acceptance and the monthly constancy testing of mammography quality control programs. Even so, evaluation by technicians still suffers limitations associated with visual inspection by human beings, such as subjectivity and long-time of benchmarking. It has been developed a computerized system that analyses radiological images of breast phantoms, using the WEKA package, to reduce subjectivity and assure the consistence of the system. Results indicate that the software *developed allows determining efficiently the visibility of structures of interest in the phantom images. Therewith, the system becomes a tool of great utility to the quality of medical diagnosis.*

1. Introdução

O câncer de mama é a segunda neoplasia mais incidente na população mundial feminina. A cada ano, cerca de 22% dos casos novos de câncer em mulheres são de mama [1]. De acordo com os dados da GLOBOCAN (*International Agency for Research on Cancer*), em 2008 foram registrados em todo o mundo 691 mil novos casos de câncer de mama em países em desenvolvimento e 692 mil em países desenvolvidos. No entanto, o número total de mortes pela doença em países em desenvolvimento foi de 268 mil, em contraste com 189 mil mortes nos países desenvolvidos [2].

No Brasil, a taxa bruta de mortalidade associada ao câncer de mama aumentou 69% entre 1979 e 2000 [3], o que pode ser parcialmente explicado pelo processo de envelhecimento populacional e exposição crescente a fatores de risco. Adicionalmente, a má distribuição de mamógrafos, os altos custos do exame e a falta de controle de qualidade das imagens tornam-se também fatores responsáveis pelo problema.

A mamografia é o método mais confiável e acessível para detecção do câncer de mama. A sua eficácia e indicação são aceitas mundialmente para o rastreamento da população com idade acima de 40 anos. Para as mulheres de grupos populacionais de risco (com histórico familiar de câncer de mama em parentes de primeiro grau), recomendam-se o exame clínico da mama e a mamografia, anualmente, a partir de 35 anos [4].

Apesar dos esforços para o aumento do número de mamógrafos para se detectar tumores não palpáveis, o tempo médio para diagnóstico e início do tratamento das pacientes com tumores palpáveis supera 120 dias [5]. A ampliação de programas de rastreamento nos países desenvolvidos tem aumentado substancialmente a carga de trabalho dos radiologistas, conduzindo à saturação iminente [6].

O mecanismo de inspeção do filme mamográfico é um procedimento custoso e predisposto a erros. Para a redução destas falhas, o Programa de Certificação da Qualidade em Mamografia do Colégio Brasileiro de Radiologia (CBR) exige que os serviços de mamografia efetuem o registro radiográfico de um simulador de mama, denominado *phantoms*.

A utilização de sistemas computadorizados de auxílio ao diagnóstico torna-se necessária para manter a qualidade das atividades e serviços clínicos. O objetivo principal da produção desses sistemas está em tentar eliminar a subjetividade na análise especialista, devido às limitações próprias do ser humano.

A automatização dos procedimentos de análise em imagens radiográficas por intermédio de sistemas especialistas de apoio ao diagnóstico, reduz significantemente a carga de trabalho de técnicos e erros médicos. Para o computador elaborar uma hipótese diagnóstica, é necessário que os especialistas informem

todos os dados relevantes observados, para que o sistema seja capaz de processar essas informações de maneira eficiente.

Este trabalho apresenta um sistema computacional que detecta as estruturas de interesse do *phantom* MAMA-CDM [7], e as correlaciona com a detecção por análise visual de especialistas, a partir de técnicas de inteligência artificial. Este software em desenvolvimento é voltado para automatização da maior parte das etapas do processo utilizado pelo Programa de Controle de Qualidade em Mamografia da Agência Estadual de Vigilância Sanitária da Paraíba (AGEVISA-PB).

2. Fundamentação Teórica

Esta seção contém a descrição dos simuladores radiográficos MAMA-CDM e dos algoritmos de árvores de decisão (J48) e modelos de aprendizagem (SMO) utilizados no desenvolvimento do software.

2.1. Simulador Radiográfico MAMA-CDM

Nos centros de radiodiagnósticos são utilizados rotineiramente simuladores radiográficos (ou *phantoms*) que contêm estruturas que representam os tecidos da mama. O principal objetivo dos *phantoms* está em simular a interação do feixe de raios X com o paciente, para avaliar a atenuação da radiação e qualidade da imagem no filme mamográfico.

O *phantom* MAMA-CDM simula uma mama comprimida entre 4cm e 5cm e possui detalhes que produzem imagens radiográficas semelhantes às estruturas presentes na mama (microcalcificações, fibras e massas tumorais).

São avaliados nas imagens dos simuladores a definição (resolução espacial), os detalhes de alto contraste (microcalcificações), os limiares de baixo contraste (discos), os detalhes lineares de baixo contraste (tecido fibroso), as massas tumorais e a densidade ótica de fundo. A Figura 1 apresenta a imagem do *phantom* MAMA-CDM e sua imagem radiográfica.

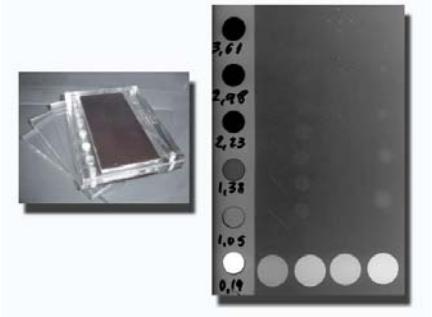


Figura 1. Simulador radiográfico MAMA-CDM e sua imagem radiográfica.

2.2. WEKA

O pacote WEKA (*Waikato Environment for Knowledge Analysis*) é formando por um conjunto de implementações de algoritmos utilizados em Mineração de Dados. Esse pacote foi desenvolvido na linguagem Java, o que garante a portabilidade do sistema, além de possuir todos os benefícios de uma linguagem orientada a objetos.

O WEKA inclui uma variedade de métodos de aprendizagem para o conjunto de dados a ser avaliado, denominados classificadores. Dentre os classificadores mais conhecidos estão as árvores de decisão e as Máquinas de Vetores de Suporte (*Support Vector Machines - SVM*).

A principal característica do WEKA está voltada na classificação de atributos por intermédio de seus algoritmos de aprendizagem. No entanto, também inclui seletores de atributos, e implementações de regras com associação e agrupamento de dados (*clusters*).

2.2.1 Algoritmo J48

As árvores de decisão têm diversas vantagens quando comparadas a outras técnicas. A principal está em sua simplicidade e facilidade na interpretação dos resultados. O J48 é a implementação otimizada do algoritmo de construção de árvores de decisão C4.5, que foi a última versão pública do algoritmo.

O algoritmo J48 é muito utilizado em classificadores envolvendo regras de decisão. Esse algoritmo constrói um modelo de árvore de decisão baseado em um conjunto de dados de treinamento e usa esse modelo para classificar as instâncias de um conjunto de testes.

Em alguns casos é possível resolver problemas complexos através de uma árvore de decisão simplificada. Muitos trabalhos comprovam a eficiência do método, mostrando que o J48 possui grande precisão e eficiência com relação a outros modelos [8].

2.2.2 Algoritmo SMO

A técnica SVM pertence a uma categoria de classificadores lineares e foi desenvolvida por Vladimir Vapnik, com o objetivo de auxiliar na solução de problemas de classificação e de reconhecimento de padrões [9].

O algoritmo de Otimização Mínima Sequencial (*Sequencial Minimal Optimization - SMO*) é utilizado na implementação otimizada do classificador SVM do WEKA. Seu custo computacional é linear para realização dos treinamentos de aprendizagem, o que permite lidar com uma quantidade de arquivos maiores nos testes.

Esse algoritmo transforma a saída do classificador em probabilidades através da aplicação de uma função sigmóide padrão [10]. Os atributos nominais são convertidos em binários e todos os valores numéricos são normalizados. Ao final do processo, o algoritmo produz um modelo de aprendizagem, utilizado na classificação dos atributos.

3. Materiais e Métodos

Uma das abordagens de inspeção visual do filme radiográfico consiste em determinar a visibilidade das estruturas de interesse presentes na imagem do *phantom* MAMA-CDM. O objetivo do trabalho está em realizar esse procedimento via software utilizando técnicas de classificação, empregando algoritmos de árvores de decisão e regras de aprendizagem através da ferramenta WEKA. A seção a seguir define as etapas de processamento para classificação das estruturas das imagens desse simulador radiográfico.

3.1. Extração de Atributos

Para a construção dos modelos que serão gerados pelos algoritmos J48 e SMO, é necessário extrair alguns atributos da imagem para realização da etapa de treinamento do sistema. É importante destacar que nem todos os atributos utilizados para treinamento serão empregados para classificação das estruturas, já que algumas das características da imagem não são consideradas relevantes pelos algoritmos.

As tabelas a seguir definem os atributos utilizados para realização da etapa de treinamento do sistema. Os atributos apresentados pela Tabela 1 serão extraídos de cada imagem, onde *p(i,j)* é o nível de cinza do pixel na posição (i,j) e, *w* e *h* são a largura e altura da imagem.

Tabela 1. Atributos globais extraídos da imagem do simulador radiográfico.

Atributo	Equação
Média dos pixels da imagem	$\mu = \sum_{i=1}^{w}\sum_{j=1}^{h}\frac{p(i,j)}{(w*h)}$
Média do módulo da diferença entre os pixels da imagem e a média da imagem	$x = \sum_{i=1}^{w}\sum_{j=1}^{h}\frac{\lvert p(i,j)-\mu\rvert}{(w*h)}$
Variância da imagem	$v = \sum_{i=1}^{w}\sum_{j=1}^{h}\frac{(p(i,j)-\mu)^2}{(w*h)}$
Desvio Padrão da imagem	$\sigma = \sqrt{v}$
Moda dos pixels da imagem	Nível de cinza mais frequente em *p(i,j)*

As estruturas de interesse do *phantom* são detectadas por meio da técnica de casamento por correlação [11]. Os filtros utilizados nas operações de correlação são compostos por duas partes, região interna e região externa, conforme ilustrado na Figura 2.

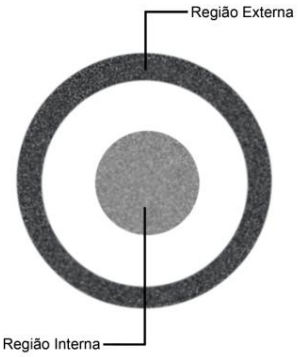


Figura 2. Exemplo de filtro utilizado para detecção e classificação de uma estrutura circular da imagem.

Para uma correlação elevada, a região interna deve ajustar-se ao interior da estrutura de interesse, enquanto a região externa deve ajustar-se ao plano de fundo da imagem. O sistema desenvolvido é capaz de extrair os atributos e localizar as estruturas de interesse com precisão. A Tabela 2 apresenta os atributos extraídos de cada estrutura, onde $p_e(i,j)$ e $p_b(i,j)$ são os níveis de cinza dos pixels da região interna (estrutura) e externa (plano de fundo) do filtro de tamanho $w * h$ na posição (i,j) da imagem.

Tabela 2. Atributos extraídos da imagem após a detecção das estruturas de interesse.

Atributo	Equação
Média dos pixels da estrutura	$\mu_e = \sum_{i=1}^{w}\sum_{j=1}^{h}\frac{p_e(i,j)}{(w*h)}$
Média dos pixels do plano de fundo	$\mu_b = \sum_{i=1}^{w}\sum_{j=1}^{h}\frac{p_b(i,j)}{w*h}$
Diferença das médias dos níveis de cinza dos pixels da estrutura e plano de fundo	$\Delta\mu = \mu_e - \mu_b$
Razão de Weber [11]	$W = \frac{\Delta\mu}{\mu_e}$

Os atributos de Haralick são frequentemente utilizados para classificação de texturas e reconhecimento de padrões [12]. Para realizar a classificação, é necessário

definir a matriz de co-ocorrência da região de interesse da estrutura detectada, que descreve as relações entre a tonalidade de cinza de um determinado pixel e os níveis de cinza dos pixels adjacentes [13]. Para o trabalho foram utilizados os atributos de contraste, correlação, entropia e segundo momento angular. A Tabela 3 descreve esses atributos de Haralick utilizados, onde: $p'(i,j)$ é contador normalizado de i seguido de j da matriz de co-ocorrência da imagem; μ_x, μ_y, σ_x, σ_y, são as médias e desvios padrões de p'_x e p'_y; $p'_x(i)$ é a i-ésima entrada da matriz de probabilidade obtida das linhas $p'(i,j)$; N_q é número de níveis de cinza quantizados da imagem.

Tabela 3. Atributos de Haralick extraídos da região de interesse das estruturas do simulador radiográfico.

Atributo	Equação
Segundo Momento Angular	$f_1 = \sum_i \sum_j \{p'(i,j)\}^2$
Contraste	$f_2 = \sum_{n=0}^{N_{q-1}} n^2 \left(\sum_{i=1}^{N_q} \sum_{j=1}^{N_q} p'(i,j) \right)$
Correlação	$f_3 = \frac{\sum_i \sum_j (i,j) p'(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Entropia	$f_9 = -\sum_i \sum_j p'(i,j) \log(p'(i,j))$

3.2. Treinamento

A etapa de treinamento consiste em produzir uma árvore de decisão e um modelo de aprendizagem para cada estrutura de interesse, que permanecerão estáticos durante o processamento das imagens pelo sistema na etapa de classificação. O treinamento é realizado utilizando imagens laudadas pelos especialistas do Programa de Controle de Qualidade de Mamógrafos da Agência de Vigilância Sanitária do Estado da Paraíba (AGEVISA-PB), de modo a ensinar a máquina para gerar a árvore de decisão e equação dos atributos pelos algoritmos J48 e SMO do pacote WEKA.

Foram utilizadas 300 estruturas para treinamento divididas entre discos de baixo contraste, fibras e massas tumorais. Essas estruturas foram classificadas por especialistas de acordo com sua visibilidade (visíveis ou não visíveis). É importante destacar que todas as imagens foram laudadas por dois ou mais técnicos especializados, de maneira a diminuir a subjetividade da inspeção visual e garantir a consistência do sistema.

Nessa etapa são criados os arquivos do formato padrão ARFF do WEKA, com todos os dados de entrada necessários para o treinamento do software. É possível selecionar os atributos relevantes para o treinamento através da própria ferramenta, na etapa de pré-processamento. Em seguida, utilizando o algoritmo J48 e SMO presentes no pacote WEKA, são gerados a árvore de decisão e equação de aprendizagem dos atributos do sistema para cada estrutura de interesse.

3.3. Classificação

Após a produção dos modelos gerados a partir do treinamento das imagens, é realizada a etapa de classificação. Nessa etapa foram implementadas as equações de aprendizagem e árvores de decisões, construídas através do processamento das imagens pelos algoritmos SMO e J48 da ferramenta WEKA.

Para os testes empíricos, foram utilizadas 100 imagens do simulador radiográfico para classificação, com um total de 1900 estruturas analisadas entre discos de baixo contraste, fibras e massas tumorais. É importante destacar que as imagens processadas na etapa de classificação são distintas das de treinamento.

O comportamento do software é visualizado através das curvas ROC (*Receiver Operating Characteristic*), em que são verificadas a sensibilidade e a especificidade do sistema, à medida que os valores dos atributos dos modelos criados variam.

4. Resultados

Para representação da eficácia do sistema, são apresentadas as tabelas de contingência (ou matriz de confusão), que indicam as taxas de acerto da classificação. A Tabela 4, 5 e 6 mostram as matrizes de confusão das massas, discos e fibras, classificadas de acordo com o modelo gerado pelos algoritmos J48 e SMO.

Tabela 4. Tabela de contingência das massas tumorais após a etapa de classificação.

		Classificado pelo Algoritmo J48		Classificado pelo Algoritmo SMO	
		Visível	Não Visível	Visível	Não Visível
Classificado pelo Especialista	Visível	0,915	0,085	0,920	0,080
	Não Visível	0,040	0,960	0,030	0,970

Tabela 5. Tabela de contingência dos discos de baixo contraste após a etapa de classificação.

		Classificado pelo Algoritmo J48		Classificado pelo Algoritmo SMO	
		Visível	Não Visível	Visível	Não Visível
Classificado pelo Especialista	Visível	0,922	0,078	0,955	0,045
	Não Visível	0,110	0,890	0,237	0,763

Tabela 6. Tabela de contingência do tecido fibroso após a etapa de classificação.

		Classificado pelo Algoritmo J48		Classificado pelo Algoritmo SMO	
		Visível	Não Visível	Visível	Não Visível
Classificado pelo Especialista	Visível	0,963	0,037	0,935	0,065
	Não Visível	0,414	0,586	0,414	0,586

A Tabela 7 e 8 apresentam as taxas de preditividade positiva e negativa, sensibilidade, especificidade, eficiência e acurácia do sistema, para a classificação de cada estrutura de interesse.

Tabela 7. Medidas de eficácia do software para classificação das estruturas utilizando o algoritmo J48.

J48	Massas	Discos	Fibras
Acurácia	0,924	0,915	0,875
Sensibilidade	0,989	0,968	0,886
Especificidade	0,738	0,758	0,820
Eficiência	0,864	0,863	0,853
Preditividade Positiva	0,915	0,922	0,961
Preditividade Negativa	0,960	0,890	0,590

Tabela 8. Medidas de eficácia do software para classificação das estruturas utilizando o algoritmo SMO.

SMO	Massas	Discos	Fibras
Acurácia	0,930	0,914	0,853
Sensibilidade	0,992	0,936	0,860
Especificidade	0,752	0,825	0,820
Eficiência	0,872	0,880	0,840
Preditividade Positiva	0,920	0,955	0,960
Preditividade Negativa	0,970	0,763	0,540

Para a construção das curvas ROC o valor do atributo $\Delta\mu$ foi alterado dentro de um certo intervalo, uma vez que este foi o atributo de maior relevância nas etapas de treinamento de todas as estruturas de interesse. Com a redução gradativa desses valores, é possível verificar o comportamento do sistema para a classificação das estruturas, conforme apresentado na Figura 3 e Figura 4.

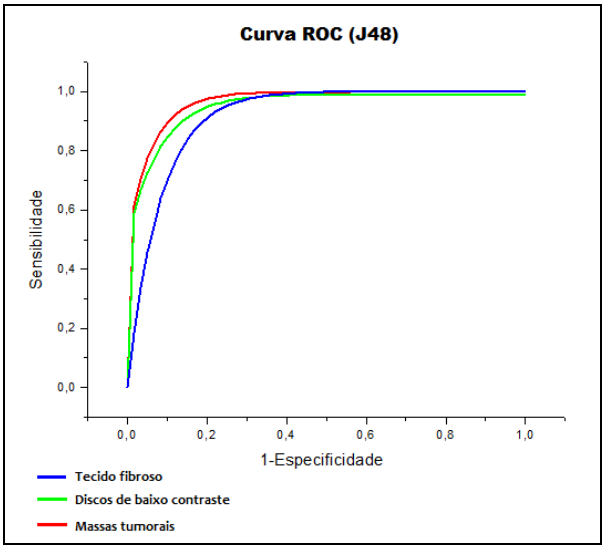


Figura 3. Curvas ROC do algoritmo J48 para classificação do tecido fibroso (Az = 0.745), discos de baixo contraste (Az = 0.887) e massas tumorais (Az = 0.892).

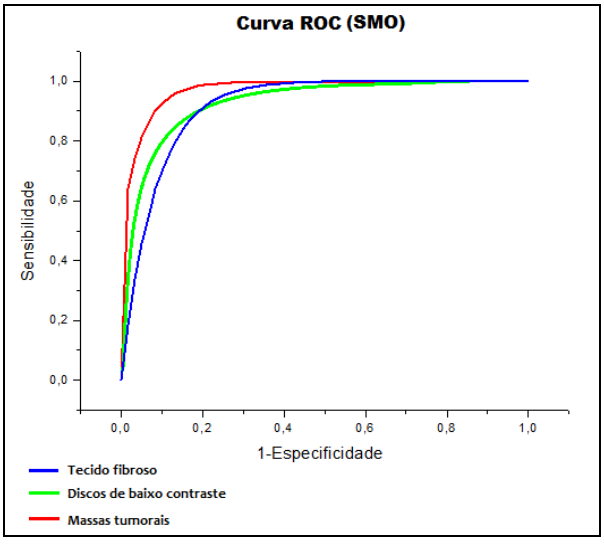


Figura 4. Curvas ROC do algoritmo SMO para classificação do tecido fibroso (Az = 0.74), discos de baixo contraste (Az = 0.859) e massas tumorais (Az = 0.945).

5. Discussões e Conclusão

As medidas de eficácia apresentadas nos resultados (Tabela 7 e 8) foram bastante semelhantes para a classificação das estruturas de interesse pelos algoritmos. Apesar de o algoritmo SMO totalizar uma maior taxa de classificação de estruturas corretas nas massas tumorais, a acurácia do algoritmo J48 é maior na maioria dos casos.

As estruturas que simulam as massas tumorais são as que apresentam maior definição em relação às outras estruturas analisadas. Por esse motivo, a decisão a respeito de sua visibilidade se torna mais fácil quando comparado as estruturas que representam os discos ou fibras. Esse fato colabora para os melhores resultados em relação às outras estruturas para classificação utilizando os algoritmos J48 e SMO.

Um dos problemas enfrentados está na presença de ruído que se assemelha a algumas estruturas. O tamanho reduzido das microcalcificações favorece a confusão entre estrutura e ruído, especialmente nos conjuntos menos evidentes. Estão sendo desenvolvidas técnicas para o tratamento de ruído dessas imagens para uma melhor detecção e classificação dessas estruturas.

Com a utilização do sistema desenvolvido irá representar uma redução do tempo e subjetividade na avaliação das imagens radiográficas, com a possibilidade das clínicas adquirirem e empregarem o sistema e, com isso, efetuarem de forma eficiente seu próprio controle de qualidade com a periodicidade adequada.

6. Agradecimentos

Ao CNPq, à CAPES e à SESU/MEC, pelo apoio na forma de bolsas de pesquisa e custeio.

7. Referências

[1] INCA, "Estimativa 2010: incidência de câncer no Brasil / Instituto Nacional de Câncer", ISBN 978-85-7318-161-6 (98), 2009.

[2] D.G. Tiezzi, "Câncer de mama: um futuro desafio para o sistema de saúde nos países em desenvolvimento"` Rev. Bras. Ginecol. Obstet. 32(6), 2010.

[3] E. Godinho, H. Koch, "Rastreamento do Câncer de Mama: Aspectos Relacionados ao Médico", Radiol. Bras. 37(2), 2004.

[4] L.M. Andrade, "Estudos in vitro de câncer de mama humano da linhagem MDAMB-231 irradiado por Co-60 & Estudos Correlato", (Mestrado: Escola de Engenharia/ Universidade Federal de Minas Gerais), 130p, 2002.

[5] L.H. Gebrim, L.G.A. Quadros, "Rastreamento do câncer de mama no Brasil", Rev. Bras. Ginecol. Obstet. 28(6), 2006.

[6] S. Astley, F. Gilbert, "*Computer-aided Detection in Mammography*", Clinical Radiol. 59:390-399, 2004.

[7] R.B. Medeiros et al., "Influência das condições de processamento no desempenho de dois sistemas tela/filme utilizados na mamografia", Revista Brasileira de Engenharia Biomédica, v. 18, n. 2, p. 57-63, mai/ago 2002.

[8] M.P.S. Silva, "Mineração de Dados – Conceitos, Experimentos e Aplicações com WEKA", Em Escola Regional de Informática RJ/ES, v.1, pp. 19-21, Sociedade Brasileira de Computação, 2004.

[9] T.C.S. Honorio et al., "Atribuição de Autoria com WEKA", IX Encontro de Extensão e X Encontro de Iniciação à Docência (UFPB), João Pessoa, 2007.

[10] D.S.C. Nascimento, "Configuração Heterogênea de Ensembles de Classificadores: Investigação em *Bagging*, *Boosting* e *MultiBoosting*", Dissertação Unifor – Mestrado em Informática Aplicada, 2009.

[11] R.C. Gonzales, R.E. Woods, "*Digital Image Processing*", 2nd ed. Prentice Hall, 2002.

[12] A.M. Puetz, R.C. Oslen, "*Haralick Texture Features Expanded Into The Spectral Domain*", *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*, 2006.

[13] R.M. Haralick, K. Shanmugam, I. Dinstein, "*Textural Features for Image Classification*", *IEEE Transactions on Systems, Man, and Cybernetics SMC-3* (6): 610–621, 1973.

Mapeamento Estatístico Paramétrico das Alterações Estruturais Cerebrais devido à Doença de Alzheimer e ao Transtorno Cognitivo Leve

Michel P. Fernandes[1], João R. Sato[2], Geraldo F. Busatto[3] e Carlos E. Thomaz[1]

[1]FEI, São Bernardo do Campo, SP, Brasil
[2]UFABC, Santo André, SP, Brasil
[3]USP, São Paulo, SP, Brasil

Resumo

This work describes the Statistical Parametric Mapping (SPM) framework to automatically extract and detect relevant statistical information between human brain images. Our aim is to show how SPM may be useful to map brain regions affected by Alzheimer's Disease (AD) and consequently improve its early detection, due to the identification of common biomarkers present in patients diagnosed with Mild Cognitive Impairment (MCI). It is known clinically that the majority of MCI patients tends to progress to AD in the future. Our experiments are based on structural magnetic resonance images provided by the international Alzheimer's Disease Neuroimaging Initiative (ADNI). The results obtained in this work are related to previous clinical findings published by others and highlight the cerebral atrophy regions present on AD, MCI and both.

1. Introdução

O estudo da Doença de Alzheimer (DA) tem se mostrado fundamental nos dias hoje. Além de ser um problema de saúde pública, pode ser também considerado como um problema de ordem econômica devido aos custos elevados associados ao seu tratamento. Por se tratar de uma doença da qual ainda não existe cura, é indispensável que se tenha o diagnóstico cada vez mais apurado dessa desordem cerebral para se dar início ao respectivo tratamento o mais cedo possível.

O caminho escolhido nas pesquisas recentes sobre o tema tem sido o de se utilizar tratamentos específicos para conter o avanço da DA logo no seu estágio inicial [24, 23], possivelmente caracterizado pelo Transtorno Cognitivo Leve (TCL). Apesar da identificação ser difícil tanto no início da demência, ou seja, no TCL, quanto na DA, atualmente existe consenso sobre os principais biomarcadores associados ao surgimento da DA [13, 12, 4, 3, 19], embora ainda a comprovação efetivamente somente seja possível através de exames *post mortem*. No entanto, pode-se acompanhar a evolução destes biomarcadores para ajudar na identificação da doença *in vivo*, cujo interesse é obviamente maior. Neste contexto, a análise de imagens de Ressonância Magnética (RM) estrutural tem sido muito utilizada por contemplar com maior clareza os primeiros sinais da doença devido à quantidade de informação existente neste tipo de dados cerebrais.

Um dos grandes desafios não somente para a identificação da DA como também para o uso de imagens médicas em geral é a transformação do conjunto de informações disponível em conhecimento clínico que poderá dar subsídio aos diagnósticos dos especialistas. Portanto, o uso de ferramentas de computação, que são capazes de analisar estatisticamente dados com dimensionalidade elevada, torna-se cada vez mais frequente e necessário à medida que a qualidade e os detalhes contidos nas imagens aumentam. Nos dias de hoje, a análise de imagens médicas é uma atividade que consome tempo dos especialistas e seu resultado efetivo nem sempre resultará no padrão completo e exclusivo que se deseja extrair. Por exemplo, em uma imagem médica de um paciente com DA objetiva-se fundamentalmente investigar o padrão de atrofia da substância cinzenta do cérebro [20] e, com menos relevancia, as correspondentes regiões associadas ao líquor ou substância branca.

Este trabalho descreve diferenças estatisticamente relevantes entre imagens de RM de DA, TCL e sujeitos sem desordens cerebrais (controles) obtidas através da aplicação do arcabouço computacional denominado Mapeamento Estatístico Paramétrico [6] (MEP). Essas diferenças são representadas no espaço original de imagens para evidenciar as informações relevantes presentes simultaneamente nas amostras de DA e TCL consideradas. As amostras foram obtidas do banco de dados ADNI (*Alzheimer's Disease Neuroimaging Initiative*) [16] que é uma iniciativa patrocinada por inúmeras empresas, em grande parte do setor farmacêutico, que visa coletar o maior número de imagens de RM estrutural em boas condições e de forma padronizada para serem analisadas em pesquisas que poderão levar a no-

vas descobertas para o tratamento dessas doenças. Atualmente, foi constatado que a utilização do MEP em conjunto com a base de dados internacional padronizada ADNI é muito pouco utilizado na comunidade científica nacional o que torna este trabalho um dos primeiros publicados em língua portuguesa neste âmbito. Na comunidade internacional, entretanto, o uso do ADNI com MEP já é maior e com pesquisas voltadas para análise de imagens por tomografia (PET)[29] e imagens por RM estrutural[11, 2, 25] que é o foco deste trabalho.

2. Mapeamento Estatístico Paramétrico

O MEP é um arcabouço computacional de análise de neuroimagens que torna possível a investigação de diferenças locais na anatomia do cérebro humano. Este arcabouço implementa o método de morfometria baseada em voxel [6, 15] que, de forma sucinta, estima a distribuição de probabilidade das intensidades dos voxels para gerar mapas estatísticos paramétricos que possam determinar diferenças relevantes ou testar hipóteses sobre efeitos regionais específicos das amostras de interesse [8].

Desenvolvido na década de 90, pioneiramente por Karl Friston e colegas [6], para análise estatística de neuroimagens funcionais e o seu uso se intensificou na última década com trabalhos acadêmicos voltados para diversas áreas clínicas e com uso não exclusivamente em exames de imagem funcional, mas também de imagem estrutural [1, 10, 27, 20].

Para a geração dos mapas paramétricos, é preciso que as imagens por RM sejam antes pré-processadas. Este pré-processamento envolve algumas etapas que tem por objetivo identificar e separar a região de interesse, torná-la comparável para amostras diferentes, e suficientemente suavizada para atender na prática a suposição estatística de que a dispersão dos voxels segue uma distribuição Gaussiana.

Mais especificamente, as etapas de pré-processamento do MEP são: (1) normalização; (2) segmentação; e (3) suavização. Na primeira etapa de pré-processamento, realiza-se a normalização espacial das neuroimagens com o intuito de corrigir possíveis erros de aquisição e permitir o registro (alinhamento) de imagens de pessoas diferentes através da estimação de parâmetros de deformação usando um espaço anatômico de referencia, como o atlas de Talairach e Tournoux [28]. Essencialmente, a normalização espacial procura minimizar a soma da diferença dos quadrados entre as características da imagem de interesse e o atlas. Este processo pode usar uma variedade de modelos para o alinhamento, dentre eles: (i) transformação afim de 12 parâmetros, que constituem uma matriz de transformação espacial (translação, rotação, escala e cisalhamento) nas direções (x, y, z), (ii) funções de baixa freqüência de base espacial (geralmente um conjunto cosseno ou de polinômios), e (iii) campo vetorial de deformação que especifica o mapeamento para cada voxel.

A segmentação da imagem bruta em substância branca, substância cinzenta e líquor serve para separar a região de interesse de acordo com a relevância clínica de cada uma dessas substâncias em um determinado estudo. Por exemplo, para o estudo da DA, os efeitos são mais evidentes e os biomarcadores mais predominantes na substância cinzenta do que nas outras regiões. A diferenciação das substâncias cerebrais é realizada através da identificação de níveis de intensidade de voxel que, posteriormente, são separados em classes específicas para cada região, de acordo com a seguinte equação:

$$P(y_i|c_i = k, \mu_k, \sigma_k) = \frac{1}{(2\pi\sigma_k^2)} exp\left(- \frac{(y_i - \mu_k)^2}{2\sigma_k^2} \right), \quad (1)$$

onde y_i é a intensidade do voxel i, e μ_k e σ_k^2 são respectivamente a média e a variância da classe k. Os parâmetros μ_k e σ_k^2 são estimados a partir de mapas de probabilidade *a priori* de tecidos sobrepostos e ilustrados na Figura 1. Esses mapas de probabilidade de tecido foram construídos pelo *International Consortium for Brain Mapping*, utilizando-se imagens por RM de cérebros adultos sem desordens cerebrais, e representam tanto uma média de intensidades quanto de posicionamentos espaciais dos voxels cerebrais.

Na última etapa de pré-processamento, a suavização tem o papel de reduzir variações devido a diferenças individuais na anatomia dos tecidos da substância branca, substância cinzenta e líquor, através da aplicação de um filtro Gaussiano. Fundamentalmente, esse processo permite que a intensidade de cada voxel da imagem seja dada pela média ponderada dos valores dos voxels adjacentes, suavizando-se as bordas correspondentes [15]. Adicionalmente, essa etapa realiza a padronização estatística dos dados para que os erros associados a cada imagem sejam normalmente distribuídos. Portanto, esta etapa tende a aumentar a eficiência do teste estatístico das diferenças, suposto teoricamente como paramétrico [7].

Para geração dos mapas estatísticos, utiliza-se comumente o teste t de *Student* para calcular a probabilidade de haver diferenças estatisticamente significantes entre os grupos de imagens pré-processadas [21]. O t-valor de cada voxel (t_k) é definido como a diferença aritmética entre as médias de cada grupo de amostras ponderada pelo desvio padrão do espalhamento das amostras, ou seja:

$$t_i = \frac{\overline{x}_{1,i} - \overline{x}_{2,i}}{\sigma_i \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}, \quad (2)$$

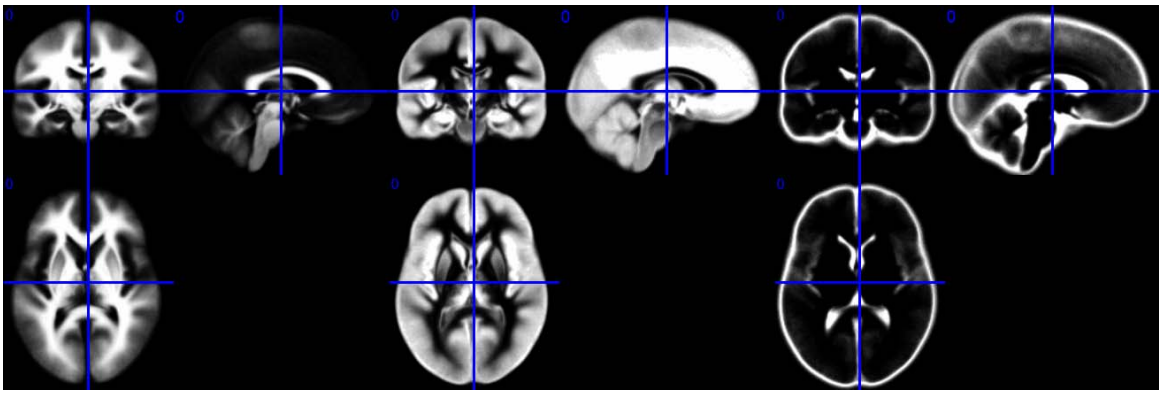


Figura 1. Mapas de probabilidades *a priori*: substância branca, substância cinzenta e líquor.

onde $\overline{x}_{1,i}$ e $\overline{x}_{2,i}$ são respectivamente as médias do voxel i para os grupos 1 e 2 de amostras, σ_i é o desvio padrão ponderado de todas as amostras para o voxel i, e N_1 e N_2 o número total de amostras do grupo 1 e do grupo 2. O desvio padrão ponderado do conjunto de amostras é definido como:

$$\sigma_i = \sqrt{\frac{(N_1 - 1)(\sigma_{1,i})^2 + (N_2 - 1)(\sigma_{2,i})^2}{N_1 + N_2 - 2}}, \quad (3)$$

onde $\sigma_{1,i}$ e $\sigma_{2,i}$ são respectivamente o desvio padrão do voxel i para os grupos 1 e 2.

Calculado os t-valores dos voxels para os grupos de amostras analisados, pode-se utilizar o conceito de hipótese nula para definir as regiões do cérebro que apresentam diferenças significativas. Para que a diferença seja significativa, deve-se rejeitar a hipótese nula (H_0) e aceitar a hipótese alternativa (H_1), descritas como:

$$H_0 : \overline{\mu}_1 = \overline{\mu}_2, \quad (4)$$

$$H_1 : \overline{\mu}_1 \neq \overline{\mu}_2. \quad (5)$$

De acordo com a teoria do teste de hipóteses, dois tipos de erro podem existir nessa análise: Erro do tipo I, isto é, rejeitar H_0 quando esta é verdadeira, ou seja, afirmar que existe diferença estatística significativa quando ela não existe; Erro do tipo II, ou seja, não rejeitar H_0 quando esta é falsa, ou seja, afirmar que não existe diferença estatística significativa quando ela existe [14]. Normalmente, determina-se um nível de significância p que representa na prática uma taxa de erro aceitável dado o grau de liberdade do estudo. Usualmente, o valor de p é fixado em 5%, 1% ou 0.1%. Assim, determinada arbitrariamente a probabilidade p de se cometer o erro de tipo I e utilizando a diferença entre a quantidade total de amostras e a quantidade de grupos como grau de liberdade, calcula-se o valor da estatística do teste para cada voxel i. Se o valor absoluto da estatística calculado com os dados das amostras for maior que o valor teórico, então considera-se que a diferença encontrada no voxel i é estatisticamente relevante.

3. Base de Dados ADNI

A base de dados ADNI [16] contempla imagens por RM estrutural e informações clínicas adicionais como, por exemplo, sexo, idade, pontuações em testes cognitivos, dentre outros, para três grupos distintos e previamente conhecidos ou diagnosticados: DA, TCL e controles. O projeto é hospedado no acervo de imagens do LONI (*Laboratory of Neuro Imaging*) IDA (*Image Data Archive*) da Universidade da Califórnia[1].

Os participantes do ADNI foram recrutados em 50 localidades dos Estados Unidos e do Canadá, sendo 233 controles cognitivamente saudáveis em idade avançada seguidos por 3 anos, 410 indivíduos com TCL amnésico seguidos por 3 anos e 200 indivíduos com DA inicial seguidos por 2 anos. Os critérios gerais de inclusão no ADNI foram: indivíduos entre 55-90 anos de idade, sem histórico de depressão ou outra condição psiquiátrica, ausência de medicações psicoativas, pontuação 4 ou menor no teste de *Hachinski* [22] modificado, e ter parceiro fixo capaz de prover informações sobre funcionalidade em atividades diárias. Para os controles saudáveis foram pré-requisitos: pontuação entre 24 e 30 no teste *Mini-Mental* e pontuação de 0 no *Clinical Dementia Rating* (CDR). Para pacientes com TCL foram pré-requisitos: queixas de problemas de memória, pontuação entre 24 e 30 no teste *Mini-Mental*, perda de memória medida com o teste *Logical Memory* II, CDR de 0.5 e sem prejuízos funcionais nas atividades diárias. Para pacientes com DA, foram pré-requisitos: estar de acordo com os critérios de provável diagnóstico do *National Institute of Neurological Disorders and Stroke e Alzheimer's Disease and Re-*

1 `http://ida.loni.ucla.edu`

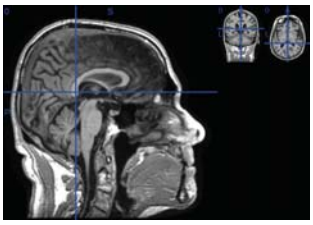


Figura 2. Imagem de RM estrutural disponibilizada pelo ADNI (sem pré-processamento).

lated Disorders Association, pontuação entre 20 e 26 no teste *Mini-Mental* e pontuação de 0.5 ou 1 no CDR.

As imagens por RM foram coletadas em equipamentos de fabricantes diferentes. Por isso, durante a fase de planejamento do ADNI, desenvolveu-se protocolos específicos para cada fabricante com o objetivo de maximizar a utilidade científica das imagens coletadas. Também foi realizado monitoramento e calibração dos equipamentos e das imagens através de correções geométricas baseadas em objetos de testes (*phantom* antropométrico) [16, 18] para assegurar a consistência das imagens adquiridas em múltiplos locais. Os protocolos foram elaborados tomando como referencia sequências 3D T1 para análises morfométricas. As imagens estão disponíveis em resoluções de 1.5T e 3.0T, e podem ser lidas no formato *Analyze* e no formato *NIFTI* (*Neuroimaging Informatics Technology Initiative*), mais atual.

4. Experimentos

Todos os experimentos deste trabalho foram realizados utilizando-se o *software* SPM5 [9] e o método de morfometria baseado em intensidade do voxel. As imagens por RM estrutural dos 3 grupos de amostras considerados (DA, TCL e controles) são do tipo T1, adquiridas em *scanners* de 1.5T e salvas no formato NIFTI. A Figura 2 ilustra uma imagem do ADNI sem pré-processamento utilizada neste trabalho.

Como o número total de imagens por RM estrutural por grupo é distinto, utilizou-se: 200 imagens para DA, 410 para TCL e 233 para controles. Para o conjunto de amostras de DA, 49% são mulheres e 52% homens; para o TCL, 36% são mulheres e 64% são homens, e para o grupo de controles, 48% são mulheres e 52% são homens. Adicionalmente, para os pacientes classificados como DA, mais da metade possui o teste CDR igual 1 e os demais com CDR 0.5. Para pacientes com TCL, a maioria (99%) possui o resultado do CDR com valor de 0.5. Todos os controles possuem CDR igual 0.0.

Nas etapas de pré-processamento do MEP, optou-se pela segmentação modulada para preservar as mesmas regiões cerebrais apesar das possíveis diferenças

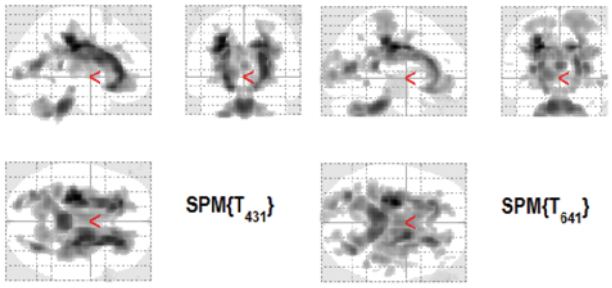


Figura 3. *Glass Brain*: Mapa de contrastes das diferenças estatísticas mais relevantes entre DA x controles (esquerda) e TCL x controles (direita).

devidos aos tamanhos das imagens, e utilizou-se o filtro Gaussiano de 8mm padrão para suavização das imagens, seguindo as heurísticas sugeridas por outros trabalhos da área [1]. Para validação das diferenças estatísticas encontradas entre as amostras de DA x controles e TCL x controles, fixou-se $p = 0.05$ e realizou-se o teste t de *Student* pareado para os correspondentes graus de liberdade de cada comparação.

5. Resultados

A Figura 3 ilustra as diferenças estatísticas entre os grupos analisados dois-a-dois, ou seja, entre DA e controles e entre TCL e controles. Este mapa de contrastes (*Glass Brain*) está projetado sobre um molde de cérebro padronizado e permite visualizar as diferenças estatísticas mais relevantes entre os grupos analisados.

Para melhor compreensão dos resultados e das regiões cerebrais demarcadas como coincidentes ou não, sobrepos-se o mapa de contrastes de cada análise em uma imagem de RM estrutural de um cérebro de controle, através do *software* MRIcro [26]. A Figura 4 mostra esses resultados. As diferenças entre DA e controles são representadas pela cor vermelha. Para as diferenças entre TCL e controles, utilizou-se a cor azul e as características comuns entre as duas análises são representadas pela mistura das cores vermelha e azul, representadas em tons de rosa.

Pode-se observar pela Figura 4 [2] que o processo de degeneração cerebral é mais evidente em DA do que em TCL, como já se esperava. Constata-se também que esse processo de degeneração em DA se manifesta fundamentalmente pela atrofia nas regiões do hipocampo, lobos tem-

2 Para melhor visualização das diferenças relevantes em formato colorido favor acessar a versão eletrônica do artigo

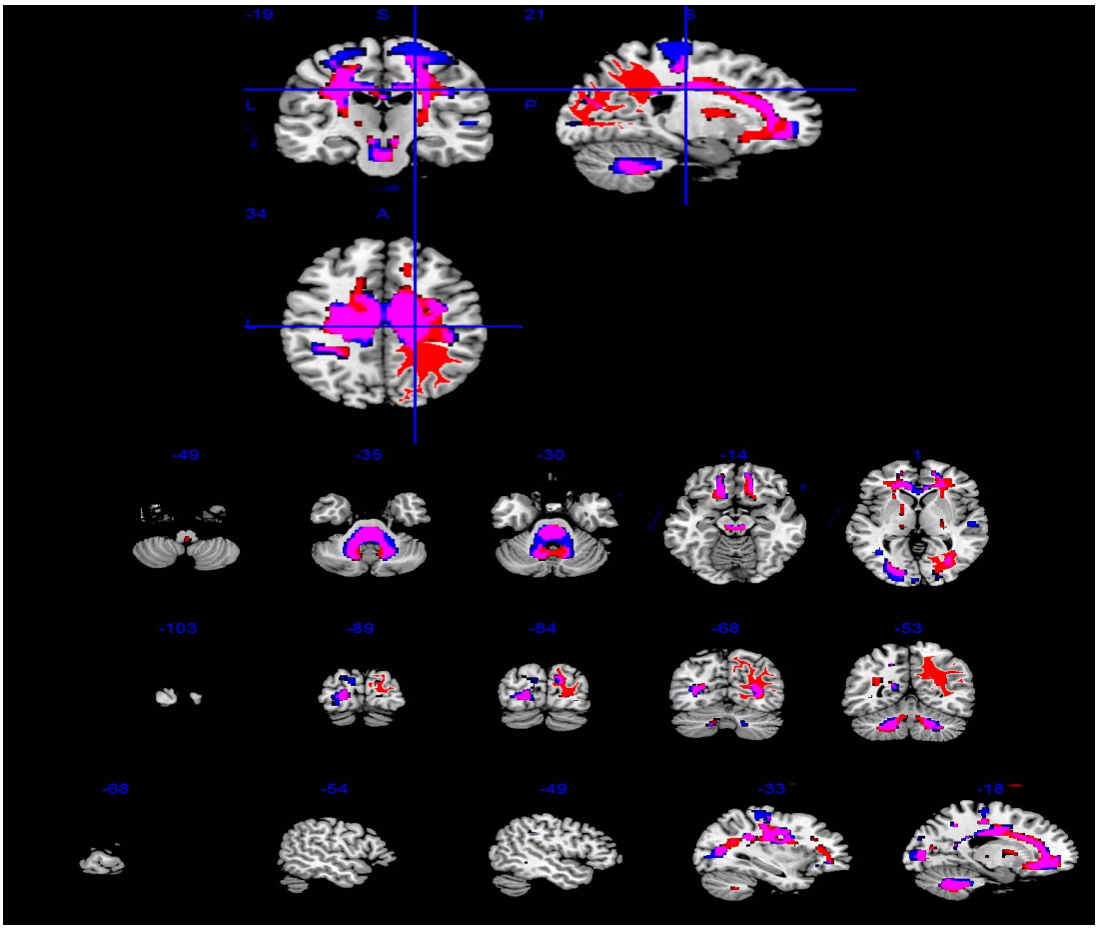


Figura 4. Diferenças relevantes devido a DA (cor vermelha), TCL (azul) e ambos (tons de rosa).

porais, corpo caloso, cerebelo e sistema ventricular. Algumas dessas regiões também são afetadas pelo TCL, porém em menor volume, indicando talvez um estágio inicial da DA se seguirem a mesma tendência de degeneração em exames consecutivos. Esses resultados de DA não são inéditos e já foram observados por outros trabalhos na literatura afim [17, 5].

6. Conclusão

Este trabalho descreveu o arcabouço computacional denominado mapeamento estatístico paramétrico utilizado para detectar e extrair automaticamente as diferenças cerebrais relevantes devido à DA e ao TCL em imagens por RM estrutural. Todas as imagens foram disponibilizadas pelo projeto ADNI. Os resultados mostram que o processo de degeneração do cérebro humano é característico em pacientes com DA e que há regiões coincidentes de atrofia cerebral entre DA e TCL. Embora tais conclusões tenham sido obtidas a partir de comparações de médias entre grupos de pacientes e controles, sem uma análise mais específica para cada indivíduo, a identificação de marcadores comuns nos pacientes de DA e TCL pode ajudar a prever precocemente e acompanhar singularmente a evolução da DA em cada paciente. Trabalhos futuros terão como objetivo quantificar as diferenças estatísticas encontradas, como por região cerebral, indicando o volume de atrofia encontrada em cada grupo de pacientes estudados. De forma adicional será utilizado além do método de diferenciação estatística univariada que foi utilizado neste trabalho, o método multivariado e a comparação entre ambos os métodos.

Agradecimentos

Os dados utilizados na elaboração deste artigo foram obtidos do banco de dados ADNI. Como tal, os pesquisadores do ADNI forneceram os dados, mas não participaram da análise e elaboração deste relatório.

Referências

[1] J. Baron, G. Chételat, B. Desgranges, G. Perchey, B. Landeau, V. d. l. Sayette, and F. Eustache. In vivo mapping of gray matter loss with voxel-based morphometry in mild alzheimer's disease. *NeuroImage*, 2001.

[2] C. Davatzikos, P. Bhatt, L. Shaw, K. Batmanghelich, and J. Q. Trojanowski. Prediction of mci to ad conversion, via mri, csf biomarkers and pattern classification. *Neurobiology of Aging*, 2010.

[3] S. M. de la Monte, L. Volicer, S. L. Hauser, and J. R. . Wands. Increased levels of neuronal thread protein in cerebrospinal fluid of patients with alzheimer's disease. *Annals of Neurology*, page 733–42, 1992.

[4] S. M. de la Monte and J. Wands. Neuronal thread protein overexpression in brains with alzheimer's disease lesions. *Journal of the Neurological Sciences*, page 153–64, 1992.

[5] N. C. Fox, E. K. Warrington, and P. A. Freeborough. Presymptomatic hippocampal atrophy in alzheimer's disease: A longitudinal study. *Brain*, page 2001–7, 1996.

[6] K. Friston, J. Ashburner, S. Kebel, T. Nichols, and P. William. *Statiscal Parametric Mapping The Analysis of Functional Brain Images*. Academic Press, 2007.

[7] K. J. Friston. Spatial registration and normalization of images, 2004.

[8] K. J. Friston, C. D. Frith, P. F. Liddle, and F. R. J. Comparing functional (pet) images: the assessment of signi?cant change. *Journal of Cerebral Blood Flow & Metabolism*, 11:690–99, 1991.

[9] K. J. Friston, A. P. Holmes, J. B. Poline, P. J. Grasby, S. C. Williams, R. S. Frackowiak, and R. Turner. Analysis of fmri time-series revisited. *NeuroImage*, pages 9675–85, 1995.

[10] K. Fritzsche, R. D. Wangenheim, Aldo von and, and R. Unterhinninghofen. Automated mri-based quantification of the cerebral atrophy providing diagnostic information on mild cognitive impairment and alzheimer's disease. *19th IEEE Symposium on Computer-Based Medical Systems (CBMS)*, 2006.

[11] S. Gabriela, N. Eini, M. Stuart, S. Oded, C. Kewei, R. E. M., J. A. M., H. Merja, T. Susanna, W. Lars-Olof, V. Ritva, K. Miia, and S. Hilkka. Whole brain atrophy rate predicts progression from mci to alzheimer's disease. *Neurobiol Aging*, 2008.

[12] D. Galasko, L. Chang, and R. Motter. High cerebrospinal fluid tau and low amyloid beta42 levels in the clinical diagnosis of alzheimer disease and relation to apolipoprotein e genotype. *Archives of Neuroloy*, page 937–45, 1998.

[13] D. Galasko, C. Clark, and L. Chang. Assessment of csf levels of tau protein in mildly demented patients with alzheimer's disea. *Neurology*, pages 632–5, 1997.

[14] L. H. Girardi and L. Cargnelutti, A. F.and Storck. Erro tipo i e poder de cinco testes de comparação múltipla de médias. *Revista Brasileira de Biometria*, page 23–36, 2009.

[15] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. A. Henson, K. J. Friston, and R. S. J. Frackwiak. A voxel-based morphometry study of ageing in 465 normal adult human brains. *NeuroImage*, page 21 – 36, 2001.

[16] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner. The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, pages 685–91, 2008.

[17] C. R. J. Jack, R. C. Petersen, and Y. C. Xu. Medial temporal atrophy on mri in normal aging and very mild alzheimer's disease. *Neurology*, page 786–94, 1997.

[18] J. Jovicich, S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. Macfall, B. Fischl, and A. Dale. Reliability in multi-site structural mri studies: effects of gradient non-linearity correction on phantom and human data. *NeuroImage*, pages 436–43, 2006.

[19] P. J. Kahle, M. Jakowec, and S. J. Teipel. Combined assessment of tau and neuronal thread protein in alzheimer's disease csf. *Neurology*, page 1498–504, 2000.

[20] S. Kloppel, C. Stonnington, C. Chu, B. Draganski, R. Scahill, J. Rohrer, N. Fox, C. Jack, J. Ashburner1, and R. Frackowiak. Automatic classification of mr scans in alzheimer's disease. *Brain*, 2008.

[21] R. D. Leão, J. R. Sato, and C. E. Thomaz. Extração multilinear de informações discriminantes em imagens de ressonância magnética do cérebro humano. *Proceedings of the 10th Workshop de Informática Médica WIM'10*, page 1740–49, 2010.

[22] C. Q. Mountjoy. Dementia and the hachinski scale. *British Journal of Hospital Medicin*, 39, 1988.

[23] R. C. Petersen and J. C. Morris. *Mild Cognitive Impairment: Aging to Alzheimer's Disease*. Oxford University Press, Inc., 2003.

[24] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen. Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neuroloy*, 56:303–8, 1999.

[25] S. Risacher, A. Saykin, J. West, L. Shen, H. Firpi, and B. McDonald. Baseline mri predictors of conversion from mci to probable ad in adni cohort. *Current Alzheimer Research*, 6, 2009.

[26] C. Rorden and M. Brett. Stereotaxic display of brain lesions. *Behavioural Neurology*, 12:191–200, 2000.

[27] Y. Sun, S. Lv, and Y. Tang. Construction and application of bayesian network in early diagnosis of alzheimer. *International Conference on Complex Medical Engineering*, 2007.

[28] P. Talairach and J. Tournoux. *A stereotactic coplanar atlas of the human brain.* Thieme Medical Publishers, 1988.

[29] I. Yakushev, A. Hammers, A. Fellgiebel, I. Schmidtmann, A. Scheurich, H. Buchholz, J. Peters, P. Bartenstein, K. Lieb, and M. Schreckenberger. Spm-based count normalization provides excellent discrimination of mild alzheimer's disease and amnestic mild cognitive impairment from healthy aging. *Neu*, 44, 2009.

Classificação Automática de Grupos de Espécies Florestais

J. G. Martins
Universidade Tecnológica Federal do Paraná
Rua Cristo Rei, 19
85902-490 - Toledo - PR - Brasil
martins@utfpr.edu.br

L. E. S. Oliveira*
Universidade Federal do Paraná
Rua Cel. Francisco H. dos Santos, 100
81531-980 - Curitiba - PR - Brasil
lesoliveira@inf.ufpr.br

P. L. Paula Filho
Universidade Tecnológica Federal do Paraná
Av. Brasil, 4232
85884-000 - Medianeira - PR - Brasil
plpf2004@gmail.com

S. Nisgoski
Universidade Federal do Paraná
Av. Pref. Lothário Meissner, 900
80210-170 - Curitiba - PR - Brasil
silnis@yahoo.com

Abstract

Forest species can be taxonomically divided into Angiosperms and Gymnosperms, but human classification can introduce errors because of this is a monotonous and stressful task. In this paper we introduce a structural features set composed by only six elements. It is capable of capturing the structure of the main image components using simple statistical measures and can be useful for an automatic forest species classification system. A series of comprehensive experiments demonstrate that the proposed feature set surpasses the traditional texture features in about 10 percentage points.

1. Introdução

Recentemente, com o advento da globalização e a redução da disponibilidade de matéria-prima, a comercialização de madeira se destaca como uma importante área de negócios. Compradores devem se certificar de que estão comprando o material correto enquanto agências supervisoras precisam garantir que não haja madeira extraída irregularmente das florestas. Tal questão envolve grandes somas nas transações comerciais realizadas e a garantia quanto à inexistência de fraudes é um importante requisito. Tais fraudes podem decorrer de misturas de diferentes tipos de madeira, das mais nobres às menos valorizadas, e da tentativa de extração e exportação de madeira proveniente de espécies sujeitas à extinção.

Neste contexto, sistemas dotados de visão computacional tornam-se muito interessantes a esta área comercial devido a uma série de vantagens providas por estes quando comparados com o processo realizado por especialistas humanos. No entanto, na última década a maioria das aplicações de visão computacional aplicada a madeira se refere a controle de qualidade, avaliação e detecção de defeitos [4, 10]. Apenas recentemente, alguns autores começaram a usar visão computacional para classificar espécies florestais. Tou, Lau e Tay [11] reportaram dois experimentos inerentes à classificação de espécies florestais que empregaram características baseadas em Matriz de Co-ocorrência de Níveis de Cinza (*Gray-Level Co-occurrence Matrix* - GLCM) para treinar um classificador baseado em rede neural artificial. Os autores empregaram imagens inteiras obtidas da base *Centre for Artificial Intelligence and Robotics* (CAIRO) e reportam taxas de reconhecimento entre 60% e 72% para cinco diferentes espécies florestais.

Khalid *et al* [5] apresentaram um sistema para reconhecer vinte diferentes espécies florestais da Malásia. Uma particularidade do método consiste do fato de que as amostras de madeira foram inicialmente cozidas e então cortadas com micrôtomos em finas seções. A aquisição das imagens foi então realizada com a utilização de uma câmera industrial de alta performance e com iluminação controlada. O processo de reconhecimento está baseado em rede neural artificial treinada com informações extraídas com GLCM. A base de dados utilizada nos experimentos contém 1.753 imagens para treinamento e somente 196 para teste. Os autores reportam taxas de reconhecimento de 95%, embora

* Esta pesquisa é auxiliada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), #306358/2008-5.

um problema destes experimentos consiste no fato de terem utilizado 90% das amostras para treinamento.

Paula *et al* [7] investigaram o uso de características extraídas com GLCM e cores para reconhecer vinte e duas diferentes espécies da flora brasileira. Os experimentos consideraram as imagens inteiras e sua divisão, com a melhor alternativa para vinte e cinco fragmentos. Foram empregadas dezoito características de cor obtidas de alguns dos canais dos modelos RGB, HSV e LUV e vinte e quatro extraídas a partir das GLCM (energia, contraste, entropia, homogeneidade, probabilidade máxima e momento de terceira ordem para as quatro direções e distância de um pixel), num total de quarente e duas. O classificador empregado foi *Multi-Layer Perceptron* (MLP) e os índices de reconhecimento foram 65% para as imagens inteiras e 82% quando estas eram divididas e se realizava votação dentre os vinte e cinco fragmentos para definir a classe à qual a instância será atribuída. Os autores afirmam que a estratégia de segmentação permitiu lidar melhor com a grande variabilidade interna das classes e que a combinação de características baseadas em cores e extraídas das GLCM gerou resultados consideravelmente melhorados.

A identificação de madeiras que já sofreram algum tipo de processamento, ou mesmo de troncos, geralmente precisa ser realizada fora de seu contexto de origem (a floresta). Esta não é uma tarefa fácil visto que não se pode contar com flores, frutos e folhas, as quais auxiliariam muito neste processo. Diante disto, a identificação precisa ser realizada por especialistas bem treinados, mas poucos alcançam boa precisão na classificação, o processo para se realizar os treinamentos é demorado e não há especialistas suficientes para atender a demanda da indústria. Um outro fator a ser considerado consiste no fato de que o processo de identificação manual toma mais tempo e obtém taxas de reconhecimento menores que o executado por máquinas. Isso devido à dispersão da atenção dos especialistas humanos diante de um processo repetitivo que pode resultar em taxas de erros que inviabilizem verificações de cargas para exportação, por exemplo.

Um modo de simplificar a tarefa de classificação de espécies florestais consiste na identificação de suas famílias ou classes taxonômicas. Neste contexto, espécies florestais podem ser categorizadas em Angiospermas e Gimnospermas. Angiospermas são as plantas que apresentam formas mais sofisticadas e se adaptaram para subreviver em um conjunto mais diversificado de ambientes e condições climáticas. Este grupo é constituído por plantas ornamentais que apresentam flores e por todos os vegetais que produzem frutos comestíveis. Além disso, os exemplares pertencentes a este grupo também compreendem espécies com valor comercial, são fonte de madeira para todo o mundo e são vulgarmente denominadas *hardwood*. Gimnospermas, de outro modo, compreendem um conjunto de exemplares considerados primitivos em razão de sua anatomia e o menor número de características, dificultando inclusive a diferenciação entre suas espécies. Também denominados *softwood* as espécies que compõem este grupo possuem menor valor comercial devido às propriedades da madeira e a suas possibilidades de aplicação. Neste grupo, a produção de madeira provém de um subconjunto denominado coníferas, dentre as quais tem-se cedros, sequóias, pinheiros e pinos. Além da produção de madeira, algumas espécies do grupo *softwoods* são empregadas para a fabricação de papel e compensados em geral, além de também serem fontes de matéria-prima para a produção de essências e resinas.

Considerando as diferenças quanto a aplicações e valores comerciais destes dois grupos, no presente trabalho é proposto um conjunto compacto de características estruturais composto por cinco características para classificar espécies florestais em Angiospermas ou Gimnospermas. A identificação deste conjunto de características explorou o fato de que Gimnospermas apresentam células mais definidas que, em geral, resultam em um número maior de pequenos componentes conexos quando são segmentadas nas imagens. Para extrair tais características empregou-se um *threshold* adaptivo seguido por operadores morfológicos binários. Considerando o fato de que este problema pode ser formalizado com um problema a ser tratado por meio de técnicas que lidem com textura, o trabalho também apresenta resultados provenientes de experimentos comparativos entre o conjunto de características proposto e algumas das características extraídas com GLCM propostas por Haralick. Foi empregado um conjunto de diferentes classificadores e as taxas de reconhecimento diferem em cerca de 10%, com vantagem para o conjunto de características proposto.

2. Base de Dados

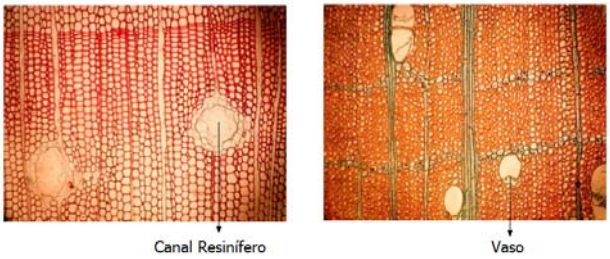
A base de dados empregada neste trabalho é composta por seiscentas images de quinze diferentes espécies florestais adquiridas no Laboratório de Anatomia da Madeira da Universidade Federal do Paraná (UFPR). Estas espécies estão relacionadas na Tabela 1, sendo cinco espécies pertencentes ao grupo Gimnospermas e dez ao grupo Angiospermas, cada uma com quarenta imagens e totalizando duzentas e quatrocentas imagens, respectivamente.

Os procedimentos adotados no processo de coleta das imagens compreenderam seis passos. O primeiro consistiu no cozimento da madeira, por tempo variado de acordo com a espécie florestal, para seu amolescimento. Em seguida foram realizados os cortes das lâminas de madeira com um micrótomo de deslizamento, cada uma com espessura de aproximadamente vinte e cinco micras[1]. No terceiro passo

Gimnospermas	Angiospermas
Araucaria Angustifolia	*Dinizia excelsa*
Pinus elliottii	*Cedrela sp*
Pinus maximinoi	*Dalbergia nigra*
Pinus taeda	*Eucalyptus grandis*
Podocarpus lambertii	*Grevillea robusta*
	Tabebuia sp
	Hymenaea sp
	Nectandra sp
	Porcelia macrocarpa
	Swietenia macrophylla

Tabela 1. Espécies que compõem a Base de Dados

realizou-se a coloração das lâminas de madeira pelo processo de tripla coloração empregando as substâncias acridina vermelha, crisoidina e azul de astra. O quarto passo realizado foi a desidratação em série alcoólica ascendente. No quinto passo montou-se a lâminas para observação, com a fixação das lâminas de madeira em lâminas de vidro. Por fim, realizou-se a coleta das imagens com o auxílio de um microscópio da marca Olympus e modelo CX40, produzindo imagens com cem vezes de aproximação e resolução de 1024×768 pixels. A Figura 1 apresenta algumas amostras da base de dados utilizada.



(a) (b)

Figura 1. Amostras da base de dados (a) Gimnospermas e (b) Angiospermas

Por fim, cabe destacar que características de cor não podem ser empregadas no processo de reconhecimento para esta base de dados devido ao uso dos corantes para melhorar o contraste no microscópio.

1 Uma micra equivale à milhonésima parte do metro ou 1×10^{-6} metro.

3. Características

Como afirmado anteriormente, poucos trabalhos relatados na literatura têm focado o problema de reconhecimento de espécies florestais, não tendo sido encontrados relatos quanto à classificação automática de grupos de espécies florestais. Os trabalhos inerentes a esta área de aplicação focam classificação de espécies florestais [5, 7, 11], considerando geralmente a formulação do problema focada no reconhecimento de textura e, consequentemente, características extraídas com GLCM. Neste trabalho é apresentado um conjunto de características estruturais que podem ser aplicadas para classificar grupos de espécies florestais (seção 3.1). Considerando a aplicação características extraídas com GLCM e a comparação de seus resultados com os obtidos pelo conjunto de características proposto, também se apresenta um breve descritivo quanto à GLCM (seção 3.2).

3.1. Conjunto de Características Estruturais

Uma característica muito frequentemente encontrada nas espécies pertencentes ao grupo Angiospermas é o elemento denominado vaso (Figura 1b), o qual constitui a maior parte do sistema de transporte de água destas plantas. De forma contrária, as espécies que compõem o grupo Gimnospermas não possuem vasos, mas comumente apresentam canais resiníferos (Figura 1a), os quais são muito similares aos vasos. Assim, esta característica por si só não discrimina ambos os grupos. Um outro aspecto das espécies pertencentes a Angiospermas é que elas possuem estruturas mais densas (células menores) que as pertencentes a Gimnospermas. Após alguns experimentos observou-se que depois da segmentação das imagens binárias de Angiospermas haviam grandes componentes conexos. Já para Gimnospermas, as imagens apresentavam grande quantidade de pequenos componentes conexos. Diante disto, a proposta para o conjunto de características para descrição das imagens consiste em detectar estes componentes conexos e descrever suas estruturas usando medidas estatísticas simples. Por esta razão, este conjunto de características foi denominado estrutural.

O primeiro passo para extrair o conjunto de características proposto consiste em binarizar a imagem. Para isso, foi utilizado *threshold* adaptivo [1] sem sobreposição de janelas. Esta fase de aplicação do *threshold* adaptivo mostrou-se importante porque permitiu a suavização dos cantos escuros decorrentes do processo de aquisição das imagens, como mostrado na Figura 2a. O algortimo aplicado para identificação deste *threshold* foi o de Otsu, um dos mais conhecidos para este propósito [6]. Foram testadas três diferentes dimensões para as janelas (pequena $[8\times8]$, média $[50\times50]$ e grande $[100\times100]$) e to-

das elas apresentaram resulados semelhantes, conforme apresentado na seção 4.

A partir deste ponto, a imagem binária foi invertida para que cada célula, vaso e canal resinífero formassem os componentes conexos. Para manter os principais componentes conexos e separá-los durante o processo de limiarização, foram aplicadas operações de erosão, um operador básico da morfologia matemática [9]. O elemento estruturante aplicado nas operações de erosão foi o quadrado, com dimensões 3×3 e número de iterações determinado empiricamente. Na seção 4 são discutidos os impactos desta operação e do número de iterações nas taxas de reconhecimento.

A Figura 2 exemplifica o processo anteriormente descrito. Como se pode observar na Figura 2c um grande número de componentes foram removidos pelas operações de erosão. Pode-se também perceber que as Angiospermas perdem mais componentes que as Gimnospermas durante este processo. Neste contexto, o número e tamanho dos componentes conexos provê informações para compor o conjunto de características que permite discriminar ambos os grupos. Especificamente neste ponto encontra-se o diferencial da proposta e o que a torna robusta para diferenciar os grupos de espécies florestais previamente identificados.

O próximo passo consiste em computar as características para os componentes conexos resultantes ao processo descrito. O vetor de características final é composto por cinco elementos: número de componentes conexos, tamanho médio dos componentes conexos, variância, curtose e obliquidade. Após computados, os valores obtidos para as características foram normalizados com a regra Min-Max, garantindo que as diferentes amplitudes de valores de cada característica não interferissem em sua representatividade.

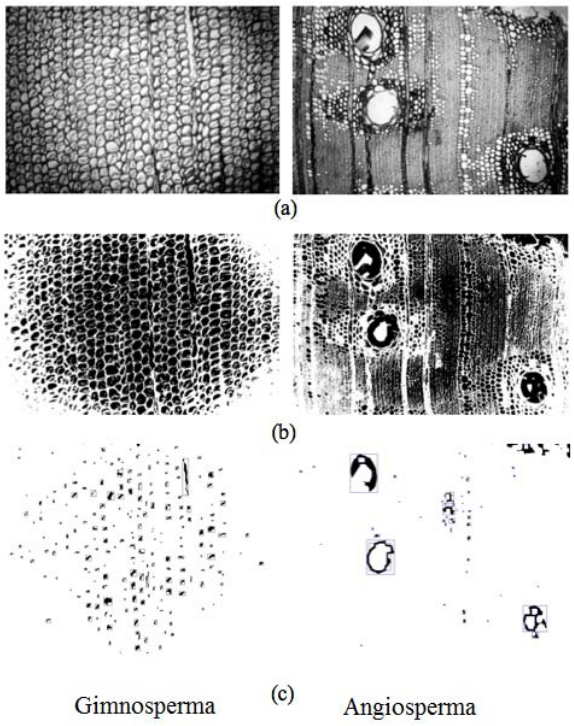


Figura 2. Extração de Características. (a) Imagens em Níveis de Cinza, (b) Imagens Binárias e (c) Componentes conexos depois da erosão

3.2. Matriz de Co-Ocorrência de Níveis de Cinza

Há vários métodos que exploram a distribuição espacial e a dependência ou correlação entre os níveis de cinza dos pixels de uma determinada área. Dentre eles, um dos mais conhecidos consiste na representação de textura baseada em Matriz de Co-ocorrência de Níveis de Cinza (*Gray Level Co-occurrence Matrix* - GLCM). Esta baseia-se nas repetições das ocorrências de padrões quanto à variação dos níveis de cinza dos pixels de uma imagem de textura e provê medidas quanto a propriedades como rugosidade, suavidade e regularidade apresentadas na imagem [12].

Sendo uma das mais usadas e com resultados muito bons em várias áreas de aplicação, formalmente, Haralick fundamenta GLCM da seguinte forma: dada uma imagem I com dimensões $N_r\times N_c$, tal que $L_r = \{1..N_r\}$ e $L_c = \{1..N_c\}$ representam os possíveis valores para linhas e colunas, respectivamente; e $L = \{1..N_g\}$ o conjunto de N_g níveis de cinza quantizados de I. A imagem I pode ser representada como uma função que associa algum nível de cinza de G a uma célula ou par de coordenadas de $L_r\times L_c$. Ou seja, $I : L_r\times L_c\rightarrow G$ [2, 3, 8].

Precursor desta técnica, Haralick definiu o uso de uma matriz $N_g\times N_g$ para representar a probabilidade P_{ij}, tal que um pixel p_i possua nível de cinza n_i e um pixel vizinho a p_i (p_j - a uma distância *d* e um ângulo α) possua nível de cinza n_j. A partir disto, os valores de uma GLCM podem empregados diretamente para formar o vetor de características ou permitir a extraição de outras informações estatísticas secundárias, dentre as quais há grande destaque para o conjunto de 14 características propostas pelo próprio Haralick [2, 3].

Embora a proposta inicial tenha definido quatorze descritores, a maioria dos trabalhos encontrados na literatura consideram difrentes subconjuntos destes e afirmam haver correlações ou redundâncias entre as informações provenientes de alguns deles. Assim, também neste trabalho optou-se por considerar o seguinte subconjunto com seis características: Energia, Contraste, Entropia, Homogenei-

dade, Probabilidade Máxima e Momento de Terceira Ordem. Estas características foram extraídas de diferentes combinações quanto a distâncias e ângulos. Os melhores resultados foram obtidos para $d = 1$ e $\theta = [0, 45, 90, 135]$, sendo que os seis descritores mencionados foram calculados para os quatro sentidos e concatenados, gerando um vetor de características com vinte e quatro componentes.

4. Experimentos e Discussões

Para avaliar a performance do conjunto de características proposto foram empregados três diferentes classificadores: *Support Vector Machines* (SVM), *Linear Discriminant Analysis* (LDA), and k*-Nearest Neighbors* (kNN). Para SVM, diferentes *kernels* foram testados e o Gaussiano produziu os melhores resultados. Os parâmetros para o *kernel*, γ e C, foram definidos empiricamente por meio de *grid search*. Todos os experimentos foram realizados considerando validação cruzada com três partições.

No primeiro experimento os três classificadores usaram as vinte e quatro características extraídas das GLCM. Os resultados são apresentados na Tabela 2 e demonstram que todos os classificadores alcançaram resultados similares com alguma vantagem para SVM.

LDA	kNN	SVM
85.6	85.5	88.6

Tabela 2. Classificadores usando características extraídas das GLCM

Com relação aos experimentos com o conjunto de características estrutural, a primeira questão consistiu em identificar o impacto do operador morfológico no processo de extração de características e, consequentemente, na performance dos classificadores. Para isso, foram extraídas características usando elementos estruturantes sob a forma de quadrado e cruz (ambos com dimensões 3×3 pixels) para o operador de erosão e diferentes quantidades de iterações para sua aplicação. Neste sentido, não foram encontradas diferenças quanto aos dois elementos estruturantes, sendo que a Figura 3 mostra os impactos em termos de performance para os três classificadores empregados. Como se pode ver, todos os classificadores apresentaram comportamentos similares para as diferentes quantidades de iterações usadas com o operador de erosão. Para um número pequeno de iterações, um grande número de componentes conexos foi considerado para a extração de características. Já para erosões mais agressivas foram removidos importantes componentes e reduziu-se drasticamente a performance de todos os classificadores. Ao final dos experimentos, os melhores taxas de reconhecimento foram alcançadas com o emprego de seis iterações.

A Tabela 3 mostra os resultados obtidos pelos três classificadores considerando as características extraídas das imagens após seis iterações do operador de erosão com o elemento estruturante quadrado. Também pode-se perceber que o tamanho da janela usada para o *threshold* adaptiva teve um pequeno impacto nas taxas finais de reconhecimento. Apesar disto, a definição das dimensões é importante, pois o tempo de processamento de cada imagem cresce à medida que as dimensões diminuem. Diante destes fatores (taxas de acerto e custo computacional), a melhor opção constitui o emprego de janelas maiores, isto é, com dimensões de 100×100 pixels.

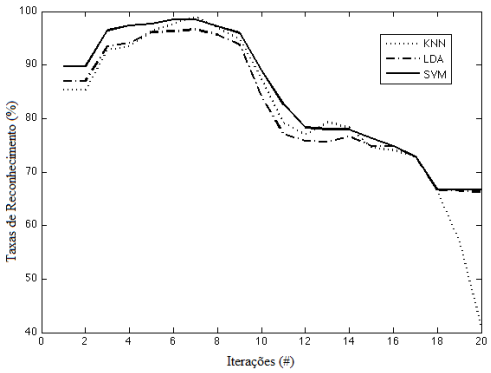


Figura 3. Impacto da erosão na performance dos classificadores.

Os melhores resultados foram produzidos pelos classificadores kNN e SVM, os quais são superiores aos resultados obtidos pelos classificadores treinados com características extraídas das GLCM em cerca de 10%. A principal justificativa para a pobre performance apresentada pelos classificadores treinados com características baseadas em GLCM é que, em alguns casos, tanto Angiospermas qunato Gimnospermas apresentam texturas similares. Nestes casos a única característica capaz de diferenciar estes grupos é a presença de vasos. Infelizmente, muito frequentemente os vasos são muito pequenos, são absorvidos pela textura e quase não têm influência ao se considerar o vetor de características como um todo.

A figura 4 compara o ROC (*Receiver Characteristic Curves*) para SVM treinado com as características estruturais e as extraídas das GLCM. Como se pode observar no gráfico, o classificador treinado com o vetor de características estruturais compacto é mais confiável ao se considerar Falsos Positivos, mesmo quando há altas taxas de *True* Positivos.

Classificadores	Tamanho da Janela		
	8×8	50×50	100×100
LDA	97.2	97.3	96.7
kNN	98.8	98.8	98.8
SVM	98.0	98.1	98.5

Tabela 3. Classificadores usando as características estruturais

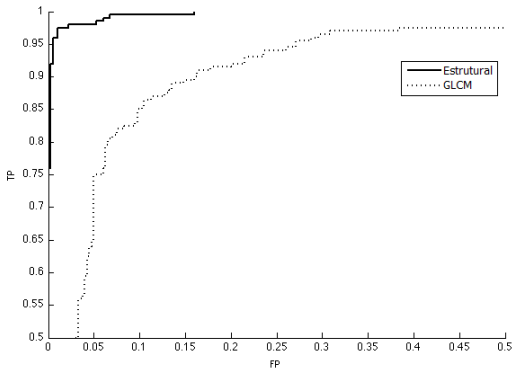


Figura 4. ROC para classificadores SVM treinados com características estruturais e GLCM.

5. Conclusões

Neste artigo foi introduzido um novo e compacto conjunto de características estruturais para diferenciar os dois grupos existentes de espécies florestais, isto é, Angiospermas e Gimnospermas. O descritor baseou-se na captura da estrutura dos principais componentes conexos da imagem usando medidas estatísticas simples. Estes componentes conexos foram identificados com a aplicação de *threshold* adaptativo e uma sequência de erosões quadradas. Os resultados dos experimentos mostraram que o conjunto de características proposto obteve melhores resultados quando comparados com características extraídas com o emprego da tradicional técnica GLCM, em cerca de 10%. Neste sentido, estes resultados apresentados são bastante promissores, pois podem ser empregados em uma fase de pré-classificação em qualquer sistema de reconhecimento de espécies florestais, reduzindo possíveis confusões entre espécies de diferentes famílias.

6. Trabalhos Futuros

Os próximos experimentos se concentrarão na ampliação da base de dados empregada, focando a validação do conjunto de características proposto diante de um número maior de espécies. Além disso, também se vislumbra a realização de testes com outros descritores, a análise de suas taxas de reconhecimento e sua comparação com os resultados apresentados neste trabalho. Por fim, um outro experimento que se mostra interessante constitui a combinação do descritor aqui apresentado com outros e subsequente análise do desempenho desta estratégia quando comparado com a aplicação individual de cada um dos descritores.

Referências

[1] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, New Jersey, 3 edition, 2008.

[2] R. M. Haralick. Statistical and structural approaches to texture. *IEEE*, 67(5):786–804, 1979.

[3] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics In Systems*, 3(6):610–621, November 1973.

[4] H. Kauppinen. A two stage defect recognition method for parquet slab grading. *15th International Conference on Pattern Recognition (ICPR'00)*, 4, Sep.

[5] M. Khalid, E. L. Y. Lee, R. Yusof, and M. Nadaraj. Design of an intelligent wood species recognition system. *International Journal of Simulation Systems, Science & Technology Special Issue on: Artificial Intelligence*, pages 9–17, 2008.

[6] L. Otsu. A threshold selection method from gray-level histogram. *IEEE Transaction on Systems Man. and Cybernetics*, 8:62–66, 1978.

[7] P. L. Paula Filho, L. E. S. Oliveira, A. S. Britto Jr., and R. Sabourin. Forest species recognition using color-based features. *20th International Conference on Pattern Recognition (ICPR2010)*, pages 4178–4181, 2010.

[8] H. Pedrini and W. R. Schwartz. *Análise de Imagens Digitais: princípios, algoritmos e aplicações*. Thomson Learning, São Paulo, 2008.

[9] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1984.

[10] L. Thomas and L. Mili. A robust gm-estimator for the automated detection of external defects on barked hardwood logs and stems. *IEEE Transaction on Signal Processing*, 55:3568–3576, 2007.

[11] J. Y. Tou, P. Y. Lau, and Y. H. Tay. Computer vision-based wood recognition system. *International Workshop on Advanced Image Technology*, pages 197–202, 2007.

[12] M. Tuceryan and A. K. Jain. *Texture Analysis in The Handbook of Pattern Recognition and Computer Vision*. World Scientific, 1998.

Segmentation Techniques based on Background Subtraction and Supervised Learning: A Comparative Study for Images of Mice and Human Skins

Bruno Brandoli Machado, Wesley Nunes Gonçalves and Jonathan de Andrade Silva
Institute of Mathematical Sciences and Computing (ICMC)
Physics Institute of São Carlos (IFSC)
University of São Paulo (USP) – São Carlos, SP, Brazil
{brandoli,jandrade}@icmc.usp.br
{wnunes}@ursa.ifsc.usp.br

Vinícius Saueia, Kleber Padovani de Souza, Bruno Toledo and Hemerson Pistori
INOVISAO and Biotechnology Department
Dom Bosco Catholic University (UCDB) – Campo Grande, MS, Brazil
{vsaueia,padovani,bc.toledo,pistori}@ucdb.br

Abstract

This paper presents a comparison between two image segmentation approaches based on background subtraction and supervised learning. Real images from two important issues, which have been studied by several computer vision research groups, were used in our experiments: namely, sign language interpretation and mouse behavior classification. According to performance measures, such as accurate rate, Jaccard coefficient, Yule coefficient, relative area error, and misclassification error, best results were obtained by background subtraction segmentators using images with complex background, otherwise, segmentators based on support vector machines outperformed when simple background were used.

keywords: *Mouse image segmentation, human skin segmentation, background subtraction, supervised learning*

1. Introduction

Segmentation is an important step in several computer vision systems. The essential idea is to split the image content into interesting and irrelevant objects. The classification of objects as interesting or irrelevant is highly dependent on the application domain. Thus, evaluating segmentation techniques in different domains is an important issue. Gesture recognition is an important element for many vision based human-to-computer interaction or computer mediated human-to-human communication [9]. Segmentation of human skin is an essential stage for vision-based gesture recognition, mainly when signers do not use additional resources, including colored gloves or markers. Another application addressed in this paper is related to behavior identification of mice under drug effect during experiments for medicine development, such as open-field experiment [10].

Different applications of image segmentation were reported in the literature. A color based segmentation method was proposed in [4], employed for a real-time system using a frame rate of 10 fps. The method was evaluated on synthetic and real images by means of outcome and manually segmented images. The metrics used were processing time, false positive, false negative, medium error of area and center of object. Terrillon and Fukamachi [15] compared nine color spaces for image segmentation. They presented two colors-based segmentation models: Gaussian model based on Mahalanobis distance and Gaussian mixture model. The aim was to segment human faces on colored images.

A segmentation algorithm for moving objects was presented in [3]. This algorithm combines the adaptive background subtraction with frames differentiation algorithm. Differentiation technique was used to determine which regions are moving and adaptive background subtraction assists on the identification of the whole moving region. In [8], a comparison of many subtraction background techniques described in the literature is showed. One of the eight compared techniques was Adaptive Gaussian mixture. For this technique, a K Gaussian mixture is built for each pixel, representing the background model. After object detection, the Gaussian parameters are updated based on an update constant. The techniques were evaluated using three parameters: distinction between foreground pixels and background

pixels, background storage and updating over time and post-processing of the objects to eliminate false positive of the resultant image.

This paper presents a comparative study of segmentation techniques applied to sign language recognition and mouse behavior analysis. The experiments were performed on a set of 40 images of mice and on another set of 240 images of Brazilian Sign Language gestures. Ground-truth segmented images were produced for all samples with the help of field specialists. Here we use five well-know measures to help us to assess the quality of the output images, thus reducing the subjectiveness of evaluations based purely on the visual analysis. The five performance measures are: accurate rate, Jaccard coefficient, Yule coefficient, relative area error and misclassification error [14]. In this sense, the main contributions of our study is to compare both groups of techniques and give directions to the systems in issue.

The remaining of this paper is organized as follows. Section 2 presents a background with regard to segmentation techniques. In Section 3 the performance measures are described. Experiments and results are discussed in Section 4. Finally, conclusions and future works are presented in Section 5.

2. Segmentation Techniques

This section presents an overview of the segmentation techniques used in this work: background subtraction, adaptive background subtraction, Gaussian models, decision trees, artificial neural networks and support vector machines. For more details of the techniques, we refer the reader to the reference papers.

2.1. Background and Adaptive Background Subtraction

Background subtraction is among the most used techniques in computer vision because it is easy to implement and demands low processing time [2, 7]. The underlying idea is to subtract an image from a reference image (a fixed background image) which does not contain interesting objects. Each new image is segmented using a pre-defined threshold. Given an image $\mathbf{I}_t$, at the instant t (frame), and $\mathbf{B}_t$ is the background image, the output image is achieved by:

$$|\mathbf{I}_t(x) - \mathbf{B}_t(x)| > \tau \quad (1)$$

where (x) and τ are the spatial position of each pixel and a predefined threshold, respectively.

However, background subtraction does not update the background information and therefore may not suit image changes on scenes, such as different illumination conditions or irrelevant objects that suddenly appear in images and no longer move. To overcome this drawback, an adaptive variant that updates the reference image over time was proposed in [6], turning this method more robust than the first one. In this variant the reference image is iteratively adjusted as follows:

$$\mathbf{B}_{t+1} = \alpha\mathbf{I}_t + (1-\alpha)\mathbf{B}_t \quad (2)$$

where $\alpha \in [0, 1]$ is an updating constant.

2.2. Segmentation based on Supervised Learning

Machine learning techniques [1] can be used for image segmentation purpose by learning a classifier from pixels of an object of interest $\mathbf{O}$. Pixels $\mathbf{p} = [r, g, b] \mid \mathbf{p} \in \mathbf{O}$ are extracted to compose the training set $\mathbf{S} = \{\mathbf{p}\}$, where r, g, b are red, green and blue components, respectively. In order to learn the classifier from $\mathbf{S}$, four supervised learning techniques are briefly described in this study:

Gaussian statistical learning: In this strategy [17], the parameters of a multivariate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ are estimated from the object pixels $\mathbf{S}$, defined by:

$$\mu = \frac{1}{|\mathbf{S}|}\sum_{\mathbf{p}\in\mathbf{S}}\mathbf{p}, \quad \Sigma = \frac{1}{|\mathbf{S}|}\sum_{\mathbf{p}\in\mathbf{S}}[\mathbf{p}-\mu][\mathbf{p}-\mu]^t \quad (3)$$

where μ is the mean, Σ corresponds to the covariance matrix, $\mathbf{p}$ is the training set vector. After the training step, a new pixel $\mathbf{p}'$ is assigned to the object of interest if its Mahalanobis distance with respect to $\mathcal{N}(\mu, \Sigma)$ is higher than a pre-defined threshold. The Mahalanobis distance is calculated by:

$$M_d = [\mathbf{p}-\mu]^t\Sigma^{-1}[\mathbf{p}-\mu] \quad (4)$$

Decision trees: Induction of decision trees [12] is a divide-and-conquer strategy for classification tasks. The algorithm splits the instance space into decision regions by generating internal or test nodes. The classification of an input instance is based on the best features that separate the data. The procedure starts by generating the root node, taking into account the whole training set. Interior nodes divide the set of instances by testing a specific feature and each child of that node will correspond to a distinct value represented by that feature. This procedure is repeated until leaf nodes are obtained. Leaf nodes will predict the class of the instances according to the path down of the final tree. Here, we use the C4.5 algorithm that splits each node on the feature with the highest information gain. The gain on a feature F with value v is defined as:

$$Gain(S, F) = Entropy(S) - \sum_{v\in F}\frac{|S_v|}{|S|}Entropy(S_v) \quad (5)$$

where S and S_v denote a set and a subset of instances, respectively. The gain is calculated by using impurity measures for quantifying the quality of the split generated by the trained model. In this study, entropy has been used.

Artificial neural networks: MLP [13] is a particular architecture of feedforward neural network which addresses data that are not linearly separable. A MLP net consists of one or more hidden layers between an input and output layer of neurons. Each neuron is fully connected from one layer to the next and can be described as a processing element which is activated by a nonlinear activation function. The neural network activation is the inner product of the input vector ($\mathbf{p}$) with the weights (connections) at hidden layers. For training the MLP, the backpropagation algorithm has been adopted (we refer to [5] for more details).

Support vector machines: SVM [16] is a widely used technique for data classification. Given a training set of two instance-label (x_i, y_i), for $i = 1, \ldots, l$ where $x_i \in R^n$ and such that $y_i = \pm 1$, a SVM finds a hyperplane that separates a couple of classes with the maximal margin in the higher dimensional space ϕ. Training a SVM requires the solution of an optimization problem with a very large quadratic programming. Alternatively, the sequential minimal optimization algorithm (SMO) [11] is a fast optimization that works by breaking the quadratic algorithm into a subset of smallest problems. For implementing the support vector machines strategy, the SMO has been used, along with polynomial basis function transformations.

3. Evaluation Methods

Visual evaluation of image segmentation algorithms is not a reliable criterion to analyze final results. Hence, there are many different ways for evaluating them. The performance can be measured through correcteness at pixel level or image regions. In this paper, we compute a total of five measures. Next we describe three of them based on pixels and two based on region properties.

3.1. Pixel-based Evaluation

In this approach, the results are based on the comparison between the ground-truth and output images. Each analyzed pixel can be classified into four possible labels, namely: (1) true positive, (2) false positive, (3) true negative, and (4) false negative. A true positive (TP) occurs when the outcome from a prediction corresponds to the ground-truth, otherwise, a false positive (FP) is found. Conversely, a true negative (TN) occurs when the outcome from a prediction is different from ground-truth image, or else, a false negative (FN) is considered.

Following these labels, it is possible to compute three quantitative measures regarding two binary images: the percentage correct classification (PCC), or as known as accurate rate, the Jaccard coefficient (JC) and the Yule coefficient (YC). Such measures are formalized in Equations 6, 7 and 8, respectively.

$$PCC = \frac{TP + FP}{TP + FP + TN + FN} \tag{6}$$

$$JC = \frac{TP}{TP + FP + FN} \tag{7}$$

$$YC = \mid \frac{TP}{TP+FP} + \frac{TN}{TN+FN} - 1 \mid \tag{8}$$

3.2. Region-based Evaluation

This approach judges the quality of the segmentation using blobs. Moreover, the performance measures are scored ranging from 0 (zero), for a totally correct segmentation, to 1 (one), for an incorrect case. The relative foreground area error (RAE) is obtained through shapes and areas from the segmented image regarding ground-truth image [14]. RAE value is 0 wheter a complete matching between overcome and ground-truth images are achieved, while the minimum matching is 1. Equation 9 defines RAE measure, where A_0 is the area of ground truth image, and A_t is the area of segmented image.

$$RAE = \begin{cases} \frac{A_0 - A_t}{A_0}, if A_t < A_0 \\ \frac{A_t - A_0}{A_t}, if A_t \geq A_0 \end{cases} \tag{9}$$

Another considered metric is defined as misclassification error (ME). It is obtained by computing the percentage of background pixels erroneously set as foreground, and foreground set as background. The ME is formalized in Equation 10, in which B_0 and F_0 refer to background and foreground of ground truth image; B_t and F_t indicate the background and foreground of the segmented image; and $|.|$ represents the cardinality for the set.

$$ME = 1 - \frac{|B_0 \cap B_t| + |F_0 \cap F_t|}{|B_0| + |F_0|} \tag{10}$$

4. Experimental Results

In order to evaluate the algorithms described in this paper, experiments are performed on two different colored image datasets. First, the datasets used for evaluation and experimental details are described. Then, the results are discussed.

4.1. Datasets

The Mouse Behavior database consists of two behaviors: vertical exploration and spatial locomotion [10]. 640 × 480 images are taken of the species Swiss and C57 within

a circular arena, resulting in a total of 40 images. Swiss are white haired animals and C57 black haired animals. We analyze the ability of segmentation algorithms in tasks where foreground objects are strongly correlated with background. In this particular case, the animal is the same color as the background. Both behaviors and two backgrounds are presented in Figure 1.

The Brazilian Sign Language database contains 240 images of postures that explore different gestures separated in 10 classes. The size of images is 800 $\times$ 600 pixels. Images are taken of 6 different signers with distinct skin tones. Each pose is captured with a static and complex background, in this case within a laboratory with artificial illumination. Further, the remaining 120 images are taken in outdoor environments with natural illumination and complex background. Some samples with all sort of actions with complex and static background are shown in Figure 2.

4.2. Experimental Setup and Sampling

Experiments are carried out with two groups of image segmentation techniques, as described in Section 2. The comparison is performed by attaining performance measures. First, we create by hand the ground truth images of both databases, a time consuming and laborious process. Additionally, for background subtraction techniques a reference image is obtained.

Samples with size of 40 $\times$ 40 pixels are collected in order to train the supervised learning models. For gesture images, samples of different parts of the body are extracted, such as parts of face, right arm, left arm and neck. For mouse behavior images, five samples of the animal and five samples of the arena are selected. Parameters for both approaches are chosen by means of empirical range for each algorithm, which result over 70,000 segmented images. Then, performance measures are calculated over segmented images. For our datasets results, the average precision is taken as the performance metric for determining the accuracy of the segmentation algorithms.

4.3. Results

Experiment 1: First, a comparison with different combinations of backgrounds and coat color of the Mouse Behavior database is shown in Figure 3. The combinations are indicated in the y axis of the graph, while the average precisions are indicated in the x axis. The highest values are obtained by the contrast between background and coat color, for instance: the black background and the Swiss mouse. For the supervised learning techniques, only the combination white background obtained good results and the remaining degrade in performance.

Experiment 2: Figure 4 shows the results achieved using the combination between backgrounds and skin tones under natural and artificial illumination. The combinations are indicated in the y axis of these graphs, while the average precisions are indicated in the x axis. We can observe that the background subtraction techniques in presence of artificial illumination do not perform well for these four combinations. There is a clear distinction in performance for simple backgrounds between these techniques and the supervised learning ones. Within the group of supervised learning, decision tree, artificial neural networks and support vector machines techniques perform much better than background subtraction and adaptive background subtraction. For natural illumination, it is also observed that the combination for both simple backgrounds perform better compared with artificial illumination. This is consistent with the shadow done for the wall. Conversely, for this case we can see that the supervised learning approach has the same behavior independent of the illumination condition, which they obtained the higher values.

5. Conclusions and Future Works

This paper has evaluated two different groups of segmentation techniques applied on two important real-world applications: gesture recognition for Brazilian Sign Language and mouse behavior analysis. We examined the challenges including illumination change and strong correlation between foreground objects and background. The evaluation was based on well known performance metrics which compare the segmented images with ground truth ones. The results presented here support that each algorithm performs successfully in a particular image database. Experiments strongly suggest that the background subtraction approach is better for mice images, mostly in cases with distinguished contrast. However, is also need to focus attention on basic features, for instance, C57 mouse is a black coat animal with short hairs, which at a young age their epidermal structures are quite visible. From the gesture recognition results, supervised learning approach achieve better results for both illumination conditions, in specific support vector machines. Future research should focus on different image segmentation techniques. We believe that another focus of attention is to explore unsupervised approaches.

Acknowledgment This work has been partially funded by UCDB, FINEP and FUNDECT. Some of the authors have received scholarships from CNPQ and CAPES during the project.

References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

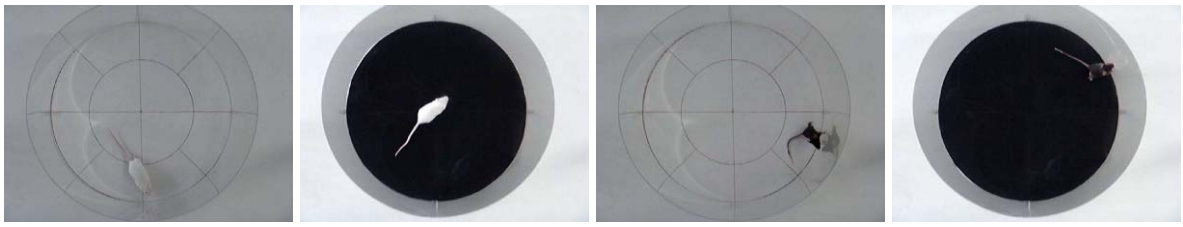


Figure 1. Examples of the two mouse behaviors: spatial locomotion and vertical exploration. From left to right, the first two images are mice in spatial locomotion, while the other two are mice in vertical exploration.

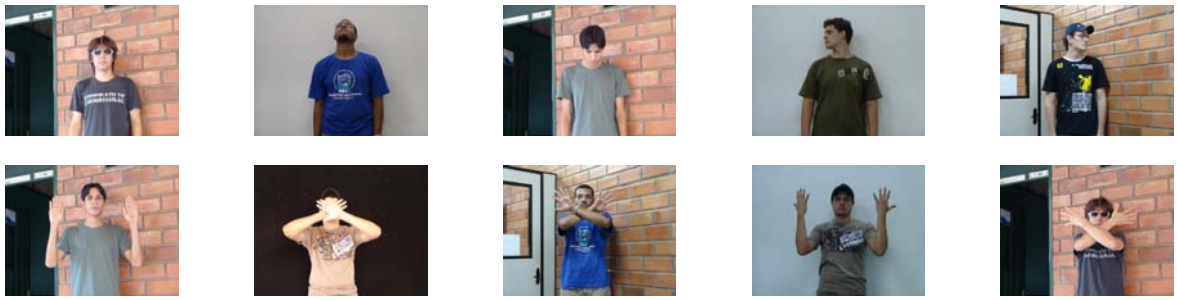


Figure 2. Ten different poses of 6 different signers taken of outdoor and indoor environments, showing the difference in illuminations and backgrounds. We are interested in evaluating the robusteness when signers wear accessories, such as sunglasses, lens and cap.

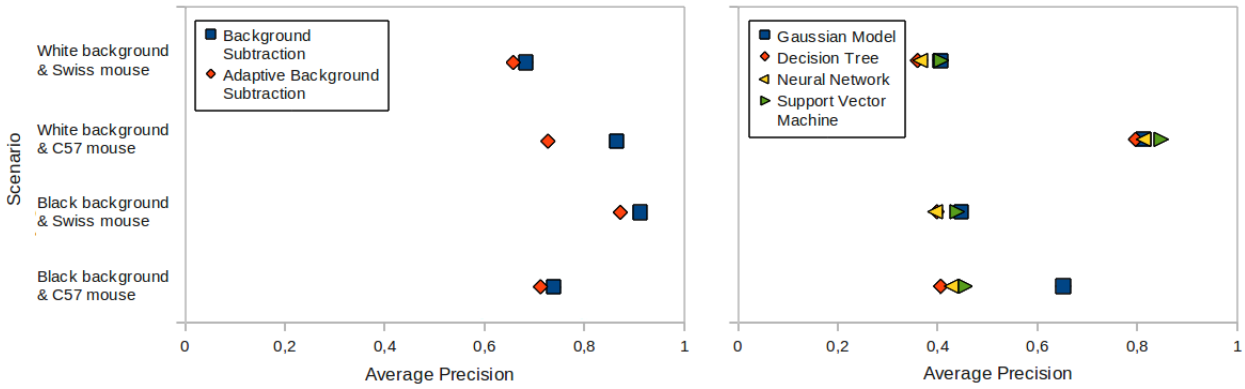


Figure 3. Mouse Behavior database evaluation influenced by combinations of background and coat color. Swiss are white haired animals and C57 black haired animals. In this experiment just artificial illumination is evaluated because experiments shall occurs in indoor environments.

[2] S.-C. S. Cheung and C. Kamath. Robust background subtraction with foreground validation for urban traffic video. *EURASIP J. Appl. Signal Process.*, 2005:2330–2340, January 2005.

[3] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa. A sys-

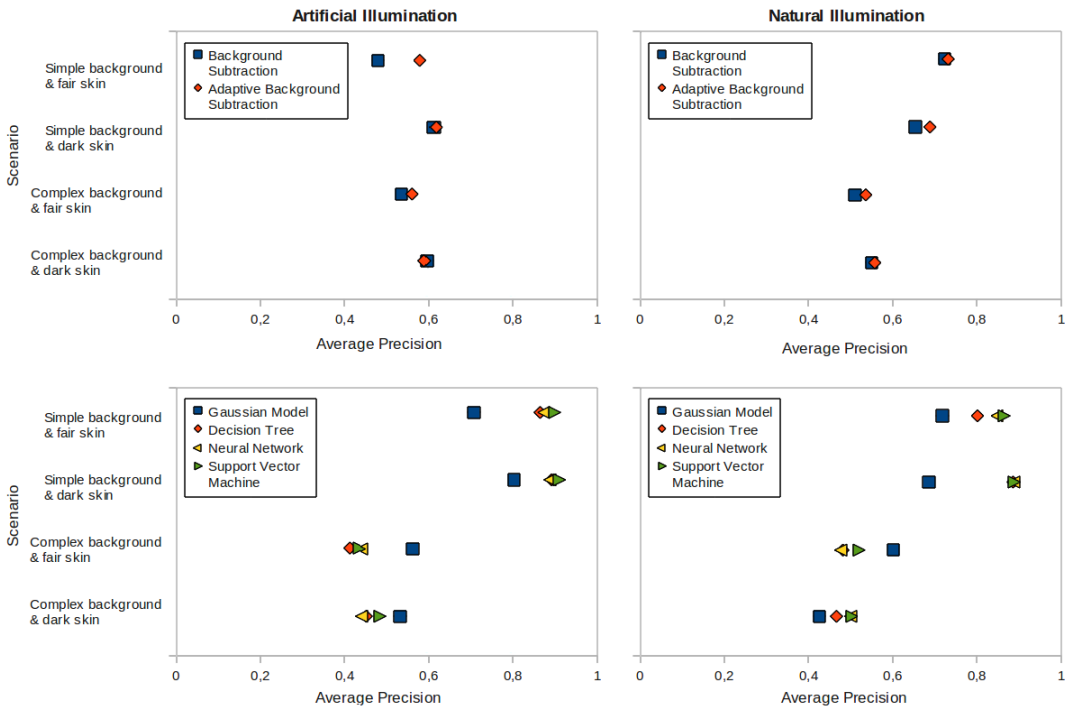


Figure 4. Evaluates combinations of backgrounds and skin tones of the Brazilian Sign Language database. We test both approaches under two illumination conditions: artificial and natural.

tem for video surveillance and monitoring. Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 2000.

[4] R. Cucchiara, A. Prati, and R. Vezzani. Real-time motion segmentation from moving cameras. *Real-Time Imaging*, 10(3):127–143, 2004.

[5] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2nd edition edition, 2001.

[6] J. Heikkilä and O. Silvén. A real-time system for monitoring of cyclists and pedestrians. In *Proceedings of the Second IEEE Workshop on Visual Surveillance*, VS '99, pages 74–, Washington, DC, USA, 1999. IEEE Computer Society.

[7] M. H. Khan, I. Kypraios, and U. Khan. A robust background subtraction algorithm for motion based video scene segmentation in embedded platforms. In *Proceedings of the 7th International Conference on Frontiers of Information Technology*, FIT '09, pages 31:1–31:6, New York, NY, USA, 2009. ACM.

[8] A. McIvor. Background subtraction techniques. In *Proceedings of Image and Vision Computing*, 2000.

[9] G. R. S. Murthy and R. S. Jadon. A review of vision based hand gesture recognition. *International Journal of Information Technology and Knowledge Management*, 2(2):405–410, July-December 2009.

[10] H. Pistori, V. V. Viana Aguiar Odakura, J. a. B. Oliveira Monteiro, W. N. Gonçalves, A. R. Roel, J. de Andrade Silva, and B. B. Machado. Mice and larvae tracking using a particle filter with an auto-adjustable observation model. *Pattern Recognition Letters*, 31:337–346, March 2010.

[11] J. C. Platt. *Fast training of support vector machines using sequential minimal optimization*, pages 185–208. MIT Press, Cambridge, MA, USA, 1999.

[12] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1:81–106, March 1986.

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning internal representations by error propagation*, pages 318–362. MIT Press, Cambridge, MA, USA, 1986.

[14] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168, 2004.

[15] J.-C. Terrillon, H. Fukamachi, S. Akamatsu, and M. N. Shirazi. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 54–63, Washington, DC, USA, 2000. IEEE Computer Society.

[16] V. N. Vapnik. *Estimation of Dependences Based on Empirical Data [in Russian]*. Nauka, Moscow, Russia, 1979. (English translation: Springer Verlag, New York, 1982).

[17] V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, New York, NY, 1998.

Computer Vision applied to Dengue's Larvae Death Rate Calculation: Preliminary Results

SOUZA, K. P.; QUEIROZ, J. H. F. S;
PISTORI, H.
INOVISAO Universidade Católica Dom Bosco
Av. Tamandaré, 6000 - Jardim Seminário
Campo Grande - MS - CEP: 79117-900
jaime.cpp@gmail.com, padovani@ucdb.br, pistori@ucdb.br

Abstract

This work investigates the applicability of some computer vision and pattern recognition algorithms in Aedes aegypt larvae behavior classification. Hidden Markov models and four different supervised learning techniques have been experimented, alone and in combination, to classify image sequences of dead and live larvae inside a recipient containing liquid substances.

1. Introduction

Dengue's epidemic aroused in recent years in Brazil reinforcing the need for more effective ways to combat its transmitter mosquito. Experiments with new larvicides involve the analysis of larvae mortality using various concentrations and combinations of different larvicides in several repetitions. Hundreds and even thousands of experiments should be performed to find a substance that produces the desired effect. During the development of some larvicide, lab tests are performed in order to analyze its substances effect. Larvae death rate is one of those tests, obtained by visual examination of recipients with larvae exposed to test substances during 24 hours. At certain times, a specialist observes recipients, identifies and records the amount of live and dead larvae. At the end of the experiment, based on these records, some conclusions are reached regarding the effect of larvicide analyzed [1].

The counting of live and dead larvae is generally performed by humans, who naturally have their limitations, such as subjectivity, inaccuracy and exhaustion. Thus some counting errors are possible. In this paper, some results related to a system that aims to automate the counting of mortality of larvae is reported. Images of larvae are captured by a fixed camera positioned above the recipients containing some larvae [2]. Subsequently, these images are processed by image processing and computer vision algorithms. Then, based on information from the previous processing, each behavior of larvae is classified and recorded. The goal of this work is to analyze the application of Hidden Markov Model (HMM) and some other supervised learning approaches, in the classification stage, exploring different initial probabilities, criteria for stopping training and associations between classifiers.

The initial probabilities used for the models were calculated using uniform distributions and direct estimation from training data. The training modes considered different stop criteria, which directly interferes in the number of iterations of the algorithm used. Finally, we used two modes of classification: the traditional, which considers the probabilities of symbols generation of models, and combined, where machine learning techniques are embedded in HMM-based classifier's classification. This combined approach can achieve superior performance in certain applications than traditional HMM's [3]. This improvement can be seen especially in applications whose classifications show considerable uncertainty.

The next section provides a brief theoretical background on the technique explored in this work, the HMM, and issues related to it. In Sections 3 and 4, we describe experiments and results, respectively. The fifth section presents the analysis of these results, discussing the performance of classifiers and pointing out possible justifications for relevant behaviors, and suggesting future works.

2. Hidden Markov Models Review

The Hidden Markov Models (HMM) are widely applied in modeling problems with temporal variations. The classification with HMMs usually occurs through the creation of an HMM for each class of problem. To classify a given sample among the known classes, the probability of each

HMM generating this sample is calculated. Finally, the sample is classified as belonging to the class associated with the HMM that presents the highest probability of generating it (most likely HMM).

In [3], Wang describes a classification approach that uses a combination of HMM with the technique of k nearest neighbors (KNN) in order to minimize the misclassification present in traditional approach, which appears especially when the difference between the values of the probabilities of the most likely HMMs is small. On this basis, they propose the use of a KNN classifier in a second classifying round for the cases where the difference mentioned above is not significant. According to the authors, this combination adds little computational cost and increases the accuracy rate of the classifier, which was increased from 82.1% to 88.3% in the recognition of facial expressions. Other works, such as [4, 5], also use similar combinations.

In this paper, we used an approach of combining HMM with machine learning techniques similar to that used in [3]. However, different machine learning algorithms, such as decision trees, support vector machines and neural networks, were used, besides KNN. When evaluating a sequence of observations, the probability of generating this sequence for each HMM's is calculated. If the difference between the highest probability and any other probability is below a certain threshold, the second level classifier is called. Otherwise, it assumes that the sample belongs to the class of the HMM that produced the higher score.

3. Experiments, Results and Analysis

The experiments in this study were conducted using sequences of images of larvae in small containers containing some liquid substance. Briefly, the aim of these experiments is to automatically identify certain behaviors of the animals in these sequences. The behaviors to be identified are alive and death. Initially, the images of larvae were captured and later processed, so that relevant information can be extracted. Then, this information was provided to the classifiers in training and classification modes.

3.1. Acquisition and image treatment

Three videos with 1300 frames each containing the image of one larva in a container with a liquid substance were produced. Two of them correspond to images of dead or dying larva and the other one to images of live larva. Because there are multiple occurrences of the behaviors in the three videos, smaller shots have been produced from the original videos, 24 shots for dead and 10 for live larva. Each shot is made up of 100 frames, approximately (the shots have different sizes). Two thirds of the shots, for each class (dead/live), were used to train and one third to test, in all experiments. Images were semi-automatically segmented using a combination of background subtraction, machine learning and manual manipulation.

Thirteen features were extracted from each segmented frame. Three of them are the X and Y coordinates of the center of mass of the larva and the first moment of Hu, which was obtained by means of statistical moments. Other six correspond to the values of a 6-bin histogram of the angles of the larva shape, which are calculated using the k-curvature algorithm. The last four are the shape descriptors: aspect ratio, form factor, roundness and compactness. Exploratory experiments showed that the shape descriptors alone were sufficient to produce good discrimination, and so, the other nine have been discarded for subsequent experiments. In order to produce a discrete sequence of symbols to feed the HMM, each frame was mapped, using a machine learning algorithm, to two states: stretched or curved larvae.

3.2. Methodology for training and classification

Traditionally, to classify a given sequence of observations, a classifier based on HMM compares the odds of generating the sequence of observations in each model. Therefore, the sequence is classified as belonging to the class whose model showed the highest probability of generating it. In the experiments, a combination of HMM and other classifiers have also been tested. A set of feature vectors composed of the output of each HMM, together with the class information (dead or alive), is used to train another classifier.

Several experiments have been conducted in order to evaluate the best configuration of the HMM for this problem, including HMM initialization and stopping strategies. Three different initialization strategies were analyzed: (1) random, (2) pre-computed automatically and (3) pre-computed manually; together with three different stopping criteria: (1) fixed number of iterations, (2) threshold and (3) output stability. In the experiments with probabilities pre-computed automatically, tests have also been used to find the best number of HMM internal states.

In the case of random initialization, an uniform distribution is assumed for the initial states, transition and emission matrices. For the ''pre-computed automatically' case, initial states and transition also assume an uniform distribution, however the cluster algorithm K-Means is used to initialize the emission matrix, using the strategy described in [6]

Finally, the initialization called "pre-computed manually' is based on direct estimates of all the parameters but requires the manual preparation of a training set that includes a mapping from each frame (observation) to an in-

ternal state. In this case, the internal states corresponds to larva in two groups of shapes that are relatively easy to be identified by visual observation: stretched and curved. Figure 1(a) illustrates an example of larva stretched and Figures 1(b) and 1(c) illustrate examples of curved larvae. These shapes may happen in both dead and live larvae.

(a) Stretched (b) Curved (c) Curved

Figure 1. Images of larvae in different postures.

The Baum-Welch iterative algorithm has been used to learn the parameters for each HMM. Some experiments have been conducted in order to find the stopping criteria for the problem. The first criterion stops iteration when the probability of the training set given the current HMM parameters is not greater than the probability given the previous parameters [7]. The second criterion stops iteration when the difference between the probability using the current and previous iteration parameters is smaller than a pre-defined threshold (in the experiments, these thresholds ranged from 10^{-1} to 10^{-6}). The third criterion uses a fixed number of iteration. The values 9 (default for JAHMM[1]), 50, 100, 500 and 1000 iterations were tested. The JAHMM's default value has been used as a baseline performance in the experiments.

Additionally, we analyzed the hit rates for this application using a combination of HMM and other classifiers. In these experiments, a second level classifier is called whenever the difference between the values returned by each HMM was bellow a predefined threshold. Decision trees (J48), k-Nearest Neighbor (IBK), support vector machines (SVM) and neural networks were tested as the second level classifier. These same classifiers have also been used directly, without HMM. As these classifiers require a fixed size attribute vector, and the sequences of frames have different sizes, the training and test set had to be preprocessed. At first, we counted the occurrences of each posture, stretched (S) or curved (C) in the sequence of frames, producing in this way two new attributes. Another four attributes were produced by counting the number of occurrences of all possible pair of postures: SC, CS, SS and CC. One last attribute corresponded to the number of times that the larvae changed its posture in the sequence.

1 JAHMM is a free open source library which contains HMM algorithms implementations.

3.3. Experiments configuration

The experiments performed in this work can be classified into four different groups: (1) initialization strategies, (2) stopping criteria, (3) combination of HMM and second level classifiers and (4) classifiers without HMM. Hit rates were produced for all the experiments and area under the ROC curves (AUC) for some of them.

3.3.1. Initialization Strategies Table 1 presents the hit rate and AUC values for all the initialization strategies. It was observed that, according to Rabiner [8], the HMM's whose initial parameters are calculated from the training set, manually or automatically, showed better results compared to HMM's with random initial probabilities. Even in this regard, the HMM's with initial probabilities precomputed automatically presented hit rates that are even higher than those computed manually. This information shows that in some cases, the manual calculation of the initial probabilities (which can be exhaustive and impractical) becomes unnecessary, given the existence of automatic mechanisms for obtaining these probabilities.

Strategy	Hit Rate(%)	AUC
Random	59%	0.625
Pre-comp. manually	82%	0.8292
Pre-comp. automat. - 2 states	88%	0.7917
Pre-comp. automat. - 3 states	91%	0.7875
Pre-comp. automat. - 4 states	88%	0.7875

Table 1. Hit Rate and AUC for different initialization strategies

3.3.2. Stopping Criteria As described earlier, we used three new criteria to compare against the standard criteria of JAHMM (which uses a fixed number of iterations equal to nine). Next, we present the results obtained considering each of the initial models. All stopping criteria were combined with all initialization strategies, as they could be dependent. With random initial probabilities, few changes were found. Using the first criterion (stop when no increase in probabilities happens), there was an increase in hit rate of 12% (in relation to JAHMM default criterion). However, this is not a significant increase, because all instances were classified as belonging to the class alive. Because this class has more instances in the training set, the final probability increased. In the second criterion (stop when difference in probability is bellow threshold), no change in hit rate were produced. The same occurred with the third criterion when the number of iterations was less than 500. From this, all instances were also classified as belonging to the class alive, as happened with the first criterion.

Whereas initial probabilities were precomputed manually, it was noticed a small change in the hit rates in two stop criteria. The first showed no changes. In the second, considering a threshold of 10^{-1}, an increase of 3% in accuracy was achieved. With thresholds equal to 10^{-2} and 10^{-3}, the increase was 12%. When the thresholds were under these values, the increase was 6%. The third criterion produced an increase of 6% in all variations of iterations. Finally, the hit rates when initialization were precomputed automatically (with two states) was reduced with stopping criteria. With the second criterion, the reduction was 17% and, with the first and third criteria of 3%. Considering all tests, the maximum increase recorded was 12% with the second stopping criterion and initial probabilities precomputed manually. However, these increase must be considered with care, as it seems that it results from overfitting the training set, which is biased toward alive larvae.

3.3.3. Combination of HMM and second level classifiers In all experiments with the combination of HMM and a second level classifier, using the best initialization strategy and stopping criteria found in the other experiments, there was no increase in hit rates, indicating that for this problem, a second level classifier is useless

3.3.4. Classifiers without HMM For comparison with previous experiments, the results of experiments performed with fixed-attribute-vector-length machine learning algorithms, using the same training and data set, are presented below . The algorithms used were IBK, J48, SVM and MLP with the three sets of features described earlier. The default parameters value from Weka[2] were used, except from IBK, which was fine tunned to a better K value.

The best results were obtained with the algorithm IBK, with a maximum AUC of 0.9792 and the worst results were obtained with the SVM algorithm, with AUC of 0.5 in all cases. This poor result with SVM may be related to the use of the default parameters in Weka. Both J48 and MLP presented better results (hit rates and AUC) than the HMMs. Although these classifiers were not initially designed to deal with temporal information, as it is the case of HMM, in these experiments, they presented much better results.

4. Conclusions and Future Work

Several experiments were performed in this work with images of larvae in recipients containing larvicides extracted from plants to evaluate the applicability of hidden Markov models to assist counting process of live and dead larvae. In the experiments, first, we analyzed the set of initial probabilities of the models. According to literature, these probabilities may significantly interfere in the training process and hence the classification. We analyzed three sets of probabilities, consisting of 1) uniform and random, 2) manually precomputed, and 3) automatically precomputed probabilities. Based on these experiments results, we conclude classifiers with models whose initial probabilities were precomputed obtained better results than those with random models. Moreover, it was observed that the results obtained with models precomputed automatically were close to those obtained with manually precomputed models, which indicates that the onerous task of obtaining initial probabilities may be unnecessary in some cases. Different stop criteria for training algorithm, Baum-Welch, were also analyzed. These experiments were used to check if the stopping criterion used in previous experiments was appropriate. In fact, few substantial changes were observed in these experiments with respect to previous ones. Thus, we can infer that, considering the computational cost and classification performance, the criteria previously used (nine iterations) was appropriate for the application.

Finally, considering classifiers using HMM's, last experiment examined the combination of such HMMs classifiers with machine learning algorithms. Such combination can improve the performance of these classifiers. In this work, the classifiers were combined with four algorithms – decision trees, k-Nearest Neighbor, support vector machines and neural networks. The parameter used for activation of the algorithms was based on the probabilities difference. In some circumstances, the auxiliary algorithms perform the final classification. No improving was detected in any experiments involving combined classifiers when the threshold used was less than 82%. In other cases, all rates were lower than those obtained with isolated classifier. For these experiments, classifiers were used with models with initial automatically precomputed probabilities with three states. These results may come from no relevant variations in the images used, especially in the images of dead larvae.

According to obtained results with classifiers based only on learning algorithms used in combinations, it was observed that the performance of classifiers based on HMM's obtained considerable mortality rates, but lower than some classifiers. Probably, this result is related to the pattern of behaviors existing in database, which might be better classified using finite numbers of features, ignoring temporal patterns in some cases. Another possible cause is the existence of noise in the entries, which may have been smoothed favoring some classifiers. Future work suggestions which can scientifically help with this application are given below.

The analysis of a larger set of different samples of live and dead larvae could find temporal patterns which can be identified by HMMs classifiers. Other algorithms for parameter reestimation may allow identification of behavior not considered by the algorithm Baum-Welch. Finally, it is also appropriate to analyze the performance of classifiers

2 Weka is a free open source library with machine learning algorithms.

based on machine learning algorithms (such as J48, IBK and MLP, which showed good results) in samples with different patterns. Additionally, the performance of other classifiers (or combinations of classifiers) in this problem can be analyzed.

5. Acknowledgment

This work has been partially funded by UCDB, FUNDECT and CPP. Some of the authors have received scholarships from CNPQ and CAPES during the project.

References

[1] W. N. Gonçalves, J. Silva, B. Brandoli, H. Pistori, and A. Souza. Markov models applied to snakes behavior identification. *Lecture Notes in Computer Science*, 4872, 2007.

[2] B. Q. Huang, C. J. Du, Y. B. Zhang, and M.-T. Kechadi. A hybrid hmm-svm method for online handwriting symbol recognition. In *ISDA '06: Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications*, pages 887–891, Washington, DC, USA, 2006. IEEE Computer Society.

[3] H. Pistori, V. V. Viana Aguiar Odakura, a. B. Oliveira Monteiro, Jo W. N. Gonçalves, A. R. Roel, J. de Andrade Silva, and B. B. Machado. Mice and larvae tracking using a particle filter with an auto-adjustable observation model. *Pattern Recognition Letters*, 31(4):337–346, 2010.

[4] K. R. A. Porto, A. R. Roel, M. M. Silva, R. M. Coelho, E. J. D. Schleder, and A. H. Jeller. Atividade larvicida do óleo de anacardium humile saint hill sobre aedes aegypti. *Revista da Sociedade Brasileira de Medicina Tropical*, 41:1–4, 2008.

[5] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, pages 42–53. IEEE Computer Graphics and Applications, 1989.

[6] Q. Wang and S. Ju. A mixed classifier based on combination of hmm and knn. In *ICNC '08: Proceedings of the 2008 Fourth International Conference on Natural Computation*, pages 38–42, Washington, DC, USA, 2008. IEEE Computer Society.

[7] J. Ye, H. Yao, and F. Jiang. Based on hmm and svm multilayer architecture classifier for chinese sign language recognition with large vocabulary. In *ICIG '04: Proceedings of the Third International Conference on Image and Graphics*, pages 377–380. IEEE Computer Society, 2004.

[8] S. J. Young, G. Evermann, M. J. F. Gales, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. C. Woodland. *The HTK Book, version 3.4*. Cambridge University Engineering Department, Cambridge, UK, 2006.

Extração de Características de Imagens Mamográficas Utilizando Redes Complexas para Detecção de Câncer de Mama

Jonathan S. Ramos
Universidade Federal de Rondônia – UNIR
Departamento de Ciências da Computação
Porto Velho, RO, Brasil
jonathansramos@gmail.com

Carolina Y. V. Watanabe
Universidade Federal de Rondônia – UNIR
Universidade de São Paulo – ICMC-USP
Departamento de Ciências da Computação
Porto Velho, RO / São Carlos, SP
carolina@unir.br

Resumo

This paper presents a method based on complex networks to shape boundary characterization of mammograms. A region of interest of mass is automatically segmented using an algorithm based on EM/MPM and the shape is modeled into a small-world complex network. To compose the feature vector, degree measurements are extracted from the network in a dynamic evolution. The experiments comparing the complex networks approach with other traditional descriptors, in detecting breast cancer in mammograms, show that our approach has gotten better accuracy. Hence, the results indicate that complex networks are well-suited to characterize mammograms.

1. Introdução

A classificação automática, no campo médico, é fundamental para diagnósticos auxiliados por computadores [7].

Os seres humanos são naturalmente adaptados a reconhecer objetos por meio de suas características visuais, tais como, tamanho, forma e cor. No entanto, dentre todos os aspectos relacionados à informação visual, a forma de um objeto possui uma grande relevância na classificação de massas em mamografias [5, 20].

Na literatura, alguns descritores de características têm sido propostos para representação de massas em mamografias. Dentre eles estão os descritores tradicionais, como o descritor de textura de Haralik e o descritor de forma de Zernike [10]. Outros baseados em forma também são utilizados, os quais extraem medidas como área, excentricidade, orientação, dentre outras [18]. Em [2] foi demonstrado que as redes complexas possuem uma grande precisão na análise da forma de objetos.

Neste trabalho é apresentada uma abordagem baseada na análise de forma que utiliza redes complexas para a caracterização de imagens de mamografia. Primeiramente a imagem é segmentada automaticamente usando um algoritmo baseado em EM/MPM [3], e depois é limiarizada, preservando a região de interesse (o tumor). Em seguida, é extraída a fronteira do tumor (i.e., o contorno), a qual é modelada como um grafo. Para diminuir o esforço computacional, a quantidade de vértices para representar o contorno é minimizada. A partir deste grafo, é obtida artificialmente, por transformações nos vértices do grafo, a topologia de uma rede Pequeno Mundo [2]. Desta rede complexa modelada são extraídas as medidas de graus, considerando a evolução dinâmica da rede, as quais irão compor o vetor de características. Por fim, a sugestão de diagnóstico em nódulos benignos e malignos é realizada por uma árvore de decisão.

Os experimentos realizados comparam o método proposto com os extratores de Zernike, Haralik, e um baseado em análise de forma proposto em [18]. Para a classificação é utilizado o algoritmo de árvore de decisão. Os resultados mostram que as redes complexas são adequadas para a representação de mamografias na detecção de câncer de mama com níveis de precisão de classificação superiores aos dos outros descritores.

Este artigo está estruturado da seguinte forma. Na seção 2 é brevemente apresentada a teoria das redes complexas e na seção 3 é apresentada a metodologia proposta. A seção 4 apresenta os resultados do experimento utilizando uma base de dados de mamografias, onde o método é comparado com outros extratores de características. Por fim, na seção 5 é apresentada a conclusão.

2. Redes Complexas

O termo redes complexas refere-se a um grafo que apresenta uma topologia não trivial, composto por um conjunto

de vértices que são interligados por meio de arestas [4, 14]. Seu estudo é um tema interdisciplinar que abrange diversas áreas de conhecimento; ciências da computação, matemática, física, biologia e sociologia [1]. O estudo de redes na forma de grafos é um dos pilares da matemática discreta e teve início em 1735, quando Euler propôs uma solução para o problema das pontes de Königsberg, originando a teoria dos grafos [13].

Apesar dos trabalhos pioneiros de Flory [11], Rapoport [16, 17] e Erdös e Rényi [8, 9], apenas mais recentemente este ramo de pesquisa tem recebido uma maior atenção. O avanço da capacidade dos computadores possibilitou uma grande melhoria na área das redes complexas. Antes se usava pequenas quantidades de vértices para análises; atualmente podem-se usar milhões ou até bilhões de vértices. Esta mudança de paradigma revelou várias características que diferem substancialmente as redes do mundo real das redes aleatórias, tidas por muitos anos como o principal modelo de redes complexas [4, 14].

A idéia principal deste trabalho é representar a forma de um tumor nos termos de uma rede de Watts e Strogatz [19] seguida da análise de suas características topológicas e dinâmicas. Este modelo de rede apresenta uma propriedade que é chamada de Pequeno Mundo, na qual todos os vértices podem ser alcançados por qualquer outro através de um número pequeno de vértices. Esse tipo de rede também apresenta um grande número de *loops* de tamanhos três, i.e., quando três vertices possuem ligações entre si.

O modelo dinâmico da rede Pequeno Mundo é obtido artificialmente por meio de limiares sequenciais aplicados nos vértices do grafo obtido da forma modelada. Desta forma, é obtida a topologia de rede Pequeno Mundo em vários estágios do crescimento da rede. O crescimento dinâmico da rede produz uma forma única, a qual pode ser utilizada no processo de análise e classificação da imagem, por meio de extração de medidas que comporão um vetor de características.

2.1. Medidas das Redes Complexas

As redes complexas possuem características topológicas específicas que caracterizam sua conectividade e que influenciam altamente a dinâmica e a função dos processos executados na rede. Nesta seção são descritas as medidas usadas neste trabalho. Mais detalhes sobre medidas em redes complexas podem ser vistos em [6, 14].

O **grau** k_i de um vértice i é o número de arestas conectadas a ele, e é definido através da matriz de adjacências A como

$$k_i = \sum_{j=1}^{N} a_{ij}, \tag{1}$$

onde N é o número de vértices existentes na rede e a_{ij} são elementos da matriz de adjacências A.

Baseado no grau dos vértices é possível extrair outras medidas da rede, tais como, o **grau máximo** (k_κ):

$$k_\kappa = max_i k_i, \tag{2}$$

e o **grau médio** (k_μ):

$$k_\mu = \frac{1}{N} \sum_{j=1}^{N} k_i. \tag{3}$$

A **evolução dinâmica** é uma característica importante das redes complexas, que afeta várias propriedades da rede [6]. Consequentemente, as medidas das redes complexas são em função do tempo, i.e., a extração de medidas da mesma rede em tempos diferentes apresenta valores divergentes. Desta forma, as redes complexas tornam-se um sistema que pode ser monitorado com o passar do tempo. Nesses casos é importante considerar as trajetórias de algumas medidas (e não somente pontos isolados no espaço de características) como um meio de analisar e classificar redes. Este fornece uma caracterização mais ampla da rede sob análise. Além disso, considerando as medidas como uma função do tempo, pode ser mais interessante tentar classificar classes de redes complexas considerando a sua dinâmica no respectivo espaço de características [6].

3. Método Proposto: Segmentação e Modelagem de Mamografias como uma Rede Complexa

Nesta seção é apresentado o método proposto para o auxílio ao diagnóstico na detecção de câncer de mama. A figura 1 mostra o *pipeline* do método proposto. Dada uma base de imagens de regiões de interesse de mamografias, na fase de treinamento é realizada a segmentação e detecção da borda do tumor e, em seguida, é extraído o vetor de características φ usando redes complexas. Esse vetor é usado na fase de construção da árvore de decisão (C4.5), que será usada na fase de sugestão de diagnóstico. Na fase de teste, uma imagem também é submetida ao processo de segmentação, detecção de borda e extração das características e então é classificada usando a árvore de decisão obtida na fase de treinamento. A saída do sistema é uma "segunda opinião" ao especialista da área, ficando a critério deste decidir a veracidade da informação dada pelo sistema. Cada etapa deste sistema está detalhada nas seções seguintes.

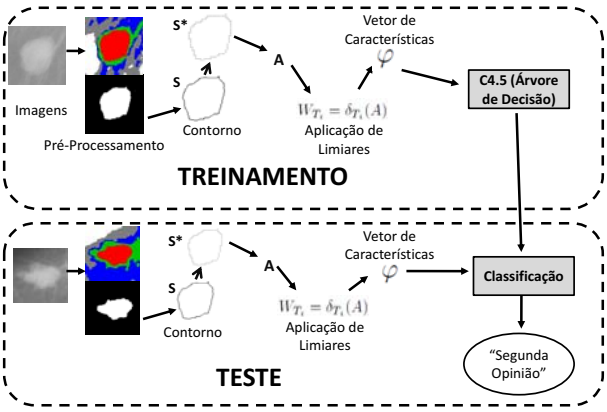
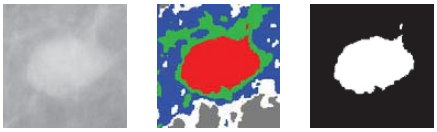


Figura 1. *Pipeline* **do método proposto.**



(a) (b) (c)

Figura 2. (a) Imagem original; (b) Imagem segmentada em 5 regiões; (c) Máscara da região principal.

3.1. Segmentação e Obtenção do Contorno da Borda da Região de Interesse

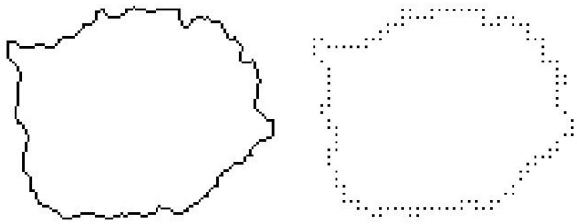
O primeiro passo é a segmentação da imagem a ser analisada. Para isso, foi aplicado o algoritmo EM/MPM (do inglês, *EM - Expectation Maximization; MPM - Maximization of the Posterior Marginals*). Este algoritmo segmenta as imagens usando técnicas que combinam um Campo Aleatório de Markov e um Modelo de Mistura Gaussiana para obter uma segmentação baseada em textura [18]. A segmentação é realizada de acordo com um número fixo de diferentes regiões. Trabalhos usando esta abordagem têm mostrado que usar cinco regiões para segmentar as imagens apresentam resultados satisfatórios [3]. Por esta razão, e por experimentos realizados previamente para o nosso estudo, neste trabalho as imagens foram segmentadas em cinco regiões. Após a segmentação, a região principal (o tumor) é escolhida para a retirada de seu contorno. A figura 2(b) apresenta a imagem de uma mamografia que contém um tumor, a qual está segmentada em cinco regiões e a figura 2(c) a máscara da região selecionada.

O contorno S da massa é obtido através da máscara da imagem (figura 2(c)) quando submetida às funções **boundaries** e **bsubsamp** descritos em [12]. O primeiro extrai S e o segundo extrai um subconjunto S^* de S, de forma que, a quantidade de pontos é menor e que preserva ao máximo a forma original de S.

Esta abordagem permite um menor esforço computacional, como será visto detalhadamente na seção 3.2.

3.2. Representação do Contorno como uma Rede Complexa

Dado S como o contorno de uma forma, onde $S = [s_1, s_2, ..., s_n]$ e s_i são vetores na forma $s_i = [x_i, y_i]$ cujos componentes são valores numéricos discretos representando as coordenadas do ponto i do contorno. Diferentemente do que foi proposto em [2], neste trabalho, alguns pontos de S são retirados gerando o subconjunto S^*, como pode ser visto na figura 3.



(a) S (b) S^*

Figura 3. Quantidade de pontos reduzido; (a) $|S| = 322$**; (b)** $|S^*| = 131$

Com o intuito de aplicar a teoria das redes complexas ao problema, uma representação do contorno S^* é construída como um grafo $G = (V, E)$. Assim $S^* = V$, ou seja, cada ponto de S^* torna-se um vértice do grafo. Portanto, o custo computacional é reduzido, uma vez que seria necessário mais vértices para a representação de S como um grafo que S^*.

O grafo é formado por um conjunto de arestas, sendo que o peso de cada aresta é dado pelo cálculo da Distância Euclidiana entre os dois vértices a serem ligados:

$$d(s_i, s_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \qquad (4)$$

Portanto, a rede é representada pela matriz de adjacências ponderada A de ordem n,

$$a_{ij} = A([a_i, a_j]) = d(s_i, s_j), \qquad (5)$$

normalizada no intervalo $[0, 1]$,

$$A = \frac{A}{max_{a_{ij} \in A}} \tag{6}$$

A princípio, o conjunto de arestas E é composto pelas ligações de todos os vértices da rede um ao outro. Isto significa que a rede possui um comportamento regular, uma vez que todos os seus vértices possuem o mesmo grau. No entanto, uma rede regular não pode ser considerada como uma rede complexa, pois não apresenta propriedades relevantes para a aplicação. Portanto é necessário obter uma rede complexa que contenha propriedades relevantes para a aplicação.

Um modo para obter a rede complexa é aplicando um limiar T_l sobre a rede regular, que produzirá um novo conjunto de arestas e consequentemente se tornará em uma rede complexa, que será representada pela matriz de adjacências W. Este limiar consiste em selecionar um grupo de arestas pertencentes a E que possua valores menores que o limiar T_l. Desta forma, é obtido uma rede complexa (que se encaixa no modelo de rede Pequeno Mundo de Watts e Strongatz [19]) por meio de uma rede inicialmente regular. As redes Pequeno Mundo são caracterizadas por duas propriedades: (i) alto coeficiente de aglomeração e (ii) a propriedade de Pequeno Mundo. O coeficiente de aglomeração é a medida de quão interconectados estão os vértices da rede, e é definido como

$$C = \frac{3N_\Delta}{N_3}, \tag{7}$$

onde N_Δ é o número de triângulos na rede e N_3 o número de triplas conectadas. Uma tripla conectada ocorre quando um vértice i está conectado ao vértice j, e o vértice j está conectado ao vértice k. Um triângulo ocorre quando os vértices j e k também estão conectados.

A propriedade de Pequeno Mundo é devida ao alto coeficiente de aglomeração, onde a maioria dos vértices estão interconectados e a maioria dos caminhos entre quaisquer dois vértices são possíveis e pequenos. Esta propriedade é caracterizada pela existência de caminhos geodésico pequenos, sendo um caminho geodésico o menor caminho existente conectando dois vértices i e j na rede, onde $i \neq j$.

Uma vez que S^* é modelado como uma rede complexa algumas de suas propriedades podem ser quantificadas. Esta tarefa, é executada como descrito nas seções 3.3 e 3.4.

3.3. Evolução Dinâmica

O uso de conceitos e ferramentas fundamentais na pesquisa das redes complexas pode providenciar informações relevantes para a caracterização de objetos e imagens. Com base nisto, neste trabalho foi usada a idéia proposta por [2], mas, ao invés de considerar todo o conjunto de pontos da fronteira do contorno do objeto para compor os vértices da rede, foi considerado apenas o conjunto S^*: dada a representação da forma de um tumor, como foi apresentada na seção 3.2, a respectiva caracterização da imagem deve ser feita, utilizando a representação, por um vetor de característica. Este vetor de características é obtido de diferentes valores de T_l de uma δ transformação. Esta operação, representada como $W = \delta_{T_l}(A)$ é aplicada a cada componente da matriz ponderada A, gerando a matriz não-ponderada W. Assim,

$$W_{T_Q} = \delta_{T_Q}(A) = \forall a \in A \begin{cases} w_{ij} = 0 & \text{se } a_{ij} \geq T_l, \\ w_{ij} = 1 & \text{se } a_{ij} < T_l. \end{cases}$$

Em outras palavras, a caracterização é feita usando várias transformações δ onde o limiar T_l é incrementado em um intervalo regular T_{inc}. Portanto, dado um conjunto T, um elemento $T_l \in T$ é definido pela função $f : T \to T$, onde

$$T_0 = T_{ini} | 0 < T_{ini} < 1$$

$$T_{l+1} = f(T_l) \ \ se \ \ T_l < T_Q < 1,$$

$$f(x) = x + T_{inc}$$

com T_{inc} e T_Q, respectivamente, o limiar inicial e final.

Dadas algumas redes obtidas pela evolução dinâmica, o uso de um vetor de características é proposto para a representação do tumor presente na mamografia. Este vetor de características é formado pelos graus médio e máximo dos vértices extraídos da rede ao longo da evolução dinâmica.

3.4. Descritor de Grau

O descritor de grau, como proposto em [2], é calculado pela matriz transformada W para cada instante T_l usando as medidas de grau máximo (equação 2) e o grau médio (equação 3). Entretanto, uma normalização do grau dos vértices pelo número de vértices na rede é necessária antes de calcular essas medidas [2], visto que para cada imagem, o número total de vértices na rede é diferente. Esta normalização é feita afim de reduzir a influência do tamanho da rede no descritor calculado, como descrito na equação 8:

$$\forall k_i = \frac{k_i}{N}. \tag{8}$$

Considerando a transformação da rede em cada instante de tempo T_l, o vetor φ é formado com a concatenação do grau máximo (k_κ) e o grau médio (k_μ) que são calculados a cada estágio da evolução da rede:

$$\varphi = [k_\kappa(T_0), k_\mu(T_0), k_\kappa(T_1), k_\mu(T_1), ..., k_\kappa(T_Q), k_\mu(T_Q)].$$

A figura 4 mostra o processo para a extração do descritor de grau de W ao longo das transformações executadas por

meio de T_l. O vetor de características apresentado, φ, pode ser utilizado como uma possível representação da forma do contorno do tumor presente na mamografia, este contém propriedades desejáveis para a análise da forma, tais como, invariância à rotação e escala e tolerância a ruídos [2].

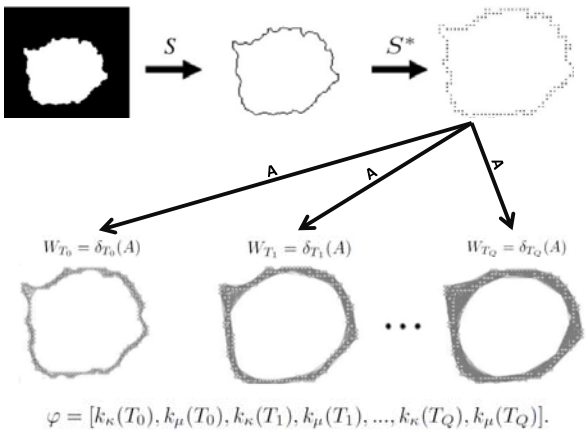


Figura 4. Obtenção de φ.

4. Experimento

Para validar a eficácia do descritor proposto foi realizado um experimento com uma base composta de 250 imagens de mamografias coletadas do *Digital Database for Screening Mammography* - DDSM dataset[1].

No experimento, o vetor de características φ é obtido como descrito na seção 3. O limiar inicial (T_{ini}) foi igual a 0.175, incrementado em um intervalo regular de 0.075 até chegar no valor de limiar final (T_Q) igual a 0.95. A base é composta por imagens de mamografias com tumores, sendo dos casos, 99 benignos e 151 malignos. Os graus máximo e médio são calculados a partir de W após a aplicação de cada limiar, totalizando 22 atributos no vetor de características (φ).

A classificação é feita utilizando o classificador C4.5 [15], este constrói uma árvore de decisão na fase de treinamento. O método é comparado com dois vetores de características tradicionais da literatura, Zernike e Haralik [10], e um vetor de características de forma proposto em [18]. Cada método possui em seu vetor de características respectivamente, 256, 140 e 11 atributos para cada imagem.

Medidas de precisão, sensibilidade e especificidade são calculadas. A precisão é a quantidade de casos que foram corretamente classificados. A sensibilidade é a quantidade de casos positivos que foram corretamente classificados. E a especificidade é a quantidade de casos negativos que foram corretamente classificados. Uma predição ótima pode alcançar 100% de sensibilidade (i.e. predizer todas as imagens do grupo maligno como maligno) e 100% de especificidade (i.e. não predizer nenhuma imagem da classe benigna como maligna). Para calcular essas medidas se faz necessário considerar os casos [18]:

- Verdadeiro Positvo (VP): tumor maligno corretamente identificado como maligno;
- Falso Positivo (FP): tumor benigno incorretamente identificado como maligno;
- Verdadeiro Negativo (VN): tumor benigno corretamente identificado como benigno;
- Falso Negativo (FN): tumor maligno incorretamente identificado como benigno.

As seguintes equações demonstram a fórmula da precisão, sensibilidade e especificidade, respectivamente.

$$\textit{Precisão} = \frac{VP + VN}{VP + VN + FP + FN} \tag{9}$$

$$\textit{Sensibilidade} = \frac{VP}{VP + FN} \tag{10}$$

$$\textit{Especificidade} = \frac{VN}{VN + FP} \tag{11}$$

4.1. Resultados e Discussão

As tabelas 1 e 2 mostram a precisão, a sensibilidade e a especificidade das Redes Complexas em comparação com os vetores de características de Zernike, Haralike e forma, utilizando as abordagens *holdout* e *cross-validation* com *10-fold*, respectivamente.

Métodos	Precisão	Sensibilidade	Especifici.
Redes Comp.	**0.8064**	0.8200	**0.7700**
Zernike	0.6935	0.7714	0.5925
Haralik	0.7096	**0.9714**	0.3703
Forma	0.7580	0.7692	0.7391

Tabela 1. Comparação usando *holdout*

Observando a tabela 1, podemos observar que as Redes Complexas detêm os maiores valores em precisão e especificidade. O método de Haralik possui um alto valor de sensibilidade, no entanto, seu valor de especificidade é muito

1 http://marathon.csee.usf.edu/Mammography/Database.html

Métodos	Precisão	Sensibilidade	Especifici.
Redes Comp.	**0.7440**	**0.8476**	0.5858
Zernike	0.6840	0.7284	**0.6161**
Haralik	0.7280	0.8211	0.5858
Forma	0.7320	0.8150	0.6153

Tabela 2. Comparação usando *cross-validation 10-fold*

baixo, ou seja, apresenta muitos falsos positivos, classificando muitos tumores benignos como malignos.

Observando a tabela 2, nota-se que as Redes complexas detêm os maiores valores em precisão e sensibilidade. A especificidade das Redes Complexas detêm valores próximos aos métodos da forma e Zernike, aproximadamente 0.03% de diferença. As Redes Complexas detêm o mesmo valor de especificidade do método de Haralik. Sendo assim não há uma diferença grande de espeficidade das Redes Complexas com os outros métodos.

5. Conclusão

Neste artigo foi apresentado um método para extração de características de mamografias utilizando a teoria das redes complexas. Foi mostrado como duas medidas, grau máximo e médio, podem ser utilizadas para a identificação de classes. Foi visto a evolução dinâmica da rede com um progressivo incremento do limiar T_l para compor um vetor de características e seu posterior uso em um classificador.

O experimento reportado ilustra que medidas, grau máximo e médio, tem um grande potencial de discriminar classes, uma vez que os resultados do experimento mostraram que o método proposto alcança altos valores de precisão quando comparado com outros métodos conhecidos (Zernike, Haralik, forma). Além do mais o método mostra um balanço entre sensibilidade e especificidade, sendo um pouco mais sensível do que específico, o que é desejável no domínio médico. Desta forma, podemos inferir que as redes complexas são adequadas para caracterização de mamografias no auxílio ao diagnóstico.

Agradecimentos

Os autores agradecem ao CNPq pelo apoio financeiro a esta pesquisa.

Referências

[1] N. J. Alves. Caracterizaçaoo de redes complexas, aplicações: Modelagem relacional entre sistemas autônomos da internet. 2007.

[2] A. R. Backes, D. Casanova, and O. M. Bruno. A complex network-based approach for boundary shape analysis. *Pattern Recognition*, 42(1):54–67, 2008.

[3] A. G. R. Balan, A. J. M. Traina, C. T. Jr., and P. M. Azevedo-Marques. Fractal analysis of image textures for indexing and retrieval by content. In *CBMS*, pages 581–586. IEEE Computer Society, 2005.

[4] A. L. Barabási. How everything is connected to everything else and what it means for business, science and everyday life. *Plume*, 2003.

[5] L. F. d. Costa and R. M. Cesar. *Shape Analysis and Classification: Theory and Practice*. CRC, December 2000.

[6] L. F. da Costa, G. T. Rodrigues, and P. R. V. Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, 2007.

[7] G. Dougherty. *Digital Image Processing for Medical Applications*. Cambridge University Press, 2009.

[8] P. Erdös and A. Rényi. On the evolution of random graphs. *Mathematical Institute of the Hungarian Academy of Science*, 5:17–61, 1960.

[9] P. Erdös and A. Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Hungarica*, 12(1-2):261–267, 1961.

[10] J. C. Felipe and A. J. M. Traina. Utilizando características de textura para identificação de tecidos em imagens médicas. In *2º Workshop de Informática Médica (WIM) 2001 - in CD-ROM*, pages in CD–ROM, Gramado, RS, 2002.

[11] P. J. Flory. Molecular size distribution in three-dimensional polymers. *Journal of the American Chemical Society*, 63:3083–3090, 1941.

[12] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital image processing using MATLAB*. Prentice Hall, 2003.

[13] J. Metz, R. Calvo, E. R. M. Seno, R. A. F. Romero, and Z. Liang. Redes complexas: conceitos e aplicações. Technical Report 290, Instituto de Ciências Matemáticas e Computação - ICMC-USP, Janeiro 2007.

[14] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, (45):167–256, 2003.

[15] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

[16] A. Rapoport. Spread of information through a population with sociostructural bias: I. assumption of transitivity. *Bulletin of Mathematical Biophysics*, 15:523–533, 1953.

[17] A. Rapoport. Contribution to the theory of random and biased nets. *Bulletin of Mathematical Biophysics*, 19(257-277), 1957.

[18] C. Y. V. Watanabe, M. X. Ribeiro, C. T. Jr., and A. J. M. Traina. Statistical associative classification of mammograms - the sacminer method. In J. Filipe and J. Cordeiro, editors, *ICEIS (2)*, pages 121–128. SciTePress, 2010.

[19] D. J. Watts and S. H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393:440–442, 1998.

[20] C.-H. Wei, C.-T. Li, and Y. Li. *Content-based retrieval for mammograms*. Hershey, 2008.

Segmentação e Restauração Digital para Eliminação de Artefatos em Imagens Frontais de Face

André Sobiecki[1], Luiz A. P. Neves[2], Gilson A. Giraldi[3], Gilka J. F. Gattás[4] e Carlos E. Thomaz[1]

[1]FEI, São Bernardo do Campo, São Paulo, Brasil
[2]UFPR, Curitiba, Paraná, Brasil
[3]LNCC, Petrópolis, Rio de Janeiro, Brasil
[4]USP, São Paulo, São Paulo, Brasil

Abstract

There is a variety of applications nowadays that has used the face image as a relevant information for recognizing subjects. However, depending on the application, such as in the problem of identifying missing people based on paper archiving, the face images can become noisy for digitalization due to its low original resolution, poor quality of the photographic paper and the presence of other artifacts like folds and stamps. In practice, all these artifacts make the process of automatically recognizing faces even harder. Therefore, we propose in this work a computational framework for segmentation and automatic restoration of image artifacts based on a statistical model built from 200 frontal faces with neutral expression. The effectiveness of the framework has been compared with other methods and shows promising results.

1. Introdução

O reconhecimento facial é o processo de identificação de uma pessoa com base em informações obtidas a partir da aparência facial [3]. O desempenho dos algoritmos de reconhecimento facial depende da qualidade do pré-processamento da imagem do rosto, da precisão da representação facial e da eficiência do classificador [7]. Por isso, é de fundamental importância que a imagem a ser utilizada em um trabalho de reconhecimento facial apresente apenas as características faciais discriminantes com o mínimo de artefatos possível.

Há aplicações que utilizam a face como informação relevante para identificação de pessoas. Por exemplo, o *site* do governo federal brasileiro de crianças e adolescentes desaparecidos [9] disponibiliza publicamente imagens faciais de pessoas sendo que muitas dessas fotos são antigas e possuem rasuras como dobraduras, arranhões, luminância irregular, bolor, carimbos e escritos diversos sobre a imagem digitalizada. Na prática, essas rasuras dificultam o reconhecimento facial automático. Uma rasura sobre uma imagem facial pode estar em diversas cores, tons, formatos e tamanhos. Informações prévias sobre a localização de características faciais como nariz, olhos, boca e cabelo auxiliam na identificação de rasuras.

Este trabalho propõe um arcabouço computacional de segmentação e restauração automática de rasuras baseado em um modelo estatístico construído a partir de 200 imagens faciais frontais. A localização das características faciais é feita por meio de uma imagem de referência gerada a partir de uma população amostral que fornece informações a *priori* sobre tons e localização espacial das características. As rasuras identificadas na imagem são posteriormente restauradas por um método de *inpainting* [13, 15]. Os experimentos deste estudo mostram que a segmentação baseada no método estatístico apresenta melhores resultados de correlação com a imagem de referência do que outros métodos implementados para comparação, como os métodos baseados em entropia e no filtro de Sobel.

2. Arcabouço Computacional

Um modelo deve cumprir os critérios de generalidade e especificidade, onde modelos gerais são aqueles que representam as fontes de variação na aparência de imagens faciais e os modelos específicos são aqueles que limitam a variabilidade permitida de modo que apenas exemplos que cumpram os devidos critérios possam ser gerados [6].

A Figura 1 representa a seqüência de métodos proposta neste artigo para identificação e eliminação de rasuras automaticamente.

Figura 1: Arcabouço computacional de eliminação de rasuras.

Todas as imagens foram antes equalizadas [2] e pré-processadas utilizando-se um filtro 3x3 de mediana para eliminar ruídos. O filtro de mediana ordena os pixel em ordem crescente e escolhe o valor mediano para igualar aos demais pixels. As etapas de identificação de artefatos, aplicação do molde e restauração digital são descritas em detalhes nas seções seguintes.

3. Identificação de Artefatos

Para identificação de artefatos, são usados três métodos de segmentação automática: o método de decisão estatística proposto neste trabalho; e mais dois métodos para fazer comparações baseados em entropia [17] e no filtro de Sobel. O filtro de Sobel é bem conhecido na literatura e não será descrito. O leitor pode consultar, por exemplo, as referências [5, 11] para mais detalhes sobre esse filtro.

3.1. Decisão Estatística

O método estatístico é baseado na teoria de inferência estatística [8, 10], onde amostras populacionais podem gerar informações a *priori* e, a partir destas informações, é possível tomar decisões dado um nível α de significância estatística desejado.

A partir de uma população amostral de 200 imagens frontais com expressão facial neutra e normalizadas espacialmente [2], calcula-se a imagem média dessas amostras através da seguinte equação:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \ , \tag{1}$$

onde x_i é o vetor n-dimensional que representa a concatenação de todos os pixels da imagem i de face, e N o número total de amostras. Neste trabalho, $N = 200$.

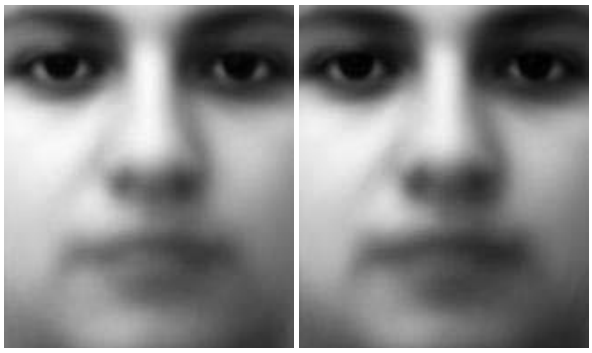
Para verificar quão válida seria a afirmação de que a média amostral descrita pela equação (1) se aproximaria da média da população de faces frontais com expressão facial neutra, calculou-se o intervalo de 99% de confiança (IC) dessa estimativa, ou seja,

$$99\%IC = \overline{x} \pm 2.58\frac{s}{\sqrt{N}} \ , \tag{2}$$

onde s é o desvio padrão das amostras, cuja variância s^2 pode ser definida como:

$$s^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(x_i - \overline{x})^2 \ . \tag{3}$$

A Figura 2 ilustra as imagens que correspondem aos limites inferior e superior do intervalo de confiança descritos na equação (2). Observa-se que visualmente as imagens são muito parecidas, pois possuem correlação de 0,998942 e 0,998962 com a imagem média amostral.



(a) (b)

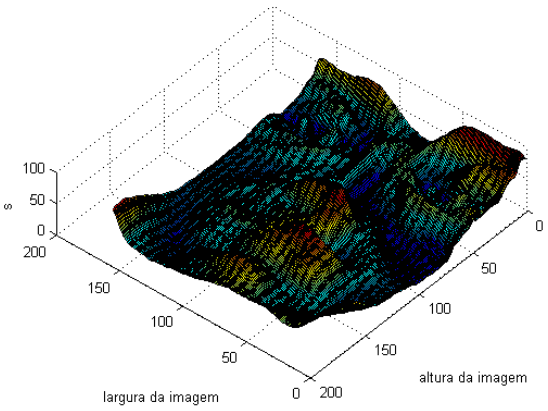
Figura 2: Intervalo de confiança da imagem média: (a) Limite inferior; (b) Limite superior.

Portanto, considerando estatisticamente válida a estimativa amostral da média, a decisão sobre a identificação de artefatos está baseada simplesmente na definição da região crítica das diferenças significantes. Em outras palavras, calcula-se o valor z da diferença entre uma imagem de face x e a imagem da média amostral $\overline{x}$, supondo que a distribuição das tonalidades dos pixels segue uma densidade de probabilidade Gaussiana com média nula e variância 1, isto é:

$$z = \frac{x - \overline{x}}{s} \ . \tag{4}$$

Quanto maior for o valor absoluto de z maior será a significância estatística dessa diferença, pixel a pixel. A Figura 3 mostra a imagem média e o desvio padrão de todos os pixels utilizados na equação (4) para a identificação das rasuras. Para os experimentos desse trabalho, o valor do nível de significância α foi fixado em 5%, que equivale a um valor z padrão de 1,96 [8].

(a)



(b)

Figura 3: Medidas estatísticas de referência: (a) Imagem da média amostral; (b) Desvio padrão de todos os pixels.

3.2. Entropia

Em processamento de imagens, uma forma de distinguir o plano de fundo da imagem do objeto de interesse, é através da aplicação de um limiar. Uma forma automática para encontrar este limiar é via a otimização da entropia, a qual fornece uma medida da quantidade de informação na imagem. Neste trabalho, considera-se duas definições para entropia: entropia de Shannon e de Tsallis [17] .

Seja então um sistema com n estados possíveis, aos quais estão associadas probabilidades (p_1, ..., p_n), onde $p_i \in [0,1]$ para todos $i = 1, ..., n$ e $\sum p_i = 1$. Shannon mediu a quantidade de informação do sistema pela expressão [16, 17, 18, 19, 20]:

$$S = -\sum_i p_i \ln(p_i) \,. \tag{5}$$

Por outro lado, Tsallis propôs a seguinte medida de informação:

$$S_q(p_1, \ldots p_k) = \frac{1 - \sum_{i=1}^{k} p_i^q}{q-1} \,. \tag{6}$$

p_i é a probabilidade de ocorrência do nível de cinza i (frequencia relativa) na imagem ($0 \le i \le 255$). Cada limiar t divide a imagem e, consequentemente, o histograma da mesma em duas regiões, por exemplo A e B: $p_1, p_2, ..., p_t$ e $p_{t+1}, p_{t+2}, ..., p_{255}$. Assim, pode-se normalizar as distribuições resultantes para obter as probabilidades associadas às regiões A e B e, consequentemente, computar as entropias S^A e S^B pela expressão (5) ou S_q^A e S_q^B pela expressão (6). Pode-se mostrar que [21]:

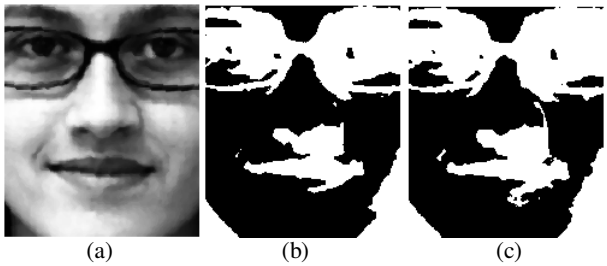
$$S^{A+B}(t) = S^A + S^B \,, \tag{7}$$

$$S_q^{A+B}(t) = S_q^A + S_q^B + (1-q) S_q^A S_q^B \,, \tag{8}$$

onde A+B representa o sistema (imagem) composto pelos subsistemas A e B. A idéia básica da aplicação do conceito de entropia para segmentacao é que o limiar ótimo é aquele que maximiza S^{A+B} (ou S_q^{A+B}), ou seja:

$$T = \max(S^{A+B}(t)) \,. \tag{9}$$

Para entropia de Tsallis, o parâmetro q é mais um grau de liberdade na teoria que foi determinado experimentalmente através de um teste em uma imagem com o q variando de 0.1 á 0.9. A Figura 4 mostra um exemplo de segmentação por esses dois métodos de entropia descritos brevemente nesta sub-seção.



(a) (b) (c)

Figura 4: Segmentação por entropia: (a) Imagem original; (b) Imagem segmentada pela entropia de Shannon; (c) Imagem segmentada pela entropia de Tsallis.

3.3. Morfologia Matemática

Para corrigir as imperfeições das segmentações baseadas na decisão estatística e nos métodos de entropia, finaliza-se a etapa de identificação de artefatos do arcabouço computacional com a utilização de operadores de morfologia matemática.

Basicamente, a morfologia matemática é um método para extração de regiões de interesse em uma imagem (binária ou em tons de cinza), tornando possível representar e descrever melhor a forma de uma região através da detecção e suavização das suas respectivas bordas [1, 4, 11].

Foi suficiente a aplicação do operador morfológico de dilatação [1] com elemento estruturante 3x3 em uma única iteração para recuperar ou aumentar o tamanho da rasura e garantir que o método de restauração digital considerasse da forma mais abrangente possível a região rasurada existente.

4. Aplicação do Molde

Como há regiões faciais que podem perder suas características naturais devido ao processo de restauração digital adotado, aplica-se um molde binário para cobrir as regiões dos olhos, do nariz e da boca, conforme Figura 5. Os pixels ativos (brancos) representam as regiões a serem desconsideradas e os pixels pretos representam as regiões que podem sofrer restauração.

Figura 5: Molde usado para desconsiderar regiões que devem manter a sua originalidade.

A criação deste molde foi possível porque as imagens usadas neste trabalho são normalizadas espacialmente [2], então tem-se como informação a *priori* a localização dos olhos, do nariz e da boca.

5. Restauração Digital

Após a segmentação automática das rasuras, é utilizado um método de restauração digital conhecido como *inpainting* [13, 15], onde a reconstituição de um pixel é baseada na assimilação das informações contidas nos seus pixels vizinhos.

Basicamente, o *inpainting* digital pode ser entendido como uma técnica de restauração de imagens ou vídeos danificados por meio de interpolação de imagens. A técnica pode ser utilizada na restauração de fotos, em efeitos especiais, remoção de objetos, recuperação de margens e difusões de curvatura orientada [14]. O método *inpainting* busca reconstruir as características básicas da imagem e preencher dados desconhecidos em uma região de interesse [13]. Neste trabalho, essa técnica utiliza a imagem molde ilustrada na Figura 5 para determinar a localização espacial das regiões faciais adequadas para processamento do *inpainting*.

Matematicamente, pode-se definir esse processamento numérico de propagação de informações na região de *inpainting* Ω pela seguinte equação [15, 22]:

$$I^{t+1}(i,j) = I^{t}(i,j) + \Delta_k I_k^t(i,j), \quad \forall (i,j) \in \Omega \quad (10)$$

onde $I^t(i,j)$ representa cada pixel da imagem na região de *inpainting* Ω, na iteração t, com $I_k^t(i,j)$ representando as atualizações dos pixels de saída para uma taxa de melhoramento Δ_k. Ou seja, a cada passo a imagem $I^t(i,j)$ é processada, gerando uma versão melhorada desta em $I^{t+1}(i,j)$ [22].

6. Experimentos

Foram testadas 30 imagens de 3 bases distintas (10 imagens por base): (a) Imagens de trabalhos anteriores [2, 23, 24]; (b) Imagens disponibilizadas publicamente pelo *site* de crianças e adolescentes desaparecidos no Brasil [9]; (c) Imagens de voluntários que concordaram em tirar uma foto ou em emprestar uma imagem de face para digitalização [23, 24]. Há rasuras como óculos, cabelo sobre a face, carimbos e arranhões.

Para avaliar quantitativamente os resultados, calculou-se o valor da correlação das 30 imagens originais e restauradas com a imagem da média amostral ilustrada anteriormente na Figura 3(a):

$$\rho_{ij} = \frac{\sum_{k=1}^{n}(i_k - \bar{i})(j_k - \bar{j})}{\sqrt{\sum_{k=1}^{n}(i_k - i)^2}\sqrt{\sum_{k=1}^{n}(j_k - \bar{j})^2}}, \quad (11)$$

onde i e j são imagens, e i_k e j_k representam o pixel k dessas imagens e $\bar{i}$ e $\bar{j}$ são os pixels médios de cada imagem. O n tem o valor correspondente a largura de 162 multiplicada pela altura de 193 de cada imagem, que no caso desse trabalho é igual então a 31266. Para determinar o parâmetro q da entropia de Tsallis, variou-se esse parâmetro de 0.1 até 0.9 e obteve-se o melhor valor de correlação para $q = 0.3$.

Foi utilizado o método da correlação para avaliar os resultados ao invés do método de pico do sinal de ruído (PSNR – Peak Signal to Noise Ratio) que é um método de avaliação de restauração de imagem mais comum, porque neste trabalho não há somente ruídos que afetam a qualidade de nitidez e resolução da imagem [25]. Imagens de face contendo artefatos como carimbos, mesmo estando em boa qualidade digital, não estão em condições de serem interpretadas por sistemas de reconhecimento facial.

7. Resultados

A Figura 6 descreve os resultados de correlação das imagens originais e restauradas com a imagem de referência amostral. Pode-se observar que a maioria das imagens restauradas pelo método estatístico teve resultados de correlação maior com o padrão frontal de faces especificado do que as imagens originais, restauradas pelos métodos de entropia e pelo filtro Sobel.

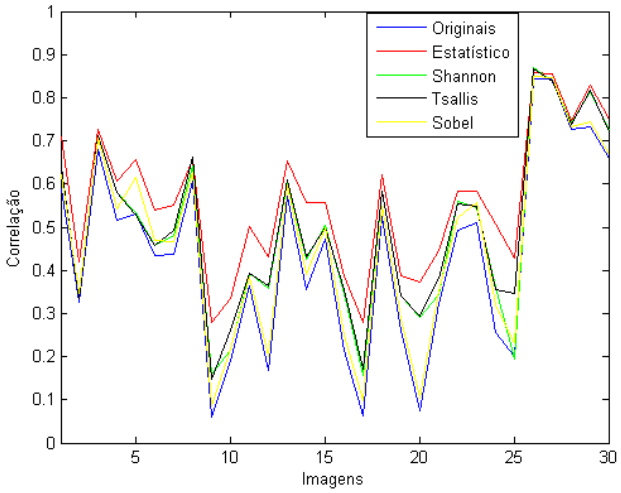


Figura 6: Correlação das 30 imagens.

As médias aritméticas dessas correlações foram: 0.4345 (originais); 0.5606 (método estatístico); 0.4915 (entropia de Shannon) e 0.5007 (entropia de Tsallis); e 0.4627 (Sobel).

A Figura 7 mostra alguns exemplos de restauração digital realizados. Pode-se perceber que o processo de restauração digital baseado no método estatístico de segmentação permitiu uma eliminação de artefatos das imagens mais conservadora e possivelmente mais realista do que os demais métodos testados.

Figura 7: Alguns resultados de restauração digital: (a) imagens originais, (b) imagens segmentadas pelo método estatístico, (c) imagens segmentadas pela entropia de Shannon, (d) imagens segmentadas pela entropia de Tsallis, (e) imagens segmentadas pelo método Sobel.

8. Conclusão

Imagens segmentadas pelo modelo estatístico obtiveram valores maiores de correlação com a imagem de referência, considerando apenas faces frontais e de expressão facial neutra, do que as imagens originais, segmentadas pela entropia e segmentadas pelo método Sobel. No modelo estatístico, cada imagem com artefatos foi comparada com a imagem de referência para um nível de significância estatístico equivalente a 95% de certeza com relação aos defeitos na imagem. Para resultados menos conservadores, pode-se assumir valores de confiança estatística menores. No entanto, nesta situação mais regiões serão necessariamente identificadas como críticas ou, no caso, como rasuras, podendo descaracterizar a singularidade de cada imagem de face.

Devido à diversidade biométrica facial da população brasileira e considerando os resultados promissores deste estudo, pode-se propor como trabalhos futuros a criação de modelos estatísticos de face que sejam mais específicos para as diferenças de raça, idade e gênero inerentes, contribuindo principalmente para uma melhor segmentação automática de rasuras em imagens frontais de face arquivadas em papel como, por exemplo, no problema de identificação de pessoas desaparecidas. Adicionalmente, seria válido a implementação de um método que caracterizasse as regiões faciais que devem manter a originalidade como a boca, os olhos e o nariz, para os casos onde a restauração digital é feita a partir de um método de *inpainting*. As entropias de Shannon e Tsallis são métodos que medem a quantidade de informação probabilística. Através destas entropias é possível segmentar automaticamente regiões de interesse, sendo mais eficiente do que a utilização de um molde construído manualmente.

Referências

[1] E. Aptoula, S. Lefèvre, “A comparative study on multivariate mathematical morphology”, *Pattern Recognition*, *Volume 40*, November 2007, pp. 2914-2929.

[2] V. Amaral, C. Figaro-Garcia, G. J. F. Gattas e C. E. Thomaz. Normalizacao espacial de imagens frontais de face em ambientes controlados e nao-controlados, Periodico Cientifico Eletronico da FATEC Sao Caetano do Sul (FaSCi-Tech), vol. 1, no. 1, Outubro 2009.

[3] O. Ayinde e Y.H. Yang, “Face recognition approach based on rank correlation of Gabor-filtered images”, *Pattern Recognition*, *Volume 35*, 2002, pp. 1275-1289.

[4] J.D.L. Santiago, A. Gamino, J. Salgado, V. Trujilo e A. Ortiz, "Operadores k-estadísticos para morfología matemática de conjuntos", *Rev.fac.ing.univ. Antioquia* [online] n.48, ISSN 0120-6230, 2009, pp. 216-227.

[5] C. Pradabpet, N. Ravinu,l S. Chivapreecha, B. Knobnob e K. Dejhan, "An Efficient Filter Structure for Multiplierless Sobel Edge Detection", *Innovative Technologies in Intelligent Systems and Industrial Applications*, 2009, pp. 40-44.

[6] G. J. Edwards, A. Lanitis, C. J. Taylor e T. F. Cootes, "Statistical models of face images — improving specificity", *Image and Vision Computing*, *Volume 16, Issue 3*,, 16 de marco de 1998, pp. 203-211.

[7] B. Jun, J. Lee, e D. Kim, "A novel illumination-robust face recognition using statistical and non-statistical method ", *Pattern Recognition Letters*, *30 September 2010*, pp. 1-8.

[8] M.R. Spiegel, L.R. Stephens, Statistics, Schaum's Outlines, Lisboa, 2008.

[9] Site de "Crianças e Adolescentes Desaparecidos" da Secretaria Especial dos Direitos Humanos da Presidência da República do governo brasileiro: (http://www.desaparecidos.mj.gov.br/Desaparecidos/)

[10] W.O. Bussab, P.A. Morettin, Estátistica Básica, Editora Saraiva, 5ª edição, São Paulo, 2002.

[11] R.C. Gonzalez, R.E. Woods, Processamento de Imagens Digitais, Editora Edgard Blücher Ltda, São Paulo, 2000.

[13] C.A.Z. Barcelos e M.A. Batista; Image restoration using digital inpainting and noise removal; Image and Vision Computing, *Volume 25, Issue 1*, *January 2007*, *Pages 61-69*

[14] T.K. Shih e R.C. Chang; Digital inpainting – survey and multilayer image inpainting algorithms; Information Technology and Applications, ICITA 2005, Page(s): 15 - 24 vol.1.

[15] M. Bertalmio, G. Sapiro, V. Caselles e C. Ballester; Image Inpainting; Proc. of SIGGRAPH 2000, Page(s): 417 - 424.

[16] M. Portes de Albuquerque, I.A. Esquef e A.R. Gesualdi Mello "Image Thresholding using Tsallis entropy", *Pattern Recognition Letters*, Science Direct, 2004, pp. 1-7.

[17] I.A. Esquef, "Técnicas de Entropia em Processamento de Imagens", Tese de Mestrado, Rio de Janeiro, 2002, pp. 1-155.

[18] S. Tong, A. Bezerianos, Y. Zhu, R. Geocadin, D. Hanley, e N. Thakor, "Monitoring Brain Injury with Tsallis Entropy", IEEE 23º Annual EMBS International Conference, Instanbul, Turkia, 2001, pp. 1-3.

[19] K. Tadaki, "The Tsallis entropy and the Shannon entropy of a universal propability", ISIT, Canadá, 6-11 de Julho 2008, pp. 1-5.

[20] M. Dobre, e L. Bolle, "Tsallis Generalized Entropy and Maximum Entropy Formalism for Spray Drops Size Distribution Function Prediction", *Journal*, ILASS, Europa, 2001, pp. 1-1.

[21] W. Shitong, e F.L. Chung, "Note on the equivalence relationship between Renyi-entropy based and Tsallis-entropy based image thresholding", *Pattern Recognition Letters*, Elsevier, 2005, pp. 1-4.

[22] U.A. Ignácio, "Aplicação de Wavelets em Inpainting Digital", *Dissertação de Mestrado em Computação Aplicada*, 2007, pp. 1-89.

[23] A. Sobiecki, C. E. Thomaz e L. A. P. Neves. Uma Proposta de Melhoria Automática em Imagens Faciais. Proceedings of the V Workshop de Visao Computacional, 5pp., Universidade Presbiteriana Mackenzie, Sao Paulo, Brazil, 2nd - 4th September 2009.

[24] A. Sobiecki, C. E. Thomaz and L. A. P. Neves. To a better digitalization and visualization of frontal face photographs. In proceedings of IV European Conference on Computational Mechanics, Image Processing and Visualization Symposium, id. 1131, p. 32, Paris, France, 16th-21st May 2010.

[25] A.N. Zamani, M.K. Awang, N. Omar e S.A. Nazeer. Image Quality Assessments and Restoration for Face Detection and Recognition System Images. Second Asia International Conference on Modelling & Simulation, 2008.

Hand Edge- Detecção da Borda da Mão pela Diferença de Luminosidade

André Luis Borsato Sanchez[1], Lucas Antonio Alves de Godoi[2], Luis Fernando Degrandi[3], Osvaldo Severino Jr[4]
Faculdade de Tecnologia de Catanduva
1-Andre.lbs.fatec@hotmail.com , 2-Lucas.godoi@fatec.sp.gov.br , 3-Luis.degrandi@fatec.sp.gov.br , 4-Osvaldo.junior2@fatec.sp.gov.br

Abstract

This paper presents a new hand edge detection method. This new methodology uses the difference in the brightness of the grayscale images obtained by the model HSM and the RGB model. The difference in brightness occurs because the RGB model using an arithmetic average to obtain the gray level while the HSM model using a weighted average. The methodology, distinguished by its simplicity, makes it easy to implement and capture the edges. The methodology was tested and showed a high performance when compared to traditional edge detectors such as Roberts, Prewitt and Sobel. Thus, this new proposal becomes a valuable tool for computer vision for feature extraction in computer vision.

.

1. Introdução

A Visão Computacional envolve a identificação e a classificação de objetos em uma imagem, desta forma, é definida pela automatização e a integração de processos e representações usadas para a percepção visual. Os processos da Visão Computacional abrangem, em geral, uma visão em baixo nível representada pelo pré-processamento da imagem, uma visão em nível intermediário representada pela extração de características e uma visão em alto nível representada pelo reconhecimento e interpretação de objetos.

A detecção de bordas é uma técnica da visão computacional utilizada para a identificação de pontos de descontinuidade de uma imagem digital. A descontinuidade é representada pela alteração repentina no nível de luminosidade dos pixels de uma imagem em nível de cinza. A detecção de borda é bastante usada na área de extração de características. Verifica-se, também, que a sua utilização reduz significativamente a quantidade de dados a serem processados, mas preserva importantes propriedades estruturais de uma imagem. Entretando, a detecção pode ser prejudicada por bordas falsas criadas por ruídos (pontos indesejáveis) na imagem provenientes da digitalização ou processo de captura da imagem.

A maioria dos algoritmos de detecção de bordas utiliza uma máscara sobre os pixels da imagem para a verificação de uma descontinuidade. Cada pixel e seus pixels vizinhos têm o seu valor de nível de cinza multiplicado por uma constante. A soma destes valores representa a máscara de resposta para aquele ponto.

Este trabalho tem como objetivo apresentar um novo método de detecção de bordas baseado na diferença de luminosidade do nível de cinza entre uma imagem obtida pelo modelo de cor RGB e uma imagem obtida pelo modelo cor HSM [1] da palma da mão denominado Hand Edge.

Os resultados obtidos demonstram o potencial dessa nova metodologia em relação as tradiconais quanto a velocidade de processamento e a detecção de bordas.

O artigo está estruturado da seguinte forma: na próxima seção serão apresentadas as características dos modelos de cores utilizados; na seção detecção de bordas serão discutidos os métodos tradicionais aplicados nas literaturas atuais, na seção Metodologia será mostrado o processo de aquisição das imagens e a aplicação da técnica proposta; na seção seguinte serão analisados os resultados obtidos e por último será abordado a conclusão sobre a metodologia apresentada.

2. Modelos de Cores

Baseando-se na Teoria dos Três Estímulos, é imediato verificar que qualquer modelo de cor deverá possuir pelo menos três parâmetros. O problema essencial da modelagem de cor está na definição de que grandezas devem ser associadas a cada um dos três parâmetros. Esta escolha origina a existência de vários sistemas de cor sendo influenciados pela natureza dos dispositivos gráficos.

2.1. Modelo de cor RGB

O modelo RGB (*Red Green Blue*) [2] é um modelo de cor concebido com base nos dispositivos gráficos com as três cores primárias: vermelho, verde e azul. Este modelo descreve as cores como o resultado da adição das três cores primárias, cada uma delas com uma intensidade que pode variar entre 0 e 1. O valor 1 corresponde à intensidade máxima com que a cor pode ser apresentada no dispositivo gráfico e o valor 0 à intensidade mínima. A

diagonal principal do cubo, que tem a contribuição de forma igual das três cores primárias, forma os tons de cinza calculados pela Equação 1. As imagens provenientes das câmeras fotográficas digitais têm este modelo de cor como padrão.

$$Cinza = \frac{R+G+B}{3} \quad (1)$$

2.2. Modelo de cor HSM

O modelo de cor HSM (*Hue Saturation Mixture*), definido pelo autor, é mais intuitivo que o modelo RGB. Com estes parâmetros, o modelo HSM aproxima-se muito do modelo que emprega os conceitos qualitativos de matiz, luz e tonalidade (Figura 1). A matiz (Hue) é definida por um ângulo entre 0 (H=0) e 360 (H=1) graus. A saturação (Saturation) é a profundidade ou pureza da cor e tem como medida a sua distância radial em relação ao eixo central com valores normalizados entre 0 e 1. O parâmetro M (Mixture), calculado pela Equação 2, corresponde à intensidade da mistura de cores e varia entre intensidade nula 0, ou seja, cor negra e intensidade máxima 1, ou seja, cor branca. Se a saturação estiver em zero (S=0), qualquer valor ao longo do eixo central (intensidade da mistura de cores) irá do preto (M=0) ao branco (M=1) passando pelos tons de cinza. Para qualquer valor de intensidade da mistura de cores e de matiz o valor da saturação varia do seu mínimo para o máximo e a mudança visualizada sai de uma cor mais escura (com mais preto) e vai para a forma mais pura da cor representada pelo valor da matiz (Figura 1).

$$M = \frac{4R+2G+B}{7} \quad (2)$$

Figura 1. Modelo de cor HSM.

.

3. Operadores para a Detecção de Bordas

Uma borda é definida por uma mudança repentina do nível de cinza, ou seja, ocorre uma descontinuidade na intensidade. Dessa forma, a descontinuidade pode ser percebida quando a gradiente da imagem tem uma variação abrupta.

Um operador de derivada permite verificar a percepção dessa alteração. Uma interpretação de uma derivada em imagens em nível de cinza reflete a taxa de mudança dos níveis de cinza da imagem [3]. Essa taxa é maior quando ocorre perto das bordas e menor em áreas constantes da imagem. Verificando os valores de intensidade da imagem e encontrando os pontos onde a derivada é um ponto de máximo, detecta-se as suas bordas.

Dado que as imagens possuem duas dimensões (comprimento e largura), é importante considerar mudanças nos níveis de cinza nas duas direções (horizontal e vertical). Por esta razão, derivadas parciais das imagens são usadas nas direções x (horizontal) e y (vertical). Uma estimativa da direção atual da borda pode ser obtida usando as derivadas x e y como os componentes da direção ao longo dos eixos (horizontal e vertical), e calcular o vetor soma. Uma imagem é uma função de duas variáveis I(x,y), então o gradiente é definido como:

$$\nabla I(x,y) = \begin{bmatrix} \frac{\partial I(x,y)}{\partial x} \\ \frac{\partial I(x,y)}{\partial y} \end{bmatrix} \quad (3)$$

e a magnitude do gradiente é dada por:

$$|\nabla I(x,y)| = \sqrt{\left(\frac{\partial I(x,y)}{\partial x}\right)^2 + \left(\frac{\partial I(x,y)}{\partial y}\right)^2} \quad (4)$$

Para determinar se um pixel é ou não um pixel de borda, a magnitude do gradiente é calculada. Se a magnitude calculada é maior do que o menor valor de entrada (definido de acordo com a natureza e qualidade da imagem que esta sendo processada), o pixel é considerado ser parte de uma borda.

A derivada x é calculada através de uma máscara (Equação 5), assim como também, a derivada y.

$$\partial x = \begin{bmatrix} m1 & m2 & m3 \\ m4 & m5 & m6 \\ m7 & m8 & m9 \end{bmatrix} \quad (5)$$

$$\frac{\partial I(x,y)}{\partial x} = \sum_{i=1}^{9} m_i p_i$$

Os detectores de bordas mais utilizados na literatura [3] são: Roberts (Equação 6), Prewitt (Equação 7) e Sobel (Equação 8).

$$\partial x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \; e \; \partial y = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (6)$$

$$\partial x = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \; e \; \partial y = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad (7)$$

$$\partial x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \; e \; \partial y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (8)$$

4. Metodologia

A detecção de borda da palma da mão denominada Hand Edge proposta por trabalho é realizada por três etapas: aquisição da imagem, obtenção da imagem diferença e detecção efetiva da borda. Para o desenvolvimento do Hand Edge foi utilizado o software MatLab versão 7 R14 na plataforma Windows XP.

4.1. Aquisição da imagem

Para a aquisição da imagem da palma da mão de uma pessoa, foi construída uma caixa de chapa de metal 0.5 mm pintada internamente de branco e encapada externamente com couro preto. A caixa possui uma base de 27x30 cm e 32 cm de altura com uma abertura na face frontal (Figura 2). A pessoa introduz a palma da mão pela abertura frontal da caixa e posiciona-a sobre a base que também foi encapada internamente com couro preto. A imagem é capturada pelo webcam LifeCam NX-6000 da Microsoft (resolução de 7.6 megapixels para fotografias com uma lente angular grande com campo de 71º) posicionada internamente no centro da tampa da caixa. Para obter uma melhor qualidade da imagem capturada, foram utilizadas quatro mini lâmpadas incandescentes, posicionadas ao redor da câmera, para garantir uma luminosidade homogênea em todos os lados da caixa, sem a formação de sombra e brilho causados pelo reflexo do flash da câmera.

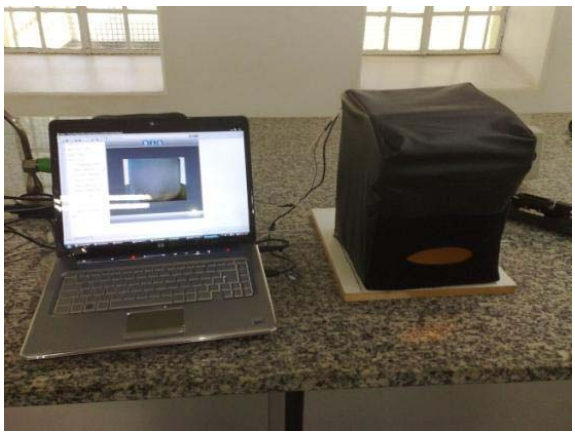


Figura 2. Dispositivo criado para a aquisição da imagem.

4.2. Obtenção da Imagem Diferença

Realizada a captura da imagem da palma da mão, a imagem foi convertida em nível de cinza no modelo RGB pela Equação 1. A mesma imagem foi convertida em nível de cinza no modelo HSM pela Equação 2. Na Figura 3, nota-se uma diferença de luminosidade na imagem em nível de cinza no modelo RGB e no modelo HSM. A imagem convertida no modelo HSM apresentava uma maior luminosidade, ou seja, possui um nível de cinza maior quando comparada com a mesma imagem obtida pelo modelo RGB. Verificou-se que a diferença entre a imagem obtida pelo HSM e a imagem obtida pelo RGB segmentava a palma da mão (Figura 3-b).

4.3. Detecção de Borda pela Hand Edge

Identificou-se que os pontos da imagem diferença que apresentavam nível de cinza igual a 5 (nível de cinza da imagem obtida pelo HSM – nível de cinza da imagem obtida pelo RGB = 5) (Equação 9) definiam a borda da palma da mão. Dessa forma, definiu-se um novo método de detecção de borda denominado Hand Edge que utiliza a diferença de luminosidade em nível de cinza entre as imagens obtida pelo HSM e pelo RGB [4] (Figura 4-b).

$$M - Cinza = 5 \quad (9)$$

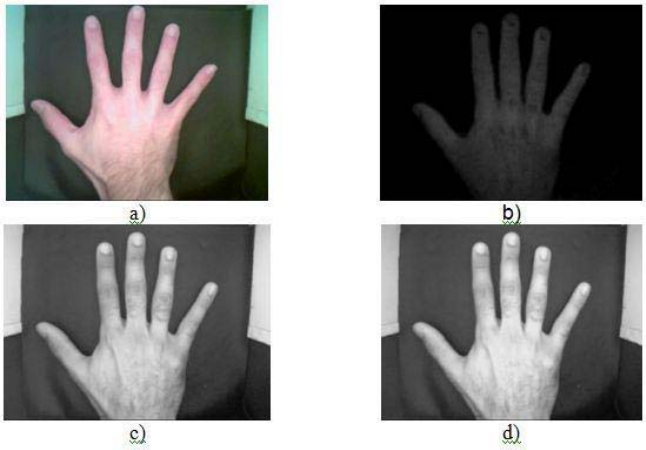


Figura 3. Captura da imagem. a) Imagem original. b) Imagem diferença. c) Imagem em nível de cinza obtida pelo RGB. d) Imagem em nível de cinza obtida pelo HSM.

5. Resultados

Para avaliar o desempenho do Hand Edge foi definido um banco de imagens da palma da mão de 50 pessoas (homens e mulheres). Inicialmente detectaram-se as bordas da palma da mão utilizando a metodologia proposta neste trabalho, o Hand Edge. Após detectou-se as bordas da palma da mão utilizando os detectores mais utilizados na literatura: Roberts, Prewitt e Sobel (Figura 4). Verifica-se que as bordas referente a base de couro preto e as laterais da caixa pintadas de branco não foram identificadas pelo Hand Edge, enquanto os demais detectores a identificaram. Como o objetivo deste trabalho era segmentar apenas a borda da palma da mão, este resultado inicial foi um grande ganho no desempenho da metodologia que conseguiu eliminar algo indesejado sem haver nenhum processamento específico para isto.

Para comprovar o ganho de desempenho obtido com a utilização da nova metodologia foi calculado o índice de mérito de Pratt [5], definido pela Equação 10, onde N é o numero de distâncias usadas, d_i é a distância entre dois pontos correspondentes e α é um parâmetro associado ao tamanho da borda.

$$F_{pratt} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\left(1+\alpha d_i^2\right)} \qquad (10)$$

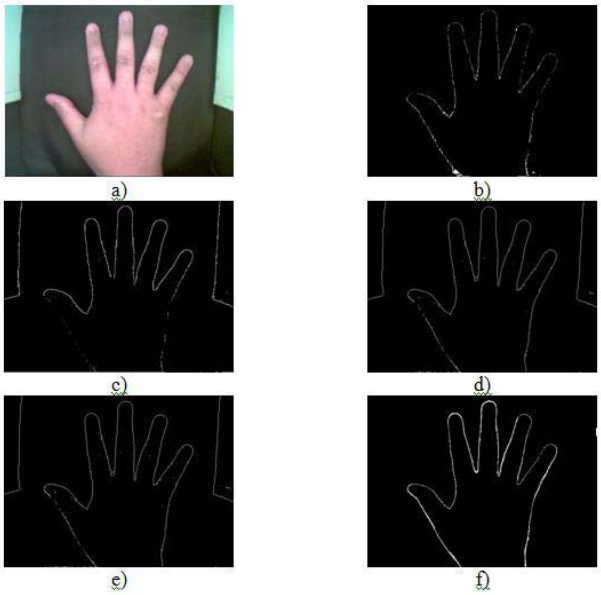


Figura 4. Detecção de bordas. a) Imagem original. b) Imagem obtida pelo Hand Edge. c) Imagem obtida pelo detector Roberts. d) Imagem obtida pelo detector Prewitt. e) Imagem obtida pelo detector Sobel. f) Imagem "Ground truth".

Para obter o índice de mérito de Pratt, foi necessário identificar qual seria realmente a borda de cada palma da mão. Portanto, foi obtida as borda de cada palma da mão de forma manual utilizando o software Adobe Photoshop CS4 definindo a imagem "ground truth" de cada palma da mão (Figura 4-f).

A imagem "ground truth" foi obtida, inicialmente, a partir da imagem diferença (Figura 5-a). Após, aplicou-se o filtro Find Edge do software Adobe Photoshop CS4 (Figura 5-b). Em seguida, apagou-se de forma manual o conteúdo interno da palma da mão deixando apenas a sua borda. Finalmente, aplicou-se uma função de inversão na imagem, tal que o pixel de cor branco tornou-se um pixel de cor negra e vice-versa (Figura 5-c).

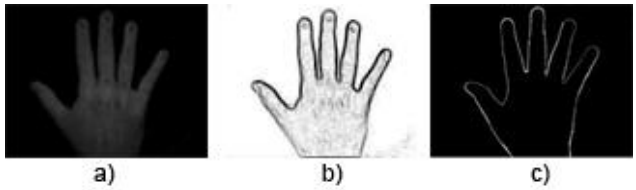


Figura 5. Obtenção da imagem "ground truth". a) Imagem diferença. b) Imagem obtida pelo filtro Find Edge do Photoshop. c) Imagem final.

O desempenho de cada metodologia utilizada (Hand Edge na cor azul, Sobel na cor vermelha, Prewitt na cor magenta e Roberts na cor preta) pode ser apreciado na Figura 6. A metodologia proposta obteve em média um índice superior a 80% de detecção de borda da palma da mão, enquanto as metodologias Sobel e Prewitt obtiveram o mesmo desempenho inferior a 70%. Já a metodologia Roberts obteve um desempenho inferior a 60%.

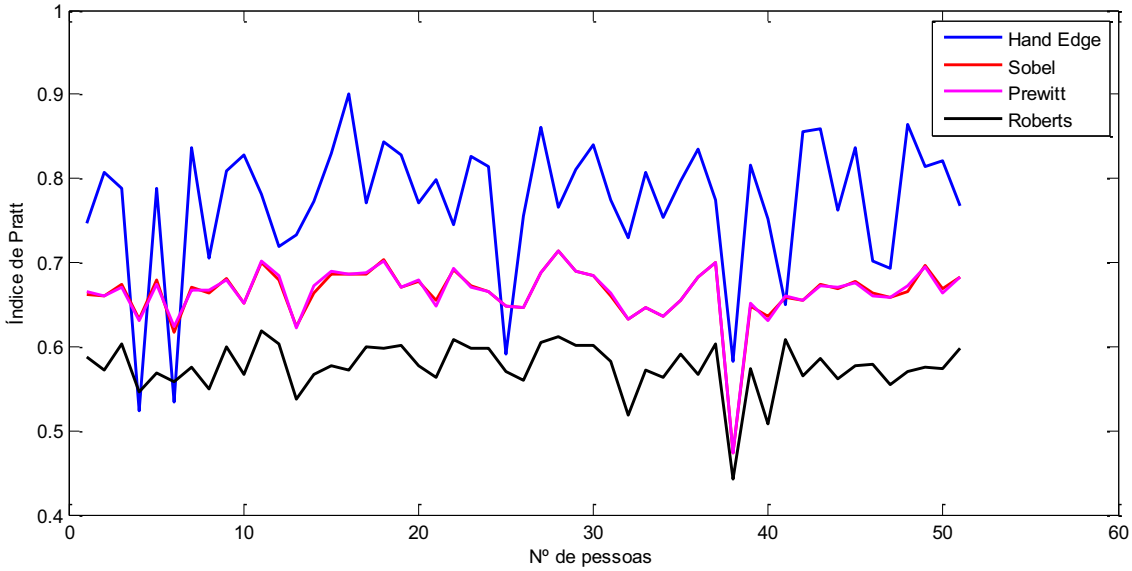


Figura 6. Comparação de desempenho das metodologias.

6. Discussão e Conclusões

Concluiu-se que o método Hand Edge obteve um desempenho superior, comparado aos demais, conforme o gráfico apresentado na Figura 6. De modo que o mesmo atingiu um índice superior a 80% de precisão enquanto os outros não ultrapassaram 70%, além disso, o modo de processamento da imagem é muito mais simples do que os detectores de Roberts, Prewitt e Sobel.Verificou-se, ainda que houve uma maior dificuldade na obtenção da borda das mãos femininas (identificadas pelos pontos de queda do Hand Edge na Figura 6), pois a grande maioria utiliza cremes, que dificulta a segmentação da borda, pois há o reflexo do flash da câmera de aquisição da imagem em lugares isolados da palma da mão. O sistema desenvolvido será de grande valia na visão computacional no processamento de imagens para a extração de características.

7. Trabalhos Futuros

Pretende-se construir um leitor biométrico baseado no reconhecimento da geometria da mão, como continuidade a este trabalho.

8. Referências Bibliográficas

[1] Severino Jr, O. (2009), *Mistura de cores: Uma nova abordagem para processamento de cores e sua aplicação na segmentação de imagens*, Tese de Doutorado – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 118 p.

[2] Gonzalez, R. C., Woods, R. E., Eddins, S. L., *Digital Image Processing using Matlab*. Pearson Education, Inc., 2004.

[3] Gonzalez, R. C., Woods, R. E., *Digital Image Processing.* Massachusetts: Addilson-Wesley, 1993

[4] Severino Jr, O., Gonzaga, A. (2009), HSM: A New Color Space used in the Processing of Color Images. *In WVC 2009*, São Paulo.

[5] ABDOU, I. E., PRATT, W. K., "Quantitative Design and Evaluation of Enhancement / Thresholding Edge Detectors", *Proceedings of IEEE,* vol. 67, no. 5, pp. 753-763, 1979.

EVEREVIS: Sistema de Detecção e Indexação Automática de Objetos em Vídeos Digitais

Bruno do Nascimento Teixeira, Júlia Epischina Engrácia Oliveira,
Fillipe Dias Moreira de Souza,Tiago Oliveira Cunha, Lucas Figueiredo Gonçalves,
Christiane Okamoto, Igor Calil Loures Oliveira, Arnaldo de Albuquerque Araújo
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - Pampulha - CEP: 31270-010
Belo Horizonte Minas Gerais - Brasil
bruno.teixeira@dcc.ufmg.br, julia@dcc.ufmg.br, fdms@dcc.ufmg.br, tocunha@dcc.ufmg.br,
christiane.okamoto@gmail.com, igorcalil@gmail.com, arnaldo@dcc.ufmg.br

Abstract

With the rapid growth of the production of multimedia content, especially after the emergence of various social networking video sharing on the Internet, a great demand has grown for the development of strategies for content navigation in voluminous videos databases. In this context, this paper proposes a system for assisting users in finding a content of interest in broadcast videos, based on video analysis applications such as face detection, dynamic and static summarization and indoor / outdoor scenes. Moreover, using this system one can set a sequence of selected video segments using the mentioned applications, which is displayed in a timeline.

1. Introdução

O crescimento da demanda por informações visuais de vídeos leva à necessidade de se criar formas adequadas de representação, modelagem, indexação, e recuperação de dados multimídias. A localização de um segmento de interesse em uma grande coleção de arquivos de vídeo é ineficiente quanto ao consumo de tempo, uma vez que é necessário assistir cada segmento de vídeo a partir do início e usar os recursos de avanço e recuo para selecionar o segmento desejado. A segmentação automática de vídeo oferece uma solução eficiente e é o primeiro passo crucial em direção a uma representação concisa e abrangente do vídeo baseada em conteúdo.

No que diz respeito à visualização dos resultados do processamento de vídeo, um desafio é fornecer uma ferramenta efetiva e eficiente para o tratamento e exibição dos dados gerados pelos algoritmos disponíveis, pois esta ferramenta deve oferecer um alto grau de interatividade, permitindo que os usuários limitem-se aos seus interesses e descartem o que não for relevante.

Neste trabalho, é proposto um sistema interativo de análise de vídeo baseado no conteúdo e que contém algoritmos específicos capazes de facilitar a navegação no vídeo de notícias. Especificamente, esses algoritmos de processamento de vídeo visam facilitar a navegação entre cenas internas e externas presentes no vídeo, fornecer uma sumarização estática e dinâmica do vídeo, navegar no vídeo através da detecção de faces, e reconhecer eventos específicos nos vídeos.

Alguns trabalhos lidam com o problema de tornar mais fácil a interação entre os usuários e o acesso ao conteúdo de um vídeo. Por exemplo, Forlines [2] descreve um protótipo de sistema para exibição de vídeo que leva em conta o conteúdo do vídeo. O protótipo proposto renderiza os quadros de um vídeo em várias regiões da tela, de acordo com a estrutura de seus conteúdos. Essa abordagem permite manter a continuidade da estória, ao passo que melhora o processo de visualização do vídeo.

Com respeito à busca de vídeos, Su *et al*. [11] apresentam um método para busca de vídeos por conteúdo que utiliza indexação baseada em padrões e técnicas de combinação. O problema da alta dimensionalidade dos vetores de características é resolvido com a indexação baseada em padrões, que também se mostra um método eficiente para encontrar os vídeos desejados entre uma quantidade grande de dados variados.

O problema de indexação e recuperação eficientes em base de dados de vídeos é abordado por Morand *et al*. [9]. Em seu trabalho, um método para recuperação de objetos escaláveis em vídeos de alta resolução é proposto, com a utilização de descritores do vídeo obtidos através de

distribuições estatísticas de coeficientes da transformada *wavelet*.

Motivado pela função de uma tabela de conteúdo em livro, [10] propôs uma técnica de construção da tabela de conteúdo em vídeos baseada em clusterização não supervisionada, facilitando o acesso ao vídeo. O método proposto é dividido em quatro módulos: detecção de cortes e quadros chaves, extração de características, agrupamento adaptativo e construção da estrutura de cenas. Diferente da TV convencional, a TV interativa permite ao usuário navegar para frente ou para trás no tempo.

Em [6], são propostas formas de navegação baseadas na indexação de episódios (EIT - *episodic indexing theory*). Para validação dos métodos e teste de sua eficiência na navegação temporal, foi usado simulador de TV interativa. Na indexação de cenas, pode-se usar faces, sons, cores e movimentos, como na interface web de busca de vídeos do projeto Open Video Project [4], facilitando o acesso a trechos do vídeo.

Na próxima seção será apresentada a metodologia utilizada para o desenvolvimento dos algoritmos integrados ao sistema interativo de análise de vídeo baseado em conteúdo. A seção três mostra o resultado do sistema e por fim, a última seção apresenta as considerações finais.

2. Metodologia

A seguir, são apresentadas as metodologias específicas para cada algoritmo proposto e como foi elaborada a interface.

2.1. Segmentação de cenas

Vídeos de notícias podem ser segmentados em cenas internas e externas. Cenas internas são descritas pela apresentação das notícias por um âncora ao passo que cenas externas caracterizam-se por qualquer cena gravada fora do estúdio.

Desta forma, é natural optar-se por utilizar descritores de cor para representar as características visuais dos vídeos. Isto porque, de uma forma geral, as cenas que exibem o âncora são mais semelhantes quanto à cor, ao contrário das cenas de entrevistas fora do estúdio de apresentação, que são consideravelmente mais dissimilares neste aspecto. Além disso, as cenas internas sofrem poucas variações, bem como apresentam certo padrão no cenário das imagens, em oposição às cenas externas que usualmente apresentam mais variações e diferentes padrões de cenário (alguns até parecidos com as cenas internas).

Os vídeos da base de dados foram caracterizados por descritores de cor (matiz como parâmetro principal), utilizando o descritor HueSIFT [12]. O descritor de cor HueSIFT é uma variação do descritor de local SIFT proposto por Lowe [7]. O SIFT descreve as informações locais de regiões (de uma imagem) através de histogramas de orientação por bordas. Por serem normalizados, os descritores SIFT, estes são invariante a mudanças de iluminação. Deste modo, o HueSIFT é a concatenação dos histogramas de matiz aos histogramas de gradientes do SIFT.

Após a caracterização dos vídeos, usa-se a representação de conjunto de palavras visuais (*Bag-of-Visual-Words* para representar os descritores de uma cena por meio de um vetor de dados de padrões visuais. Esses padrões são frutos de um procedimento de medição de similaridade (usando métricas de distância, tal como a Euclidiana) entre os descritores de características da base de dados de treinamento.

A fim de classificar as cenas, foi aplicado o algoritmo de máquina de vetores de suporte (SVM – *Support Vector Machine*) [13]. Após a rotulação dos vetores de características extraídos da base de dados de treinamento como cenas internas e externas, o próximo passo é o treinamento do classificador tendo estes dados anotados como entrada. Uma vez que o aprendizado do classificador esteja finalizado, resultando um modelo de predição, os dados de entrada são submetidos ao processo de classificação. Finalmente, tem-se a classificação das cenas dos vídeos como internas ou externas com uma acurácia influenciada pelo modelo de treinamento utilizado. Por ser um problema binário, foi utilizado um *kernel* linear.

2.2. Sumarizações estática e dinâmica

Sumarização de vídeo é o processo de extração de um resumo do conteúdo original do vídeo, cujo objetivo é fornecer rapidamente a informação concisa do conteúdo do vídeo, preservando a mensagem do vídeo original. Sumários de vídeos apresentam-se como uma solução para permitir a identificação prévia sobre conteúdos de vídeos, a fim de evitar o tempo despendido na análise manual desta tarefa. O resumo possibilita ao usuário julgar se o vídeo é realmente relevante para o fim desejado, auxiliando-o na tomada de decisões.

Existem dois tipos de resumos de vídeos: resumos estáticos e resumos dinâmicos. O primeiro é formado por um conjunto de imagens consideradas as mais representativas do vídeo, enquanto o segundo é formado por um conjunto de segmentos do vídeo dispostos em ordem temporal. Em geral, este último é mais atrativo por incorporar elementos de movimentação e aúdio.

Particularmente, neste trabalho foi utilizado um método concebido para ser simples e eficiente na geração de resumos estáticos. Histogramas de cor e perfis de linha [5] são utilizados para representar as imagens do vídeo. Após a extração das características visuais, as imagens são agru-

padas pelo algoritmo *k-means* [8]. O quadro-chave de cada grupo é identificado, e por fim as imagens são dispostas em ordem cronológica, produzindo então o resumo estático do vídeo. Um passo adicional foi incorporado à abordagem com a finalidade de produzir resumos dinâmicos. Ao fim do processo de geração do resumo estático é realizada a detecção das tomadas do vídeo e as tomadas que possuem imagens presentes no resumo estático são então selecionadas, formando assim o resumo dinâmico do vídeo.

2.3. Detecção de faces

Os algoritmos para a detecção de faces possuem como tarefa encontrar os locais e tamanhos de um número conhecido de rostos quase que freqüentemente na posição frontal.

O algoritmo deste trabalho utiliza o método Viola-Jones [14], que é implementado na biblioteca *OpenCV* (*Open Source Computer Visio*n) [1]. Esse método de detecção de objetos em imagens é baseado em quatro conceitos: características simples e retangulares chamadas *Haar features*, uma imagem integral para rápida detecção de características, o método *Adaboost* [3] de aprendizado de máquina, e um classificador em cascata para combinar eficientemente as características.

Uma pequena janela é verificada através de toda a imagem e um classificador é aplicado a cada janela, retornando ou não uma face em cada localidade, e isto é repetido em múltiplas escalas. Viola-Jones combina classificadores fracos como se fossem uma cadeia de filtros, o que é especialmente eficiente para a classificação de regiões em uma imagem.

2.4. Interface

A interface (Figuras 2 e 3) consiste em três áreas principais: navegação, exibição e montagem. A área de navegação é o primeiro ponto focal do usuário e apresenta os resultados dos algoritmos. Cada algoritmo possui uma aba separada e em seu interior, imagens que representam seus segmentos. Essas imagens são clicáveis para a exibição no *player* e arrastáveis até a área de montagem. A área de exibição apresenta um *player* com comandos simples de reprodução e pode ser redimensionado sem alterar a razão de aspecto do vídeo. Por fim, a área de montagem apresenta uma linha do tempo inicialmente vazia que pode ser preenchida arrastando-se os segmentos disponíveis da área de navegação. Através do botão de exibição, os segmentos são exibidos seqüencialmente e sem intervalos. Também, os segmentos podem ser reordenados facilmente na linha temporal, formando uma nova edição para o vídeo.

Além dessas áreas, há um menu na parte superior da tela no qual o usuário pode se registrar, escolher um vídeo disponível para visualização ou enviar um vídeo para o processamento dos algoritmos.

A interface é uma camada completamente isolada do módulo de processamento de vídeos. A comunicação entre ambos ocorre somente por meio de um banco de dados. O banco de dados utilizado é o MySQL[1] e os dados são manipulados via PHP, que é uma linguagem apropriada para desenvolvimento *Web*, constituindo o lado servidor da aplicação. O lado do cliente é baseado em Flash, construído através do programa Adobe Flash Builder e das linguagens MXML e Action Script.

3. Resultado do sistema

O diagrama do sistema interativo de análise de vídeo com base no conteúdo proposto é mostrado em alto nível na Figura 1. Como pode ser visto, foi implementada uma aplicação *Web*, que é uma das formas mais efetivas de permitir que usuários reais interajam com o sistema. A arquitetura interna segue os modelos usuais de comunicação entre módulos de uma aplicação *Web*.

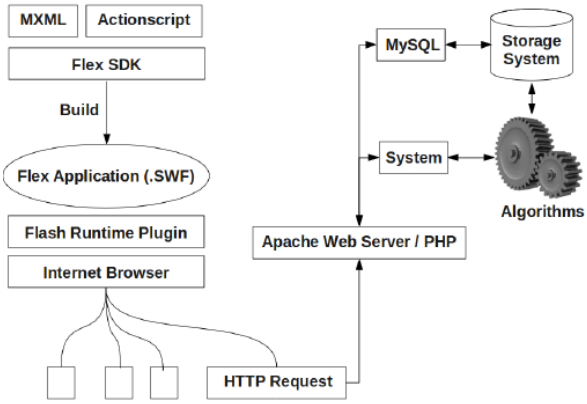


Figure 1. Diagrama de operação do sistema proposto.

A Figura 2 apresenta a tela inicial da interface. Um vídeo é enviado por um usuário e é processado usando os algoritmos disponíveis no sistema. Depois disso, o usuário pode navegar por esse vídeo particular, ou por qualquer outro da base de dados do sistema. A Figura 3, mostra o resultado da segmentação do vídeo usando algoritmos de detecção de faces, classificação em cenas internas e externas e sumarização estática.

1 http://www.mysql.com/

4. Conclusões

Neste trabalho, foi apresentada uma interface para navegação em vídeos de notícias. O sistema de navegação é composto por sumarização estática, detecção de faces e pela classificação das cenas em internas e externas. As diversas combinações de módulos estão sendo exploradas para determinar a hierarquia que maximiza a navegação. Métricas que determinam a eficiência da navegação serão propostas validando as técnicas e a estrutura proposta.

5. Agradecimentos

Os autores agradecem ao CNPq, à CAPES e à FAPEMIG pelo suporte financeiro do projeto.

References

[1] OpenCV: Open Computer Vision library. http://sourceforge.net/projects/opencvlibrary.

[2] C. Forlines. Content aware video presentation on high-resolution displays. In *Proceedings of the working conference on Advanced Visual Interfaces*, pages 57–64. ACM, 2008.

[3] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Proceedings of the Second European Conference on Computational Learning Theory*, pages 23–37, London, UK, 1995. Springer-Verlag.

[4] G. Geisler, G. Marchionini, B. M. Wildemuth, A. Hughes, M. Yang, T. Wilkens, and R. Spinks. Video browsing interfaces for the Open Video Project. In *ACM SIGCHI Conference on Human Factors in Computing Systems, CHI 2002: Extended Abstracts*, pages 514–515. 2002.

[5] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital Image Processing using Matlab*. Prentice-Hall, 2003.

[6] J. Kim, H. Kim, and K. Park. Towards optimal navigation through video content on interactive tv. *Interact. Comput.*, 18:723–746, July 2006.

[7] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–11–, 2004.

[8] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. L. Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.

[9] C. Morand, J. Benois-Pineau, J.-P. Domenger, J. Zepeda, E. Kijak, and C. Guillemot. Scalable object-based video retrieval in hd video databases. *Signal Processing: Image Communication*, 25:450–465, 2010.

[10] Y. Rui, T. S. Huang, and S. Mehrotra. Constructing table-of-content for videos. *ACM MULTIMEDIA SYSTEMS JOURNAL, SPECIAL ISSUE MULTIMEDIA SYSTEMS ON VIDEO LIBRARIES*, 7:359–368, 1998.

[11] J.-H. Su, Y.-T. Huang, H.-H. Yeh, and V. S. Tseng. Effective content-based video retrieval using pattern-indexing and matching techniques. *Expert Systems with Applications*, 37:5068–5085, 2010.

[12] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *PAMI*, 32(9):1582–1596, 2010.

[13] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, New York, 1995.

[14] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, 2001.

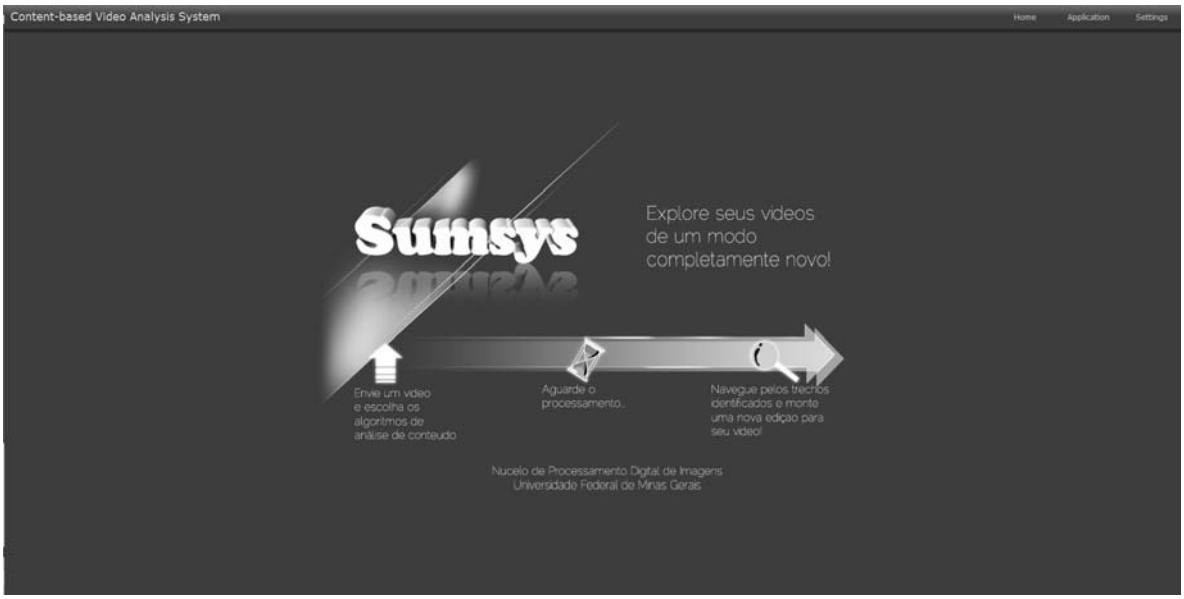


Figure 2. Tela inicial do sistema.

(a)

(b)

(c)

(d)

Figure 3. (a) Navegação pelas faces; (b) Navegação pelo sumário estático; (c) Navegação pelas cenas internas; (d) Navegação pelas cenas externas.

Análise Discriminante e Classificação de Imagens 2D de Ultrassonografia Mamária

Albert C. Xavier[1], João R. Sato[2], Gilson A. Giraldi[3] e Carlos E. Thomaz[1]

[1]FEI, São Bernardo do Campo, SP, Brasil
[2]UFABC, Santo André, SP, Brasil
[3]LNCC, Petrópolis, RJ, Brasil

Abstract

Research in detection and treatment of cancer using images, including breast cancer, has improved in the last years. Clinical experts say that breast cancer is the most common cause of cancer death among women worldwide. This paper approaches this problem by carrying out a statistical discriminant analysis of mammary tumors in 2D ultrasound images. To analyze the images, univariate and multivariate statistical methods have been investigated with the aim of extracting discriminant information for classification purposes. Our results show a good classification performance of the multivariate statistical methods when using preprocessed and segmented 2D ultrasound images as a whole. Additionally, our approach highlights the most statistically significant differences between the tumors, ranking in the original image space the simplest and most difficult cases of classification.

1. Introdução

Na área médica, a classificação de imagens é utilizada para auxiliar no diagnóstico de vários tipos de doença, incluindo o câncer de mama. O câncer de mama tem registrado crescimento muito preocupante nos últimos anos e se posiciona como maior causador de morte por câncer na população feminina mundial. No Brasil, 33 mulheres morrem por dia em decorrência do câncer de mama [1].

Nesse contexto, os recursos computacionais atuais e o processamento de imagens, aliados à estatística, podem contribuir muito para o esclarecimento e um diagnóstico mais preciso do câncer de mama. Há um crescente avanço nos métodos de diagnóstico por imagem incluindo a utilização de ultrassom. Na imaginologia mamária, a ultrassonografia representa uma modalidade de diagnóstico muito significativa por ter menor custo e ainda evitar o contato prejudicial dos pacientes com radiação [7].

Este trabalho tem como objetivo principal analisar as alterações morfológicas estatisticamente relevantes para classificação das imagens de tumor de mama. Tal análise é essencial para o estudo das diferenças entre os grupos de tumores benignos e malignos. São realizadas análises no espaço de características, definido por quantidades geométricas e de textura, bem como no espaço de imagens, comparando-se os resultados.

2. Análise Discriminante Multivariada

Técnicas em estatísticas multivariadas como LDA (Linear Discriminant Analysis) e SVM (Support Vector Machine) podem ser utilizadas na etapa de classificação [2, 4, 5, 11, 14]. Nestas aplicações , cada amostra é representada por um ponto no espaço n-dimensional, onde n é o número de variáveis do problema em questão. A denominação multivariada corresponde às técnicas que utilizam todas as características (variáveis) para interpretação do conjunto de dados.

2.1. LDA

A proposta do método LDA, também conhecido como método de Fisher [2], é encontrar o hiperplano de maior separação entre os grupos analisados. O cálculo desse hiperplano de separação considera o conhecimento prévio da classe ou grupo de cada amostra. A análise LDA é paramétrica, ou seja, considera que a distribuição de probabilidade das amostras é conhecida e pode ser representada pela média e dispersão das amostras.

O método baseia-se na diminuição do espalhamento das amostras com relação ao grupo a qual pertencem e, também, na maximização da distância da média entre estes grupos [2]. Em outras palavras, calcula-se as matrizes de espalhamento inter-classes e intra-classes com objetivo de discriminar os grupos de amostras pela maximização da separabilidade entre classes enquanto minimiza-se a variabilidade dentro das mesmas.

Matematicamente, as matrizes de espalhamento inter-classes (S_b) e intra-classes (S_w) são definidas como:

$$S_b = \sum_{i=1}^{g} N_i(\overline{\mathrm{x}}_i - \overline{\mathrm{x}})(\overline{\mathrm{x}}_i - \overline{\mathrm{x}})^T, \quad (1)$$

$$S_w = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (\mathrm{x}_{i,j} - \overline{\mathrm{x}}_i)(\mathrm{x}_{i,j} - \overline{\mathrm{x}}_i)^T, \quad (2)$$

onde g é o número de grupos analisados, N_i a quantidade de amostras do grupo i, $\overline{\mathrm{x}}$ e $\overline{\mathrm{x}}_i$ são a média total e a média das amostras do classe i, respectivamente, e $\mathrm{x}_{i,j}$ é a amostra j do grupo i. E a relação proposta por Fisher que deve ser maximizada é dada pela equação (3) a seguir:

O principal objetivo do método LDA é encontrar a matriz de projeção P_{LDA} que maximiza a razão entre o determinante da matriz de espalhamento inter-classes S_b e o determinante da matriz de espalhamento intra-classes S_w, conhecido como critério de Fisher e descrito matematicamente por

$$P_{LDA} = \arg\max \frac{|S_b|}{|S_w|}. \quad (3)$$

O critério de Fisher descrito pela Equação 3 é satisfeito quando a matriz de projeção P_{LDA} é composta, no máximo, pelos $(g-1)$ autovetores de $S_w^{-1}S_b$, cujos autovalores correspondentes são não-nulos.

Na prática, essa razão somente pode ser calculada se a matriz de espalhamento intra-classe S_w for não-singular. Exatamente nesse momento deve ser observado a proporção entre o número de amostras e de variáveis. Em cenários onde o número total de amostras N é bem menor que o número de variáveis n, ocorre uma instabilidade no cálculo da matriz inversa de S_w [3]. A quantidade de amostras necessárias para evitar essa instabilidade no cálculo da matriz inversa de S_w deve ser igual ou superior a 5 vezes a quantidade de variáveis destas [6]. Considerando o cenário do trabalho desenvolvido aqui, certamente ocorreria esse problema, pois a quantidade de amostras é bem inferior a quantidade de variáveis de cada amostra, representada por imagens de tumores.

Portanto, para o tratamento do problema de instabilidade no cálculo da inversa da matriz S_w, utilizou-se o método denominado MLDA (*Maximum uncertainty Linear Discriminant Analysis*) [13]. Essa técnica consiste em substituir S_w por outra matriz regularizada S_w^*, gerando um aumento no espalhamento dos dados e mantendo as variações mais relevantes existentes nas amostras. A nova matriz regularizada S_w^* pode ser calculada por meio dos seguintes passos:

1. Selecionar os autovetores Φ e autovalores Λ de S_p, onde $S_p = \frac{S_w}{N-g}$;
2. Calcular a média dos autovalores $\bar{\lambda}$;
3. Gerar uma nova matriz de autovalores baseada na dispersão dos maiores $\Lambda^* = diag[max(\lambda_1, \bar{\lambda}), ..., max(\lambda_n, \bar{\lambda})]$;
4. Calcular a matriz de espalhamento intra-classes regularizada $S_w^* = (\Phi\Lambda^*\Phi^T)(N-g)$.

Com a matriz S_w^* calculada, substitui-se S_w da equação (3) por S_w^* e regulariza-se o critério de Fisher para problemas onde $N \ll n$.

2.2. SVM

Com o mesmo propósito final do LDA, o método SVM também visa encontrar o hiperplano de maior separação entre os grupos de amostras. De forma análoga ao LDA, o cálculo do hiperplano de separação considera o conhecimento prévio da classe de cada amostra j investigada, ou seja, $y_j \in \{-1, 1\}$. Porém, o método SVM é não-paramétrico, ou seja, não considera a distribuição de probabilidade das amostras.

O SVM é uma técnica de reconhecimento de padrões com sólido embasamento teórico e que tem apresentado resultados muito satisfatórios mesmo quando comparado a métodos clássicos como redes neurais e árvores de decisão. Trata-se essencialmente de um classificador de duas classes mas que também pode ser estendido para tratamento de mais de duas classes [8]. Este método é baseado na teoria de aprendizado estatístico, pioneiramente descrita por Vapnik e colaboradores [14]. Como vantagens, o SVM apresenta boa capacidade de generalização, robustez em grandes dimensões e convexidade da função objetivo. Para encontrar a solução ótima do classificador é usada uma função quadrática, em que não há presença de vários mínimos locais, e sim apenas um mínimo global, o que facilita a obtenção do valor ótimo.

O hiperplano SVM pode ser definido resumidamente como:

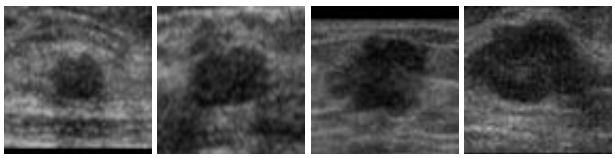
$$w_{SVM} = \sum_{j=1}^{N} \alpha_j y_j \mathrm{x}_j, \quad (4)$$

onde α_j são os coeficientes de Lagrange não-negativos obtidos pela solução de um problema de otimização quadrático com restrições de desigualdade linear [8]. As observações de treinamento x_j, com α_j não-zero, ficam na fronteira da margem e são chamadas de vetores de suporte [11]. O SVM pode fazer uso de vários tipos de *kernel*, incluindo o polinomial, o Gaussiano-RBF e o linear. Para os experimentos desenvolvidos no presente trabalho, foi adotado somente o *kernel* linear.

3. Arcabouço Experimental

3.1. Banco de Imagens e Características

Foram utilizadas 250 imagens ultrassonográficas de tumores mamários e um conjunto de valores referentes às características de circularidade, sombra acústica e heterogeneidade destes tumores. As imagens de tumores benignos totalizam 100 e as de tumores malignos 150. Como existem 5 imagens diferentes do mesmo tumor, tem-se 20 diferentes tumores benignos e 30 diferentes tumores malignos. A Figura 1 apresenta exemplos das imagens investigadas aqui. As images foram adquiridas por meio de um equipamento modelo Voluson 730 (General Electric, USA) com um transdutor S-VNW5-10. As especificações técnicas deste são: frequência de varredura de 5-10 MHz, largura de varredura de 40 mm e ângulo de varredura de 20 a 30 graus.



(a) (b) (c) (d)

Figura 1. Exemplos de imagens investigadas com tumores (a,b) benignos e (c,d) malignos.

Os valores geométricos de circularidade foram obtidos a partir de um ponto central da região de interesse (tumor). Calculou-se a distância entre cada ponto da borda da lesão e o centro da imagem. Os valores também foram normalizados pela área total da imagem. Em geral, as lesões malignas apresentam valores mais altos de desvio padrão em relação a distância média quando comparadas às lesões benignas. As outras duas características, heterogeneidade e sombra acústica, são de textura. Os valores de heterogeneidade, utilizando imagens em escala de cinza, foram calculados por meio da entropia BGS (*Boltzman-Gibbs-Shannon*). Geralmente as lesões malignas são mais heterogêneas que as benignas. Já a sombra acústica tem relação direta com a região inferior da imagem. No caso das imagens de tumores mamários, a sombra acústica está extremamente relacionada ao tipo de tumor. Na maioria dos casos de tumor benigno ocorre a formação de um reforço acústico abaixo do região do tumor em decorrência da existência de muitas partículas de água. Os tumores malignos, que são geralmente mais sólidos, tendem a apresentar uma sombra acústica. Nos tumores malignos a sombra acústica é mais intensa (cor mais branca) que o reforço acústico presente nos tumores benignos. Então, para calcular os valores dessa característica foram comparados os histogramas da região da lesão e da região logo abaixo desta. Quanto mais escuro é a região abaixo da lesão, maior é a probabilidade desta ser benigna [10]. Os valores de pré-processamento das características foram atribuídos por radiologistas e calculados em estudos anteriores [9, 4].

3.2. Pré-processamento

Com o propósito de ajustar as diferenças de resolução, as imagens foram submetidas a uma etapa de pré-processamento. As imagens fornecidas, em escala de cinza, possuem resoluções diferentes que variam de 57 a 161 pixels na altura e de 75 a 199 pixels na largura. Independentemente da resolução, a maioria dos tumores está localizada no centro das imagens. Cada pixel (ponto) da imagem é representado dentro de uma escala de 0 a 255, que são as variações de tons de cinza em um sistema de representação de luminância com 8 bits de resolução. A grande maioria das imagens foi e recortadas para adequação ao processo de análise. A resolução de 70 x 70 pixels, escolhida após testes, visou evitar a perda de informações significativas no recorte das imagens. No entanto, 20 imagens de um total de 250 tiveram que ser redimensionadas pois possuiam resolução inferior a 70 pixels na altura. A maior diferença encontrada, em 5 dessas 20 imagens, foi de 13 pixels. Nas demais imagens redimensionadas o maior ajuste foi de 5 pixels.

3.3. Experimentos

As imagens foram submetidas primeiramente a um processo de redução de dimensionalidade utilizando a técnica PCA (*Principal Component Analysis*) [3], porém preservando todas as componentes com autovalores não-nulos [12]. Na sequência, as imagens projetadas no espaço do PCA foram então projetadas no espaço dos classificadores lineares MLDA e SVM para se obter a separação dos tumores malignos e benignos.

Neste processo de separação linear empregando os classificadores MLDA e SVM, foram adotados dois tipos de parâmetros de entrada. No primeiro a classificação das imagens de tumores mamários considera a intensidade dos pixels, ou seja, a análise é feita na imagem como um todo considerando todas as componentes do PCA com autovalores não-nulos. Já no segundo processo os classificadores são executados utilizando as características extraídas dos tumores nas imagens (circularidade, heterogeneidade e sombra acústica). Os valores atribuídos a cada característica são tratados separadamente nos classificadores.

O cálculo da acurácia dos classificadores foi baseado na abordagem de validação cruzada (*cross-validation*). Nos experimentos aqui realizados, foi adotada a forma extrema da abordagem de validação cruzada denominada *leave-one-out* [3]. Este método foi aplicado em decorrência do pequeno número de amostras rotuladas.

4. Resultados

4.1. Análise Estatística Univariada

A Figura 2[1] apresenta as imagens médias das amostras dos grupos benigno e maligno. Claramente observa-se uma diferença no formato e tamanho dos tumores nas imagens. Na imagem média das lesões benignas, vista na Figura 2(a), o tumor é menor e mais concentrado. Já o tumor das imagens de lesões malignas, visto na Figura 2(b), parece ser maior e mais espalhado. Na sequência, por meio de uma operação matemática simples, foi feita a subtração das imagens médias dos dois grupos com o propósito de visualizar possíveis regiões discriminantes. O resultado apresentado na Figura 2(c) indica que existe uma possível diferença entre as imagens na região inferior e também na região central direita. Provavelmente a sombra do tumor exerce influência sobre estas áreas. Em complemento às diferenças observadas na subtração das imagens médias, foi implementado o teste de hipóteses baseado na distribuição de probabilidade de *t-student*. Nesse experimento, a finalidade foi obter as variações mais relevantes estatisticamente entre as imagens dos grupos de tumores malignos e benignos. Na Figura 2(d), em cor azul, pode-se observar as regiões mais discriminantes selecionadas de acordo com um nível de confiança de 99.9% (ou $p < 0.001$) da tabela *t-student*. Do ponto de vista estatístico, estas são as regiões de maior discriminância entre os dois grupos de imagens de tumor mamário, avaliadas pixel-por-pixel. Essas regiões mais discriminantes foram projetadas sobre uma imagem de referência que representa aqui a subtração das imagens médias dos dois grupos.

4.2. Desempenho dos Classificadores

A Tabela 1 apresenta a acurácia dos classificadores nas análises das imagens como um todo e das características geométrica e de textura. O desempenho dos classificadores considerando a intensidade de todos os pixels da imagem simultaneamente foi notadamente superior ao desempenho desses mesmos classificadores analisando as características extraídas das imagens. Esses resultados indicam que os fatores determinantes na classificação dos grupos vão além de características como circularidade, heterogeneidade e sombra acústica, principalmente para separações lineares.

	MLDA			SVM		
	Total	Benignos	Malignos	Total	Benignos	Malignos
Imagem	100%	100%	100%	100%	100%	100%
Circularidade	78%	79%	77%	76%	68%	82%
Heterogeneidade	64%	63%	65%	60%	46%	70%
Sombra Acústica	73%	60%	82%	70%	61%	75%

Tabela 1. Taxas de classificação.

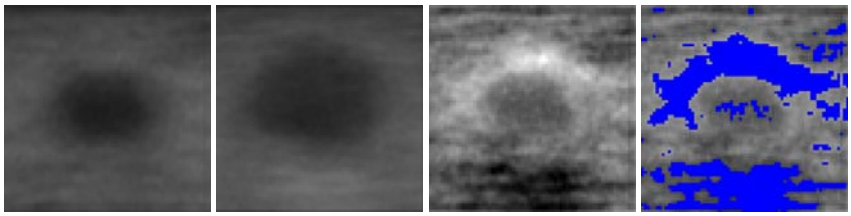
Pode-se constatar que os dois classificadores separaram muito bem os grupos na análise simultânea de todos os pixels da imagem. A taxa de acerto foi de 100% em ambos classificadores. Em contraste, a classificação considerando os valores das características apenas não apresentou resultado satisfatório. A maior taxa de acerto, de 78%, foi obtida na avaliação da característica circularidade pelo classificador MLDA.

A Figura 3 apresenta os hiperplanos descritos pelos dois classificadores com o intuito de observar que a ordem de classificação das amostras, no entanto, ficou diferente entre esses classificadores. A imagem 95, de tumor benigno, ficou mais próxima dos casos malignos no hiperplano classificador do SVM. Uma possível explicação para tal posicionamento seria a presença de características inerentes aos tumores malignos como a sombra acústica e o formato irregular. Já a imagem 93, também do grupo de tumores benignos (lado direito), foi ordenada como um dos extremos do grupo indicando uma classificação mais fácil e/ou mais simples. De forma análoga, também destaca-se a imagem 206 do grupo de tumores malignos. Esta ficou posicionada no extremo do grupo de tumores malignos (lado esquerdo). É valido destacar ainda a posição da imagem 166 no hiperplano gerado pelo classificador MLDA. Esta imagem foi ordenada como próxima aos tumores malignos, sugerindo uma maior complexidade de classificação. Observando a imagem 166, nota-se que esta apresenta características dos dois grupos de tumores, ou seja, esta apresenta características de tumor benigno como a ausência de sombra acústica, mas também apresenta formato irregular, geralmente característico dos tumores malignos. A proximidade da fronteira de decisão definida pela ordem das imagens pode ser interpretada como uma maior dificuldade de classificação. Já a maior distância desta pode indicar uma maior facilidade de classificação.

4.3. Navegação nos Hiperplanos

A Figura 4(a) apresenta uma padronização dos formatos de nódulos mamográficos proposta pelo BI-RADS (*Breast Image Reporting and Data System*). Nesta é possível observar a transição do nódulo benigno para o maligno. A

1 Para visualizar as imagens em cores, favor acessar a referência eletrônica do artigo.



(a) (b) (c) (d)

Figura 2. Análise estatística univariada: (a) Imagem média de tumores benignos; (b) Imagem média de tumores malignos; (c) Imagem da subtração de (a) por (b); (d) Imagem das diferenças estatísticas ($p < 0.001$).

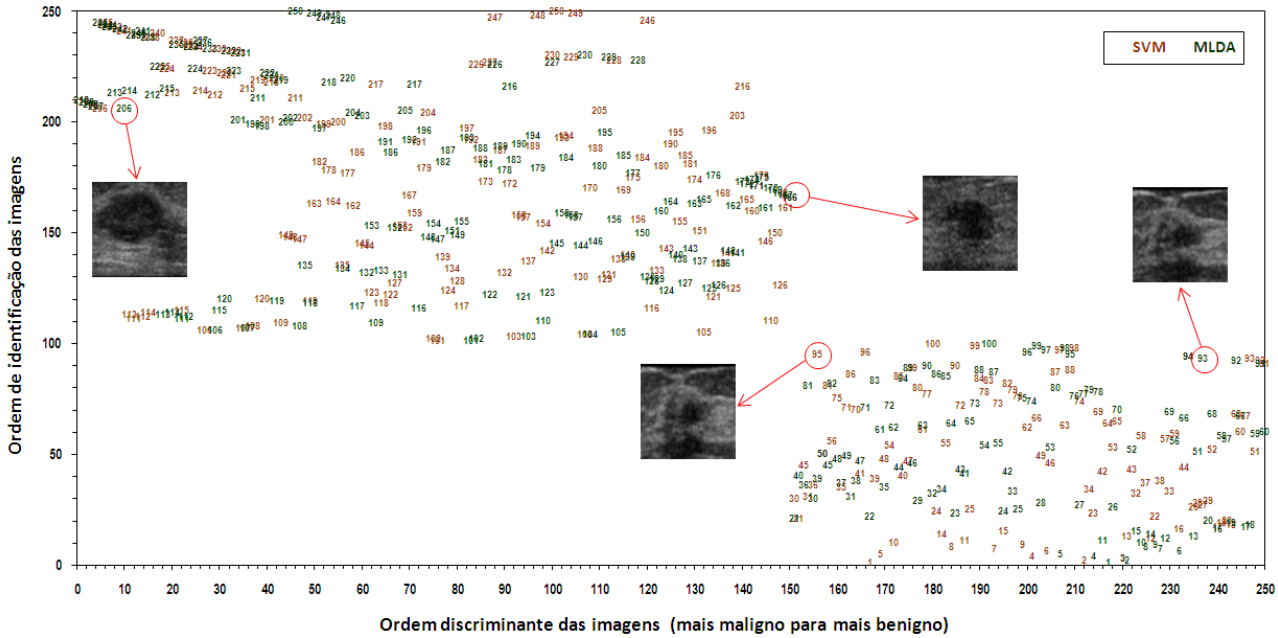
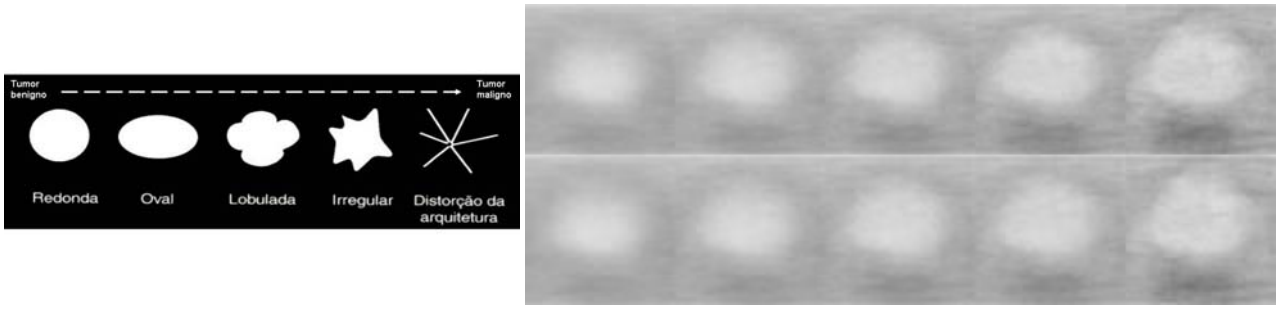


Figura 3. Hiperplanos MLDA e SVM de separação: maligno a esquerda e benigno a direita.

diferença no formato dos nódulos é bastante expressiva. Em âmbito geral, trata-se de um bom guia para o estudo dos tumores mamográficos em imagens de ultrassonográficas.

Para avaliar essa transição, apresentada pelo BI-RADS, foi gerada a navegação nos hiperplanos MLDA e SVM entre os grupos orientada a protótipos. Tal navegação foi construída com as imagens médias calculadas a partir de intervalos estabelecidos na ordenação das imagens, obtida pelos métodos MLDA e SVM. O intervalo adotado foi de cinquenta imagens, ou seja, a cada cinquenta imagens ordenadas calcula-se a imagem média. Como a ordenação é composta pelas 250 imagens amostrais, obtém-se então cinco imagens para navegação dos protótipos ordenados do mais benigno para o mais maligno.

A Figura 4(b), em modo negativo, ilustra a transição das imagens entre os grupos de tumores benignos e malignos para os classificadores MLDA e SVM. É possível notar, apesar do número limitado de imagens considerado, diferenças no formato dos tumores e também na região inferior destes. A região mais clara na parte inferior das imagens é mais intensa nas imagens que pertencem ao grupo dos tumores malignos (lado direito). O tamanho dos tumores também apresenta variação na transição entre os grupos. O tamanho dos tumores benignos (lado esquerdo) é menor e segue aumentando a medida que a navegação se aproxima dos tumores malignos (lado direito).



(a) (b)

Figura 4. (a) Formato BI-RADS dos tumores mamários; (b) Navegação nos hiperplanos MLDA (linha superior) e SVM (linha inferior) por meio de protótipos (dos mais benignos para os mais malignos).

5. Conclusão

O sucesso da aplicação de métodos estatísticos lineares é fortemente dependente das fases de pré-processamento e segmentação das imagens. Portanto, muitas vezes a aplicação de técnicas de melhoramento de imagens, na fase de pré-processamento, pode favorecer um determinado tipo de análise mas prejudicar outro. Os resultados obtidos na análise multivariada considerando a intensidade de todos os pixels da imagem simultaneamente, por meio dos classificadores MLDA e SVM, alcançaram desempenho superior perante a mesma análise utilizando características extraídas das imagens como circularidade e sombra acústica. Já a análise univariada indicou as regiões central e inferior da imagem como mais diferentes estatisticamente. Na navegação entre os grupos de tumores benignos e malignos, foi possível notar as diferenças entre os grupos justamente nessas regiões indicadas pela análise univariada. Muito provavelmente as características de circularidade e sombra acústica exerceram influência na discriminação dos grupos. Entretanto, a análise final dos resultados indicou que as diferenças entre as imagens de tumores benignos e malignos não são concentradas em determinadas regiões da imagem. Trata-se de uma ferramenta útil para o diagnóstico médico e identificação de casos mais simples e difíceis de classificação.

Referências

[1] Instituto Nacional de Câncer. Estimativa 2010: incidência de câncer no Brasil. 2009.

[2] R. A. Fisher. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(7):179–188, 1936.

[3] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, second edition, 1990.

[4] G. A. Giraldi, P. S. Rodrigues, E. C. Kitani, J. R. Sato, and C. E. Thomaz. Statistical Learning Approaches for Discriminant Features Selection. *Journal of the Brazilian Computer Society*, 14(2):7–22, 2008.

[5] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, second edition, 2009.

[6] A. K. Jain and B. Chandrasekaran. Dimensionality and sample size considerations in pattern recognition practice. In P. R. Krishnaiah and L. N. Kanal, editors, *Handbook of Statistics*, volume 2, pages 835–855, The Netherlands, North-Holland, 1982. Eds. Amsterdam.

[7] D. B. Kopans. *Imagem da Mama*. Medsi, second edition, 2000.

[8] A. C. Lorena and A. C. P. L. F. Carvalho. Uma Introdução às Support Vector Machines. Technical report, 2007.

[9] P. S. Rodrigues, R.-F. Chang, and J. S. Suri. Non-Extensive Entropy for CAD Systems of Breast Cancer Images. *Computer Graphics and Image Processing, Brazilian Symposium on*, 0(3):121–128, 2006.

[10] P. S. Rodrigues, G. A. Giraldi, R.-F. Chang, and J. S. Suri. *Automatic Classification of Breast Lesions in 3-D Ultrasound Images*. Biomedical Imaging and Bioinformatics, October 2006.

[11] J. R. Sato, A. Fujita, C. E. Thomaz, M. da Graça Morais Martin, J. Mourão-Miranda, M. J. Brammer, and E. A. Junior. Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. *NeuroImage*, 46(1):105 – 114, 2009.

[12] C. E. Thomaz, J. P. Boardman, S. Counsell, D. L. G. Hill, J. V. Hajnal, A. D. Edwards, M. A. Rutherford, D. F. Gillies, and D. Rueckert. A Multivariate Statistical Analysis of the Developing Human Brain in Preterm Infants. *Image and Vision Computing*, 25(6):981–994, 2007.

[13] C. E. Thomaz, E. C. Kitani, and D. F. Gillies. A Maximum Uncertainty LDA-based approach for Limited Sample Size problems - with application to Face Recognition. *Journal of the Brazilian Computer Society*, 12:7–18, September 2006.

[14] V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.

Segmentação de texturas dinâmicas: um novo método baseado em caminhadas determinísticas

Wesley Nunes Gonçalves, Bruno Brandoli Machado, Odemir Martinez Bruno
Instituto de Física de São Carlos (IFSC)
Instituto de Ciências Matemáticas e Computação (ICMC)
Universidade de São Paulo (USP) – São Carlos, SP, Brazil
wnunes@ursa.ifsc.usp.br, brandoli@icmc.usp.br, bruno@ifsc.usp.br

Resumo

Recently, dynamic texture has attracted steadily growing interest due to the number of multimedia databases. Dynamic texture emerged as a new field of research that extends the concept of self-similarity of texture to the spatiotemporal domain. In this paper, we propose a new approach for dynamic texture segmentation based on deterministic walks and k-means algorithm. Although k-means presents interesting results, it only guarantees its convergence to a local minimum, which can affect the segmentation. To overcome this issue, this paper also presents a comparative study of six initialization methods of k-means applied to the dynamic texture segmentation problem. Experimental results on 300 databases demonstrate the effectiveness of the proposed method compared to the state-of-the-art methods.

1. Introdução

Devido ao grande número de base de dados multimídias, as texturas dinâmicas têm atraído um crescente interesse da comunidade de visão computacional. Texturas dinâmicas podem ser definidas como texturas em movimento, no qual o conceito de imagem de textura é estendido para o domínio espaço-temporal. Exemplos de texturas dinâmicas incluem cenas de tráfego, escada rolante, fogo, e fumaça. Diversas aplicações utilizando texturas dinâmicas foram reportadas na literatura, como recuperação de vídeos [12, 11], modelagem de músicas [3], análise de movimento [15], análise de tráfego de carros [5], entre outras. Assim, existe uma crescente demanda por métodos efetivos e eficientes para a modelagem de texturas dinâmicas.

Grande parte das aplicações em visão computacional requer uma etapa de segmentação de imagens ou vídeos. Um caso particular, a segmentação de texturas dinâmicas consiste em dividir o vídeo em regiões homogêneas no espaço e no tempo. A maioria dos métodos existentes para segmentação de texturas dinâmicas são baseados em sistemas dinâmicos [8]. Assim, a principal diferença dos métodos da literatura está na manipulação desse sistema. Em [8], a segmentação de texturas dinâmicas é abordada através de uma formulação *level-set*. Descritores *Ising* de segunda-ordem são utilizados em [9] para segmentar vídeos por meio da minimização da variância espaço-temporal. Recentemente, Chan [6] propôs um método que estima os parâmetros de k sistemas dinâmicos lineares através do bem conhecido algoritmo *Expectation Maximization* (EM). Uma versão aprimorada foi proposta em [4], o qual representa as texturas dinâmicas como uma coleção de camadas estocásticas. Cada camada modela uma textura dinâmica amostrada de um sistema dinâmico linear.

Neste artigo, nós propomos um método para segmentação de texturas dinâmicas baseado em caminhadas determinísticas [2] e no bem conhecido algoritmo *k-means*. Para modelar características de movimento e aparência presentes no vídeo, caminhadas determinísticas são executadas em três planos ortogonais da sequência de imagens: planos XY, XT, e YT. O primeiro plano modela características de aparência, enquanto os outros dois planos modelam características de movimento. Após a extração de características, os vetores obtidos para cada pixel são agrupados utilizando o algoritmo *k-means*. Embora o *k-means* forneça resultados robustos, é amplamente reportado na literatura que o seu desempenho depende da escolha inicial dos centróides. Para superar essa limitação, experimentos utilizando seis métodos de inicialização foram conduzidos em 300 bases de dados. Cada base consiste de um vídeo composto por três ou quatro texturas dinâmicas. Resultados experimentais ilustram que nosso método supera os principais métodos existentes, pois faz a modelagem de textura dinâmica mais eficaz e independente da escolha inicial de centróides.

O restante do artigo é organizado como segue. Seção 2 descreve como aplicar caminhadas determinísticas em vídeos. Na Seção 3, nós apresentamos o método proposto para segmentação de texturas dinâmicas. Seção 4 descreve os principais métodos de inicialização para o algoritmo *k-means*. Experimentos e resultados são apresentados na Seção 5, o qual é seguida pelas conclusões na Seção 6.

2. Caminhadas Determinísticas em Sequência de Imagens

Recentemente, um novo método para análise de texturas em imagens foi proposto em [2]. Este método considera agentes independentes caminhando nos pixels da imagem. Dado que um agente está no pixel p_i, ele se move para um dos seus pixels vizinhos p_j que minimiza a diferença de intensidades $w_{ij} = |I(p_i) - I(p_j)|$, onde $I(p_i)$ é o nível de cinza do pixel p_i. Além disso, o agente evita retornar para pixels visitados recentemente armazenados em uma memória M de tamanho μ. Considerando $\eta(p_i)$ como o conjunto de pixels vizinhos de p_i, os passos iterativos do agente são descritos na Equação 1 para o critério de movimento $din = min$. O agente pode também caminhar em direção da máxima diferença de intensidade ($din = max$) de acordo com a Equação 2.

$$p_j = \arg\min_{j=1}^{N}\{w_{ij} \,|\, p_j \notin M \text{ e } p_j \in \eta(p_i)\} \qquad (1)$$

$$p_j = \arg\max_{j=1}^{N}\{w_{ij} \,|\, p_j \notin M \text{ e } p_j \in \eta(p_i)\} \qquad (2)$$

Após alguns passos, o agente produz uma trajetória que pode ser dividida em duas partes: uma parte de transiente inicial t que termina em um ciclo de período p chamado atrator. O atrator consiste em um grupo de pixels cujas intensidades formam um caminho que o agente não consegue escapar. Figura 1 apresenta uma caminhada determinística. Os pixels em verde representam o transiente $t = 3$, enquanto o atrator $p = 6$ é representado pelos pixels em laranja. Para modelar uma imagem de textura, cada pixel é utilizado como uma condição inicial de uma trajetória. O transiente e atrator de cada trajetória são combinados em uma distribuição conjunta $S_{\mu,din}(t,p)$. A distribuição conjunta define a probabilidade de uma trajetória possuir transiente t e atrator p (Equação 3). A partir do estudo dessas distribuições, é possível obter características capazes de discriminar imagens de textura [2].

$$S_{\mu,din}(t,p) = \frac{1}{N}\sum_{i=1}^{N}\begin{cases} 1, & \text{se } t_i = t, p_i = p \\ 0, & \text{caso contrário} \end{cases} \qquad (3)$$

onde t_i e p_i são respectivamente o transiente e atrator de uma trajetória iniciada em um pixel i.

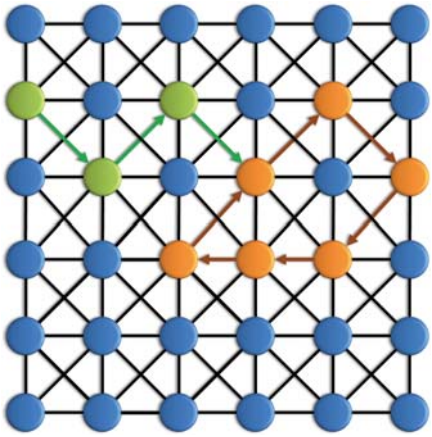
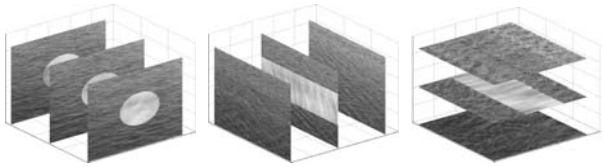


Figura 1. Exemplo de uma caminhada com transiente $t = 3$ e atrator $p = 6$.



(a) Plano XY. (b) Plano XT. (c) Plano YT.

Figura 2. Três planos obtidos do vídeo.

Para aplicar a idéia descrita acima na modelagem de texturas dinâmicas, considere $I(p_i) \,|\, p_i = (x_i, y_i, t_i)$ como a intensidade de um pixel p_i do vídeo, onde x_i e y_i são os índices espaciais e t_i é o índice temporal. Para modelar características de movimento e aparência, nós aplicamos caminhadas determinísticas em três planos ortogonais do vídeo. Os três planos, denominados planos XY, XT e YT, podem ser visualizados na Figura 2.

Os planos visam definir o conjunto de pixels vizinhos $\eta(p_i)$ e consequentemente o comportamento do agente. O plano XY constrói uma vizinhança tal que o agente só pode caminhar em pixels pertencentes ao mesmo frame, caracterizando a aparência dos frames do vídeo. Por outro lado, os planos XT e YT contem características de movimento do vídeo. A definição da vizinhança dos pixels para os planos XY, XT e YT são apresentados nas Equações 4, 5 e 6, respectivamente.

$$p_j \in \eta(p_i) \text{ if } \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq \sqrt{2} \text{ e } t_i = t_j \qquad (4)$$

$$p_j \in \eta(p_i) \text{ if } \sqrt{(x_i - x_j)^2 + (t_i - t_j)^2} \leq \sqrt{2} \text{ e } y_i = y_j \quad (5)$$

$$p_j \in \eta(p_i) \text{ if } \sqrt{(y_i - y_j)^2 + (t_i - t_j)^2} \leq \sqrt{2} \text{ e } x_i = x_j \quad (6)$$

Para combinar características de aparência e movimento, caminhadas determinísticas são aplicadas separadamente em cada plano descrito acima utilizando a nova definição de vizinhança. Assim, três distribuições conjuntas são obtidas: $S^{XY}_{\mu,din}(t,p)$, $S^{XT}_{\mu,din}(t,p)$ e $S^{YT}_{\mu,din}(t,p)$.

2.1. Vetor de Características

A partir da distribuição conjunta, um histograma $h^{\phi}_{\mu,din}(t+p)$ é construído para sumarizar essas informações (ϕ representa um dos três planos). Esse histograma representa o número de trajetórias que possuem tamanho igual a $(t + p)$. Em seguida, n características são selecionadas para compor o vetor $\psi^{\phi}_{\mu,din}$ – Equação 7.

$$\psi^{\phi}_{\mu,din} = [h^{\phi}_{\mu,din}(\mu + 1), \ldots, h^{\phi}_{\mu,din}(\mu + n)] \quad (7)$$

A distribuição conjunta depende do tamanho da memória μ e critério de movimento din. Para obter informações de diferentes escalas e fontes, um vetor φ^{ϕ} considerando diferentes valores de μ é apresentado na Equação 8. Esse vetor de característica é composto pela concatenação dos vetores $\psi^{\phi}_{\mu_i,din}$.

$$\varphi^{\phi} = \left[\psi^{\phi}_{\mu_1,din}, \psi^{\phi}_{\mu_2,din}, \ldots, \psi^{\phi}_{\mu_M,din}\right] \quad (8)$$

onde din é o critério de movimento no qual pode ser max ou min.

A utilização de múltiplas memórias é uma estratégia efetiva para caracterização de padrões de textura e possui um grande potencial como método de classificação de imagens [2]. Cada memória está relacionada com um escala da imagem, tornando o método multi-escala. Para obter características de aparência e movimento, uma estratégia combinando diferentes planos ortogonais é também utilizada. Assim, o vetor de característica φ (Equação 9) contem características extraídas dos três planos: XY, XT e YT.

$$\varphi = \left[\varphi^{XY}, \varphi^{XT}, \varphi^{YT}\right] \quad (9)$$

3. Segmentação de Texturas dinâmicas

Nesta seção, nós descrevemos o método proposto para segmentação de texturas dinâmicas baseado em caminhadas determinísticas e algoritmo *k-means*. O método proposto pode ser sumarizado pelos seguintes passos: (1) vídeo é dividido em blocos; (2) Para cada bloco, um vetor de característica é obtido utilizando caminhadas determinísticas aplicadas em três planos ortogonais; (3) Algoritmo *k-means* agrupa os vetores, resultando na segmentação do vídeo.

3.1. Blocos

Dado um vídeo de T frames com $w \times h$ pixels, ele é dividido em $N = (w \times h \times T)$ blocos. Para cada pixel j, um bloco β_j com dimensão $p \times p \times q$ centrado no pixel j é extraído para compor o conjunto de blocos. É importante notar que p e q devem ser grandes o suficiente para modelar as características de textura e movimento, e ao mesmo tempo, pequeno o suficiente para não misturar texturas dinâmicas diferentes. Se os limites das texturas não mudarem com o tempo, a dimensão temporal q pode ser setada com a duração do vídeo - $q = T$.

3.2. Extração de Características

Para cada bloco β_j, um vetor de características φ_j é extraído usando caminhadas determinísticas descritas na seção anterior. Seguindo [2], nós utilizamos os seguintes parâmetros: critério de movimento $din = max$, 5 memórias ($\mu = 1$ a 5), $n = 5$ descritores do histograma e os três planos XY, XT, e YT. Com esses parâmetros, um vetor de características 75-dimensional é obtido para cada pixel.

3.3. Algoritmo K-means

Após a extração de características, os vetores φ_j são agrupados por meio do algoritmo *k-means*. O *k-means* é um algoritmo de agrupamento bem estabelecido que é computacionalmente eficiente e não requer a especificação de muitos parâmetros. Basicamente, o *k-means* é um algoritmo iterativo que minimiza as distâncias intra-grupos de cada centróide $\overline{x_i}$ movendo-o repetitivamente para a média aritmética do seu conjunto Voronoi.

4. Métodos de Inicialização

Embora o *k-means* seja muito utilizado em problemas de visão computacional, é amplamente reportado na literatura que o seu resultado depende da escolha inicial dos centróides. Devido a essa limitação, o *k-means* pode convergir para uma partição que é significantemente inferior ao ótimo global. Diversas heurísticas foram propostas para encontrar inicializações robustas do algoritmo. Os tópicos seguintes descrevem os principais métodos de inicialização.

Aleatório: Apesar das limitações e por geralmente fornecer uma solução local, o método de inicialização aleatório é um dos mais utilizados devido a sua simplicidade. Neste método, k vetores da base são aleatoriamente escolhidos

como centróides. Cada vetor da base possui a mesma probabilidade de ser escolhido como um centróide inicial.

Astrahan: Astrahan [1] propôs um método de inicialização baseado em densidade. Primeiro, ele define um raio de vizinhança $r = \frac{1}{n(n-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d(x_i, x_j)$. Então, calcula-se a densidade de cada vetor da base. A densidade de um vetor x_i é o número de vetores que estão a uma distância menor que r. O vetor com maior densidade é escolhido como o primeiro centróide. Os $k-1$ centróides restantes são escolhidos em ordem decrescente de densidade, desde que eles estejam a uma distância mínima r dos outros centróides já escolhidos.

KKZ: Uma estratégia baseada em distância de pares de vetores foi proposta por Katsavounidis [10]. A idéia do método KKZ é selecionar vetores que estejam mais distantes um dos outros, seguindo a premissa que esses vetores provavelmente pertencem a grupos diferentes. Neste método, o primeiro centróide é aquele que possui a maior norma ($\overline{x_1} = \arg\max ||x_j||$, onde $||.||$ é a norma de um vetor). Então, calcula-se a menor distância d_j entre cada vetor x_j e os centróides iniciais já definidos. O vetor com a maior distância é adicionado ao conjunto de centróides iniciais. Repetem-se os passos descritos acima até que o número de centróides iniciais seja igual a k.

Múltiplas Execuções do K-means - ME: Uma estratégia comum que apresenta soluções interessantes é inicializar o *k-means* aleatoriamente M vezes e selecionar os centróides que forneceram o melhor agrupamento. Seguindo [14], o índice de silhueta é escolhido como critério para avaliar a solução obtida pelo *k-means* porque ela fornece alta correlação com a partição ideal dos dados.

PCA: O método PCA [13] é baseado na aplicação iterativa da análise de componentes principais - PCA. Inicialmente, todos os vetores pertencem ao mesmo grupo. Então, escolhe-se o grupo C_i (inicialmente somente um grupo está disponível) com a maior variância intra-grupo para realizar a divisão. Para a divisão do grupo, os vetores $x_j \in C_i$ são projetados em direção da primeira componente resultante do PCA. Os vetores projetados são denominados y_j. Em seguida, o grupo C_i é dividido em dois sub-grupos $C_i^{(1)}$ e $C_i^{(2)}$ de acordo com a regra: se $y_j \leq \overline{y}$, adiciona x_j para $C_i^{(1)}$; senão, adiciona x_j para $C_i^{(2)}$. Repetem-se os passos acima até que o número de grupos seja igual a k. Os centróides iniciais são os centróides de cada um dos grupos.

Inicialização Manual: Em segmentação de imagens, é comum a realização de segmentação semi-supervisionada. O usuário manualmente divide as regiões e o algoritmo de segmentação utiliza essa segmentação inicial para refinar a solução. Neste trabalho, as bases de dados fornecem uma segmentação inicial que é utilizada nesse método de inicialização do *k-means*.

Métodos	$k = 3$	$k = 4$
Manual	0.95($\pm$0.08)	0.94($\pm$0.07)
ME	0.93($\pm$0.10)	0.92($\pm$0.10)
PCA	0.83($\pm$0.16)	0.85($\pm$0.11)
KKZ	0.79($\pm$0.15)	0.78($\pm$0.11)
Astrahan	0.79($\pm$0.15)	0.83($\pm$0.12)
Aleatório	0.77($\pm$0.15)	0.77($\pm$0.10)

Tabela 1. Comparação de métodos de inicialização para bases com k = 3 e 4.

5. Experimentos e Resultados

Para avaliar o método proposto, experimentos foram conduzidos em 300 bases de dados, cada uma composta por um vídeo de texturas dinâmicas obtidas em [6]. Os vídeos possuem 60 frames com resolução de 160×110 pixels e são compostos por $k = 3$ ou 4 grupos (texturas dinâmicas). Cada grupo foi aleatoriamente selecionado de um conjunto de 12 texturas dinâmicas, o qual incluem grama, mar, água fervendo, escada rolante, fogo, fumaça, agua e plantas em movimento. Em todos os experimentos, blocos de dimensão $p = 9$ e $q = 60$ foram utilizados para extrair vetores de características utilizando o método proposto.

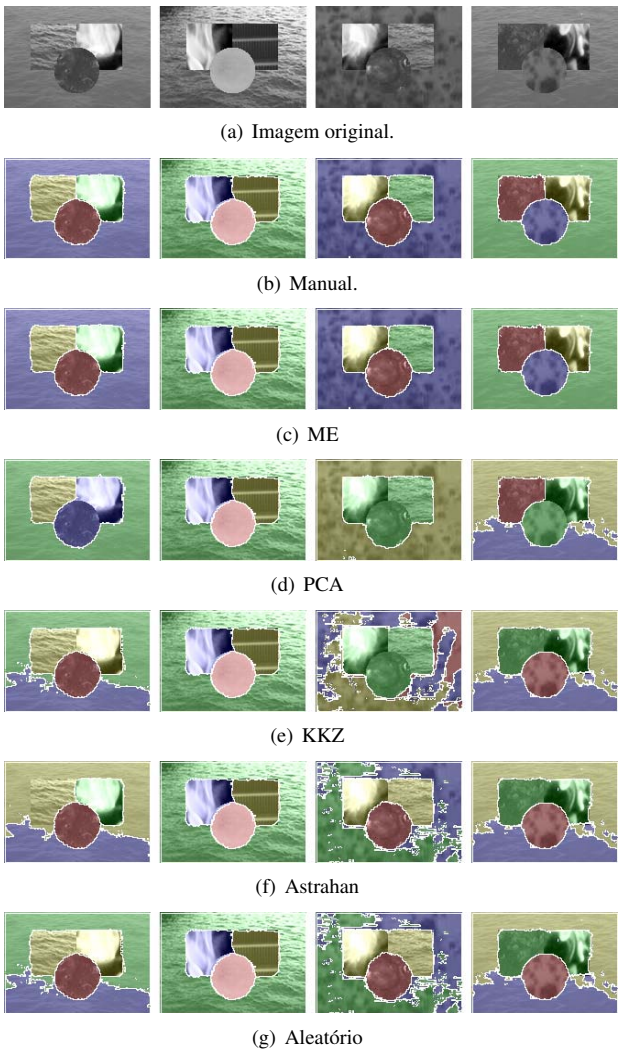
5.1. Análise dos Métodos de Inicialização

Nesta seção, nós avaliamos a qualidade da segmentação final obtida pelo *k-means* quando inicializado com cada um dos seis métodos de inicialização. Para quantificar os resultados de segmentação, *rand index* [14] foi utilizado em todos os experimentos. *Rand index* avalia o grau de concordância entre a solução do método proposto e a partição real. Para o método de múltiplas execuções do *k-means* (ME), o número de iterações foi setado em 50.

A partição final fornecida pelos diferentes métodos de inicialização são quantitativamente avaliados na Tabela 1. Como podemos ver, a inicialização manual alcançou os maiores valores de *rand index*. Desconsiderando o tempo computacional devido as múltiplas iterações, o método ME alcançou resultados comparáveis ao método de inicialização manual. Assim, o método ME pode ser utilizado como um método automático para inicialização do *k-means* em segmentação de texturas dinâmicas. Embora o método PCA não tenha fornecido resultados comparáveis aos dois melhores métodos, ele pode ser considerado um método alternativo para segmentação de texturas dinâmicas em tempo real. Os demais métodos apresentaram resultados similares ao método aleatório, com uma pequena vantagem para o método Astrahan.

Para estatisticamente comparar os métodos de inicialização, teste de Friedman [7] seguido pelo teste post-hoc Nemenyi foi realizado no *rand index* com $\alpha = 0.05$. Tabela 2 apresenta uma comparação entre pares de métodos de inicialização para base de dados com 3 e 4 grupos. O símbolo de cada posição s_{ij} da tabela possui os seguintes significados: $\odot$- não houve diferença significativa entre os métodos i e j, $\triangle$ - houve diferença significativa com o método i sendo melhor que o método j, $\triangledown$ - houve diferença significativa com o método i sendo pior que o método j.

De acordo com os resultados, o método de inicialização manual (1) e ME (2) não apresentaram diferenças significativas entre eles. Entretanto, eles apresentaram resultados significativamente superiores aos demais métodos de inicialização. Para as bases com $k = 3$, método PCA (3), Astrahan (4), KKZ (5) e Aleatório (6) não apresentaram resultados com diferença significativa. Entretanto, para as bases com $k = 4$, método PCA alcançou resultados significativamente superiores aos resultados obtidos pelos métodos KKZ e Aleatório. Como ilustração, a segmentação final obtida por diferentes métodos de inicialização é apresentada na Figura 3.



(a) Imagem original.

(b) Manual.

(c) ME

(d) PCA

(e) KKZ

(f) Astrahan

(g) Aleatório

Figura 3. Resultado de segmentação.

5.2. Comparação com Métodos da Literatura

Além dos resultados dos métodos de inicialização, nós comparamos o método proposto com os três principais métodos de segmentação de texturas dinâmicas: Layered Dynamic Texture - LDT [4], Multiple Dynamic Texture - MDT [6] e Generalized Principal Component Analysis - GPCA [15]. Nestes experimentos, os métodos da literatura foram inicializados manualmente.

Na Tabela 3, nós apresentamos a média e desvio padrão do *rand index* para os métodos de segmentação de texturas dinâmicas. Como podemos ver, o método proposto utilizando inicialização manual fornece os melhores resultados para as bases com $k = 3$ e 4. Além disso, o método proposto utilizando inicialização automática - ME obteve resultados excelentes comparado com os demais métodos que foram inicializados manualmente. Estes resultados experimentais demonstraram que nosso método é um método de segmentação de textura dinâmica eficaz comparado com o estado-da-arte.

Métodos	$k = 3$	$k = 4$
Método proposto - Manual	0.95($\pm$0.08)	0.94($\pm$0.07)
Método proposto - ME	0.93($\pm$0.10)	0.92($\pm$0.10)
LDT	0.89($\pm$0.13)	0.92($\pm$0.10)
MDT	0.85($\pm$0.13)	0.87($\pm$0.11)
GPCA	0.55($\pm$0.11)	0.55($\pm$0.07)

Tabela 3. Comparação entre métodos de segmentação de texturas dinâmicas.

A Figura 4 apresenta o resultado de segmentação do método proposto em uma aplicação real. Nesse exemplo, o método foi aplicado com $k = 2$ e 3. Para $k = 3$, é possível notar que o método consegue segmentar regiões do vídeo com velocidades diferentes.

$k=3$						
	1	2	3	4	5	6
1	⦿	⦿	△	△	△	△
2	⦿	⦿	△	△	△	△
3	▽	▽	⦿	⦿	⦿	⦿
4	▽	▽	⦿	⦿	⦿	⦿
5	▽	▽	⦿	⦿	⦿	⦿
6	▽	▽	⦿	⦿	⦿	⦿

$k=4$						
	1	2	3	4	5	6
1	⦿	⦿	△	△	△	△
2	⦿	⦿	△	△	△	△
3	▽	▽	⦿	⦿	△	△
4	▽	▽	⦿	⦿	⦿	⦿
5	▽	▽	▽	⦿	⦿	⦿
6	▽	▽	▽	⦿	⦿	⦿

Tabela 2. Comparação entre pares de métodos paras as bases com k=3 e 4.

(a) (b) (c)

Figura 4. Exemplos de segmentação em (a) imagem real utilizando (b) k = 2 e (c) k = 3.

6. Conclusão

Neste artigo, nós propomos um novo método para segmentação de texturas dinâmicas baseado em caminhadas determinísticas e algoritmo *k-means*. Nesse método, um vetor de características é extraído para cada pixel utilizando caminhadas determinísticas e então, esses vetores são agrupados com o algoritmo *k-means*. O algoritmo *k-means* é dependente da inicialização dos centróides. Assim, este artigo também apresentou um estudo comparativo de seis métodos de inicialização aplicados ao problema de segmentação de texturas dinâmicas.

Resultados promissores foram obtidos pelo método de inicialização manual e múltiplas execuções do *k-means*. Apesar dos resultados inferiores, nós sugerimos o método de inicialização PCA para aplicações em tempo real. Além disso, resultados experimentais indicaram que o nosso método supera os mais importantes métodos da literatura para segmentação de texturas dinâmicas.

Agradecimentos. Este trabalho recebeu apoio da FAPESP (2010/08614-0) e CNPq (306628/2007-4 e 484474/2007-3).

Referências

[1] M. M. Astrahan. Speech analysis by clustering or the hyperphome method. In *Stanford Artificial Intelligence Project Memorandum AIM-124*. Stanford University, 1970.

[2] A. R. Backes, W. N. Gonçalves, A. S. Martinez, and O. M. Bruno. Texture analysis and classification using deterministic tourist walk. *Pattern Recognition*, 43(3):685–694, 2010.

[3] L. Barrington, A. B. Chan, and G. Lanckriet. Modeling music as a dynamic texture. *Trans. Audio, Speech and Lang. Proc.*, 18(3):602–612, 2010.

[4] A. B. Chan, M. Morrow, and N. Vasconcelos. Analysis of crowded scenes using holistic properties. In *Performance Evaluation of Tracking and Surveillance workshop at CVPR 2009*, pages 101–108, Miami, Florida, 2009.

[5] A. B. Chan and N. Vasconcelos. Classification and retrieval of traffic video using auto-regressive stochastic processes. In *IEEE Intelligent Vehicles Symposium*, pages 771–776, June 2005.

[6] A. B. Chan and N. Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(5):909–926, May 2008.

[7] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal Machine Learning Research*, 7:1–30, December 2006.

[8] G. Doretto, D. Cremers, P. Favaro, and S. Soatto. Dynamic texture segmentation. In *ICCV*, pages 1236–1242, 2003.

[9] A. Ghoreyshi and R. Vidal. Segmenting dynamic textures with ising descriptors, arx models and level sets. In *Dynamical Vision Workshop in the European Conference on Computer Vision*, pages 127–141, 2006.

[10] I. Katsavounidis, C. Jay, and Z. Zhang. A new initialization technique for generalized lloyd iteration. *Signal Processing Letters, IEEE*, 1(10):144–146, 1994.

[11] R. Peteri and D. Chetverikov. Qualitative characterization of dynamic textures for video retrieval. In *In Proc. International Conference on Computer Vision and Graphics*, volume 32, pages 33–38, 2006.

[12] A. Ravichandran and R. Vidal. Video registration using dynamic textures. In *ECCV*, volume II, pages 514–526, 2008.

[13] T. Su and J. Dy. A deterministic method for initializing k-means clustering. In *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence*, pages 784–786. IEEE Computer Society, 2004.

[14] L. Vendramin, R. J. G. B. Campello, and E. R. Hruschka. On the comparison of relative clustering validity criteria. In *SDM*, pages 733–744. SIAM, 2009.

[15] R. Vidal and D. Singaraju. A closed form solution to direct motion segmentation. In *CVPR*, pages II: 510–515, 2005.

Uma Abordagem para a Localização e o Reconhecimento de Placas de Licenciamento Veicular por meio de Operadores Morfológicos e Busca por *Template*

Danilo de Sousa Abreu
Universidade Nove de Julho
Diretoria dos Cursos de Informática
danilo.danilosousa@gmail.com

Wonder Alexandre Luz Alves
Universidade Nove de Julho
Diretoria dos Cursos de Informática
wonder@uninove.br

Sidnei Alves de Araújo
Universidade Nove de Julho
Programa de Mestrado em
Engenharia de Produção
saraujo@uninove.br

André Felipe Henriques Librantz
Universidade Nove de Julho
Programa de Mestrado em
Engenharia de Produção
librantz@uninove.br

Abstract

In this work is presented a method for location and recognition of vehicle license plates. The proposed method consists of two major stages: (1) extraction of license plate region *and (2)* recognition of license plate*. The first stage consists in extracting the license plate region from the input image, using a robust scheme of morphological operators. In the second stage is performed a generation process of license plates candidates using template matching technique and, after that, the license plate recognition is determined by using heuristics. Preliminary results show that the method can be a good alternative for applications involving the recognition of vehicle license plates.*

Resumo

Neste trabalho é apresentado um método para localização e reconhecimento de placas de licenciamento veicular. O método proposto consiste de duas etapas principais: (1) extração da região da placa *e (2)* reconhecimento da placa*. Na primeira etapa, extrai-se a região da placa na imagem de entrada, utilizando um esquema robusto de operadores morfológicos. Na segunda etapa é realizado um processo de geração de placas candidatas por meio de busca por template, para depois, determinar a placa reconhecida utilizando heurísticas. Resultados preliminares mostram que o método proposto pode ser uma boa alternativa para aplicações que envolvam o reconhecimento de placas de licenciamento veicular.*

1. Introdução

O reconhecimento de placas de licenciamento (LPR, *License Plate Recognition*) em imagens e vídeo é um problema importante em visão computacional e está presente em muitas aplicações, tais como: controle de tráfego [17], controle de caixa em estacionamento [18], entre outras [11, 13, 15]. Consequentemente, muitas metodologias foram desenvolvidas e propostas ao longo das últimas décadas [15, 8, 13, 10, 11, 12] para a localização e o reconhecimento de placas de licenciamento veicular em imagens e vídeos.

A despeito de tais estudos, ainda não é fácil projetar um método geral de LPR. Isso ocorre porque existem muitas variações que tornam o problema de LPR difícil, uma vez que as imagens contendo a placa de licenciamento veicular podem apresentar diversas variações em relação aos modelos das placas e as condições de tempo e iluminação. Além disso, pode haver situações de oclusão devido a defeitos físicos ou problemas de posicionamento da placa [11, 8, 15]. De forma geral, o problema de LPR pode ser composto por três subproblemas: **(1)** localização da placa, **(2)** segmentação dos caracteres da placa e **(3)** reconhecimento da placa.

Dadas essas considerações, diversas técnicas vêm sendo empregadas com frequência nos subproblemas de LPR. Técnicas como operadores morfológicos [8, 3], análise de regiões conexas [11], análise por textura [14], vêm sendo utilizadas no subproblema de localização de placa. No subproblema de segmentação dos caracteres da placa são utilizadas as técnicas de projeção horizontal [5], filtro ho-

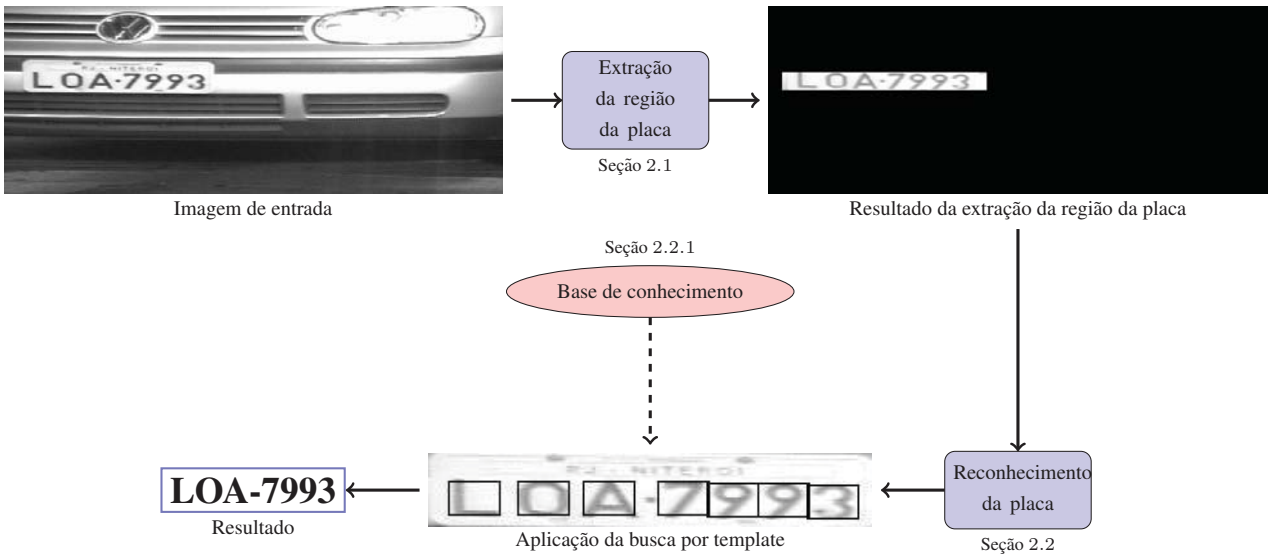


Figura 1. Visão geral do método proposto.

momórfico [3], entre outras. Por fim, no subproblema de reconhecimento da placa têm-se utilizado busca por *template* [8], máquina de suporte vetorial [6], redes neurais artificiais [7], entre outras técnicas de reconhecimento de padrões.

Neste trabalho é proposto um método para localizar e reconhecer placas de licenciamento de veículos em imagens, por meio de operadores morfológicos e busca por *template*. Primeiramente, extrai-se a região da placa na imagem de entrada, depois utiliza-se a busca por *template* para o reconhecimento dos caracteres da placa para e por fim, utiliza-se heurísticas para determinar a placa.

O restante deste trabalho está organizado da seguinte forma: na Seção 2 é apresentada a abordagem proposta para a localização e o reconhecimento de placas de licenciamento veicular. Na Seção 3 são descritos os experimentos realizados para validar o método proposto. E finalmente na Seção 4, são apresentadas algumas conclusões deste trabalho.

2. Método proposto

O método proposto consiste de duas etapas principais (ver Figura 1), sendo: (1) *extração da região da placa* e (2) *reconhecimento da placa*. Na primeira etapa, Seção 2.1, o objetivo é extrair a região da placa na imagem de entrada, por meio de uma combinação de operadores morfológicos. Na próxima etapa, Seção 2.2, é realizado o reconhecimento dos caracteres presentes na região da placa, para depois, determinar o resultado do reconhecimento da placa utilizando heurísticas.

2.1. Extração da região da placa

A extração da região da placa da imagem de entrada é feita por meio da aplicação de uma serie de operadores da Morfologia Matemática (MM). A MM refere-se a um ramo do processamento e da análise de imagens. Seu principal objetivo é obter informações significativas de uma imagem com base na análise da sua geometria e da forma de seus objetos [19, 4].

Para extrair a região da placa de uma imagem de entrada f, é proposto um esquema morfológico [2, 1], com base em hipóteses que levam em conta o contraste e a geometria das regiões dos caracteres da placa. Tais hipóteses estão classificadas a seguir.

1. **Contraste**
   (a) Existe contraste entre as regiões dos caracteres da placa em relação aos seus *backgrounds*.
   (b) Os níveis de cinza dos caracteres de uma mesma placa são similares
2. **Geometria da fonte**
   (a) Os caracteres de uma mesma placa têm suas larguras e alturas similares
   (b) Os caracteres de uma mesma placa estão alinhados ao longo de uma reta
   (c) A distância entre quaisquer dois caracteres adjacentes em uma mesma placa são similares

Primeiramente é aplicado um filtro de abertura e fechamento *top-hat* ([19, 4]) na imagem de entrada f, utilizando um elemento estruturante (EE) disco com raio equivalente a

espessura média dos caracteres da placa, gerando assim as imagens de saída f_w e f_c, respectivamente. Depois toma-se o máximo entre entre as imagens f_w e f_c, obtendo a imagem de saída f_m e em seguida, binariza-se a imagem f_m pelo método de Otsu [16] gerando a imagem f_b. A imagem f_b contém todas as componentes conexas menores ou igual a espessura dos caracteres da placa. Depois é aplicado na imagem f_b aberturas baseadas em atributos de altura, largura e área (produzindo a imagem f_a) para selecionar as regiões candidatas dos caracteres da placa. Após essas operações, uma dilatação por um EE no formato de linha é aplicada na imagem f_a obtendo a imagem f_d contendo as potenciais regiões de placas. Por fim, extrai-se a região da placa da imagem f_d por meio de abertura com base em atributo de retangularidade [2]. Na Figura 2 é apresentado um exemplo da extração da região da placa.

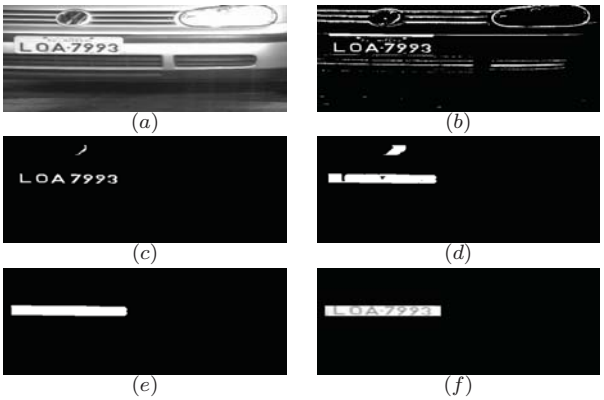


Figura 2. Exemplos de imagens geradas na etapa de extração da região da placa. (a) imagem de entrada f; (b) imagem binária f_b; (c) imagem f_a produzida após a abertura com base em atributos; (d); imagem f_d após a aplicação da dilatação que forma as potenciais regiões de placas; (e) imagem obtida após a seleção da região da placa na imagem f_d por meio de abertura por retangularidade; (f) imagem com a região da placa extraída.

2.2. Reconhecimento da placa

Após a extração da região da placa é realizada a etapa de reconhecimento da placa, a qual é composta de duas sub-etapas, a extração de placas candidatas e a validação da placa. Sendo que na primeira sub-etapa é utilizada a busca por *template* para criar um conjunto ζ de placas candidatas e na segunda é definida a placa reconhecida por meio de heurísticas.

2.2.1. Base de conhecimento

Antes de criar o conjunto ζ de placas candidatas foi construído uma base de conhecimento de caracteres de placas para serem utilizados posteriormente na busca por *templates*. A base de conhecimento foi construída a partir da base de imagens catalogada e disponibilizada através site do *Laboratório de Processamento de Sinais e Imagens do Centro Brasileiro de Pesquisas Físicas*. A base foi construída utilizando 57 imagens de veículos sob diversas condições. Estas imagens foram subdivididas em dois grupos, um com as imagens de treinamento e outro com as de teste. A partir das imagens de treinamento foi construído um conjunto de imagens *templates* $\mathcal{Q}$ de forma que os *templates* $q \in \mathcal{Q}$ estão rotulados de $A-Z$ e $0-9$. No total foram extraídos 198 imagens de templates de 29×19 *pixels*, como mostrado na Figura 3. Depois da construção do conjunto $\mathcal{Q}$ foi

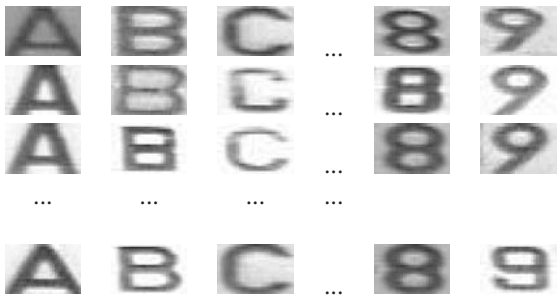


Figura 3. Exemplos de imagens de *templates* que compõem a base de conhecimento.

realizado uma busca exaustiva com intuito de encontrar limiares correspondentes para cada classe de caractere. Estes valores caracterizam-se por números de ponto flutuante entre 0 e 1, que corresponde ao valor de correlação esperado para determinar a região como candidato a um caractere.

2.2.2. Extração de placas candidatas

Para formar o conjunto ζ de placas candidatas utilizou-se a busca por *template* para gerar as possíveis placas candidatas. A Busca por *template* é um método que extrai informações de uma imagem f e de uma imagem *template* q afim de verificar a presença de determinados padrões na imagem f. Assim como em outros processos de reconhecimento de padrões, a Busca por *template* utiliza medidas de similaridades para detectar a presença de determinados padrões na imagem analisada f [9]. Neste trabalho, o coeficiente de correlação foi usado como medida de similaridade. A busca por *template* é aplicada em cada *pixel* da região da placa f_d utilizando os *templates* de uma dada classe de caractere $\mathcal{C}$ presente em $\mathcal{Q}$ da seguinte forma:

$$\mathcal{F}_1(x,y) = \sum_{i=0}^{W}\sum_{j=0}^{H}\Big(f_d(x+i,y+j) - f_m\Big),$$

$$\mathcal{T}_1 = \frac{1}{|\mathcal{C}|}\sum_{i=0}^{W}\sum_{j=0}^{H}\sum_{q\in\mathcal{C}}\Big(q(i,j) - q_m\Big),$$

$$\mathcal{F}_2(x,y) = \sum_{i=0}^{W}\sum_{j=0}^{H}\Big(f_d(x+i,y+j) - f_m\Big)^2,$$

$$\mathcal{T}_2 = \frac{1}{|\mathcal{C}|}\sum_{i=0}^{W}\sum_{j=0}^{H}\sum_{q\in\mathcal{C}}\Big(q(i,j) - q_m\Big)^2,$$

$$\Big[cc(f_d,\mathcal{C})\Big](x,y) = \frac{\mathcal{F}_1(x,y)\cdot\mathcal{T}_1}{\sqrt{\mathcal{F}_2(x,y)}\cdot\sqrt{\mathcal{T}_2}}, \qquad (1)$$

onde f_m é a média dos níveis de cinza da imagem f_d sobre a região em analise; q_m é a média dos níveis de cinza dos *templates* do conjunto $\mathcal{C} \subset \mathcal{Q}$; W e H são, respectivamente, a largura e a altura da região do *template* $q \in \mathcal{C}$.

Assim, a busca por *template* é aplicada em cada *pixel* da região da placa da imagem f_d utilizando cada classe de caractere presente em $\mathcal{Q}$. Cada *pixel* que tiver o valor de correlação igual ou superior ao limiar estabelecido para sua classe é selecionado para compor a placa candidata. Dessa forma, é construído o conjunto das placas candidatas ζ com todas as possíveis placas que podem ser formadas a partir dos caracteres selecionados.

2.2.3. Validação da placa

Esta etapa tem como finalidade selecionar a melhor placa presente ζ por meio de heurísticas. A primeira validação realizada consiste na verificação do caractere encontrado mediante sua localização, ou seja, os primeiros três *templates* selecionados precisam ser de classes do tipo letra. Do mesmo modo, os quatro últimos *templates* selecionados precisam ser de classes do tipo número. O principal objetivo é reduzir a confusão que normalmente ocorre entre "B" e "8", "I" e "1", "O" e "0". A segunda validação é a seleção dos *templates* sobrepostos. Para determinar se os *templates* estão sobrepostos calcula-se a distância Euclidiana entre eles. Caso haja sobreposição escolhe-se o rótulo do *template* de máxima correlação.

Na Figura 4 é apresentado o resultado da aplicação do método proposto em algumas imagens utilizadas para criar a base de *templates*. O método proposto reconheceu 86% das 29 imagens utilizadas para o treinamento. Sendo que dos 203 caracteres presentes nas 29 placas, 195 foi corretamente reconhecido, ou seja 96% de acerto.

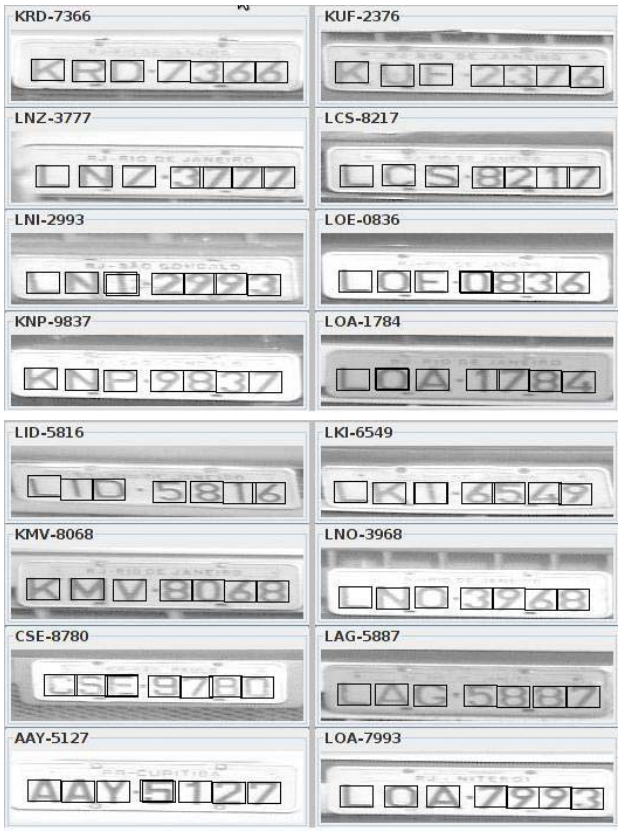


Figura 4. Exemplos da aplicação do método proposto em algumas imagens utilizadas para criar a base de *templates*.

3. Experimentos e resultados

Para avaliar o método proposto, foram utilizadas as 28 imagens do conjunto de teste. Assim como ocorre em outros métodos semelhantes, o método proposto foi avaliado com base nas taxas de acerto e de precisão. A avaliação foi realizada sob três aspectos: (1) a extração da região da placa, (2) o reconhecimento dos caracteres das placa e (3) o reconhecimento da placa.

No aspecto da extração da região da placa o método proposto extraiu corretamente 25 de um total de 28 placas, obtendo assim, uma taxa de acerto de 89% e uma taxa de precisão de 100%. Levando em consideração o aspecto do reconhecimento dos caracteres, o método proposto reconheceu 173 de um total de 175 caracteres das 25 placas extraídas, chegando a uma taxa de acerto e precisão de 98%. Por fim, sob o aspecto do reconhecimento da placa, a abordagem proposta reconheceu 22 das 25 placas extraídas obtendo 88% de taxa de acerto e precisão. Considerando todas as placas do conjunto de testes, o método obteve 78%

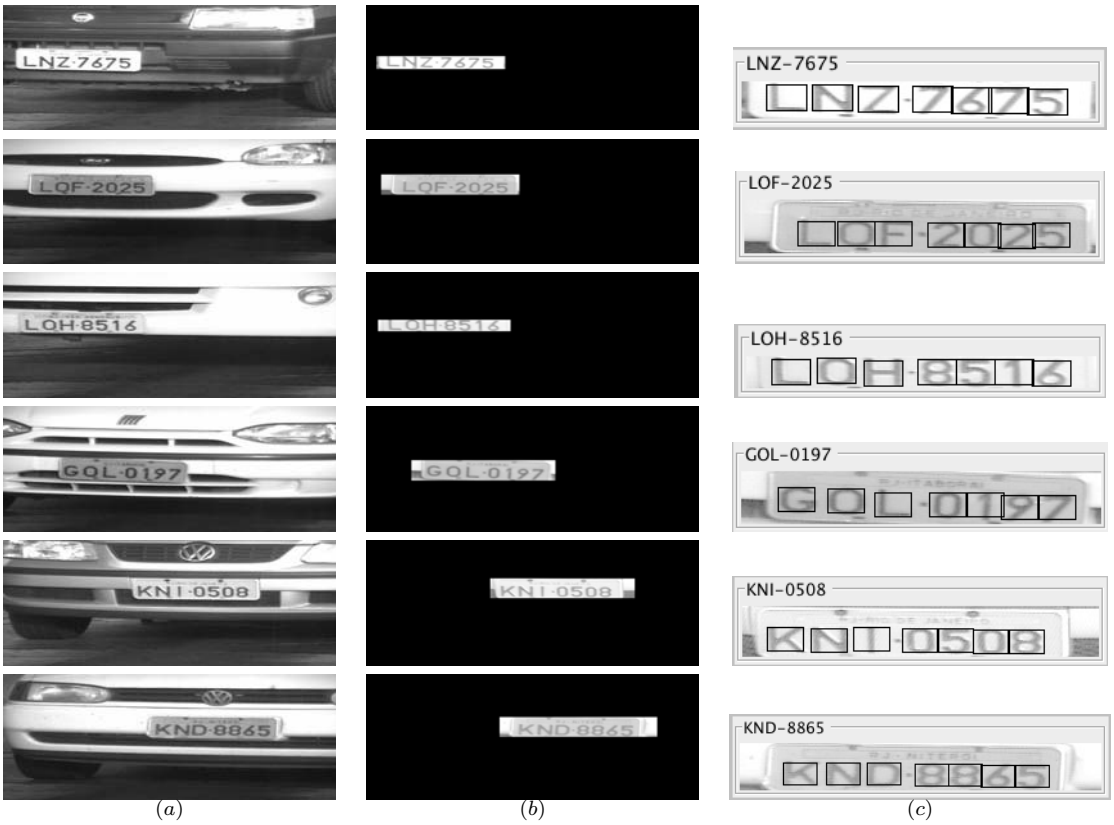


Figura 5. Exemplos da aplicação do método proposto. (a) **imagens de entrada;** (b) **imagens obtidas após etapa de extração da região da placa;** (f) **imagens mostrando as placas segmentadas e reconhecidas.**

de taxa de acerto e 88% de precisão. Na Tabela 1 são apresentados os resultados do método proposto considerando todas as imagens do conjunto de teste.

Tabela 1: Resultados obtidos pelo método proposto.

	Acerto	Precisão
Extração da região da placa	89%	100%
Reconhecimento dos caracteres	88%	98%
Reconhecimento da placa	78%	88%

Na Figura 5 são apresentadas algumas imagens resultantes da aplicação do método proposto.

4. Conclusão

Neste trabalho foi apresentado um método para localização e reconhecimento de placas de licenciamento veicular composto por duas etapas principais: (1) *extração da região da placa* e (2) *reconhecimento da placa*. Em nossos experimentos o método proposto obteve um bom desempenho, sendo que 89% das imagens tiveram suas placas extraídas corretamente, 88% dos caracteres das placas foram reconhecidos e 78% das placas foram corretamente identificadas. Esses resultados evidenciam que a abordagem proposta pode ser uma boa alternativa para aplicações que envolvam o problema de reconhecimento de placas de licenciamento.

Agradecimentos

Os autores agradecem à Universidade Nove de Julho (Uninove) pelo apoio financeiro.

Referências

[1] W. A. L. Alves and R. F. Hashimoto. Classification of regions extracted from scene images by morphological filters in text or non-text using decision tree. In *18th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision - WSCG 2010 Proceedings*,

Plzen, Czech Republic, february 2010. Vaclav Skala - Union Agency.

[2] W. A. L. Alves and R. F. Hashimoto. Text regions extracted from scene images by ultimate attribute opening and decision tree classification. *Graphics, Patterns and Images, SIBGRAPI Conference on*, 0:360–367, 2010.

[3] V. P. de Araújo, R. D. Maia, M. F. S. V. D'Angelo, and G. N. R. D'Angelo. Automatic plate detection using genetic algorithm. In *Proceedings of the 6th WSEAS International Conference on Signal, Speech and Image Processing*, pages 43–48, Stevens Point, Wisconsin, USA, 2006. World Scientific and Engineering Academy and Society (WSEAS).

[4] E. R. Dougherty and R. A. Lotufo. *Hands-on Morphological Image Processing*. SPIE Press, Bellingham, WA, 2003.

[5] X. He, L. Zheng, Q. Wu, W. Jia, B. Samali, and M. Palaniswami. Segmentation of characters on car license plates. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pages 399 –402, oct. 2008.

[6] R. Huang, H. Tawfik, and A. Nagar. License plate character recognition based on support vector machines with clonal selection and fish swarm algorithms. *Computer Modeling and Simulation, International Conference on*, 0:101–106, 2009.

[7] Z. Ji-yin, Z. Rui-rui, L. Min, and L. Yin. License plate recognition based on genetic algorithm. *Computer Science and Software Engineering, International Conference on*, 1:965–968, 2008.

[8] S. Kasaei, S. Kasaei, and S. Monadjemi. A novel morphological method for detection and recognition of vehicle license plates. *American Journal of Applied Sciences*, pages 2066–2070, 2009.

[9] H. Y. Kim and S. A. de Araújo. Grayscale template-matching invariant to rotation, scale, translation, brightness and contrast. In *Proceedings of the 2nd Pacific Rim conference on Advances in image and video technology*, PSIVT'07, pages 100–113, Berlin, Heidelberg, 2007. Springer-Verlag.

[10] K.-B. Kim and D. S. Kim. Recognition of car license plates using morphological information and som algorithm. *JACIII*, 8(4):385–389, 2004.

[11] H. Kwasnicka and B. Wawrzyniak. License plate localization and recognition in camera pictures. In *Symposium on Methods of Artificial Intelligence*, 2002.

[12] R. Lotufo, A. Morgan, and A. Johnson. Automatic number-plate recognition. *Image Analysis for Transport Applications, IEE Colloquium on*, 1990.

[13] H. Mahini, S. Kasaei, F. Dorri, and F. Dorri. An efficient features - based license plate localization method. In *Proceedings of the 18th International Conference on Pattern Recognition - Volume 02*, ICPR '06, pages 841–844, Washington, DC, USA, 2006. IEEE Computer Society.

[14] L. Matos, Y. T. dos Passos, and R. Maciel. Detecção de placas de licença veicular utilizando segmentação por texturas. *In Workshop de visão computacional*, 2008.

[15] A. Mohsin, A. H. Hassin, and I. Q. AbdulJaleel. An automatic recognizer for iraqi license plates using elman neural network. *Journal of Software Engineering and Applications*, pages 1163–1166, 2010.

[16] N. Otsu. A threshold selection method from graylevel histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.

[17] C. Setchell. *Applications of Computer Vision to Road-traffic Monitoring*. PhD thesis, Department of Computer Science, University of Bristol, September 1997.

[18] T. Sirithinaphong and K. Chamnongthai. The recognition of car license plate for automatic parking system. In *Signal Processing and Its Applications, 1999. ISSPA '99. Proceedings of the Fifth International Symposium on*, volume 1, pages 455 –457 vol.1, 1999.

[19] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2003.

Uma ferramenta visual para analisar as imagens de microscopia de força atômica AFM

Carolina Guillén Valencia[(1)], Marcelo de Assumpção Pereira da Silva[(1) (2)], Roberto Mendonça Faria[(1)], Odemir Martinez Bruno[(1)]
(1)Instituto de Fisica de São Carlos, Universidade de São Paulo
2) UNICEP – Centro Universitário Central Paulista – São Carlos - SP
krolgv@ursa.ifsc.usp.br, maps@ifsc.usp.br, faria@ifsc.usp.br, bruno@ifsc.usp.br

Resumo

A técnica de microscopia de força atômica (AFM) permite o estudo das morfologias dos materiais na escala nanométrica. Um importante material de estudo são os polímeros. Eles tem a habilidade de auto-organização e autoestratificação, notando-se a formação de padrões regulares. A aplicação destas estruturas de poliestireno é considerada como potencial alternativa para área de electroelectrônicos, e outras ramas de tecnologia como a microfluídica e a biotecnologia. Este artigo apresenta uma ferramenta computacional desenvolvida com o propósito de analisar e visualizar as imagens obtidas pela técnica AFM, para um processo de quantificação e modelagem, o que permite o auxílio para a compreensão de fenômenos ligados a elas. Foi conduzido um estudo com a ferramenta das estruturas SEBS (Styrene, Ethylene, Butylene, Styrene), os resultados reportam a eficiência da ferramenta para caracterizar e visualizar as estruturas, em comparação com outras ferramentas.

Palavras-chaves: Análise de imagens, visualização, microscopia de força atômica, SEBS, polímeros, open-source software.

1. Introdução

A técnica de microscopia de força atômica (AFM) faz parte de uma família de instrumentos, usado amplamente no estudo das morfologias dos materiais tais como os semicondutores, polímeros, metais, etc. Os materiais poliméricos sintéticos sofreram um grande desenvolvimento nas áreas laboratorial (síntese e processamento) e industrial (aplicações em setores de bens de consumo). Especialmente a causa de suas propriedades semicondutoras, uma importante linha de aplicação é a tecnologia, mediante os dispositivos de filmes finos. Ao mesmo tempo outros tipos de materiais poliméricos têm aplicações na área de coberturas e revestimentos como tintas, matérias anticorrosivos, etc.

Os polímeros SEBS (Styrene, Ethylene, Butylene, Styrene) tem a habilidade de auto-organização e auto estratificação na escala nanométrica, notando-se uma periodicidade singular e com morfologias similares ás esféricas, cilíndricas, e de lâminas, tendo a influência de parâmetros como: a razão da rigidez maciez, o grau de polimerização, a seletividade do solvente, a temperatura do recozimento, o desmolhado, e o processo de fabricação para obter as amostras, estes são fatores decisivos para seu ordenamento estrutural nos substratos sólidos. E recentemente as soluções obtidas pelo desmolhamento de filmes finos de SEBS, mediante o processo da evaporação do solvente tolueno, apresenta uma formação de padrões em forma de gotas e listras submicrométricas [10]. O tolueno é um solvente apropriado para os polímeros, já que não influencia as interações moleculares iniciais, as ações da dinâmica molecular e a formação final da estrutura dependem das interações específicas entre cada componente dos blocos do copolímero e o subtrato.

A aplicação destas estruturas de poliestireno está sendo consideradas como potencial alternativa para a litografia macia na escala nanométrica, em especial na área de electroelectrônicos e outras ramas de tecnologia como a microfluídica e a biotecnologia [11] [12]. É por isto, nosso interesse em construir uma ferramenta computacional para assistir a análise e visualização das imagens das estruturas microscópicas, para apoiar ao processo de quantificação e modelagem da formação dos padrões regulares na morfologia destes filmes finos, necessárias em muitas pesquisas para aplicações tecnológicas. Assim mesmo visualizar as estruturas, já que a técnica de microscopia de força atômica possibilita a informação para sua análise e reconstrução tridimensional das superfícies com rugosidades extremamente pequenas na escala nanométrica. A ferramenta fornece com diferentes técnicas, em especial para o estudo da morfologia da formação de diversos padrões regulares.

Neste artigo apresentamos o uso da ferramenta desenvolvida, utilizando as imagens das estruturas SEBS [10], para amostrar sua eficácia na extração de medidas e visualização delas, pretendendo contribuir ao estudo dos padrões das nanoestruturas, centrando-se na análise das formações geométricas e morfológicas dos terraços. A ferramenta obtêm importantes dados quantitativos, o que ajuda ao estudo da dinâmica da evolução das estruturas

SEBS, substituindo outros métodos tradicionais providos por outros softwares comerciais de microscopia. Além de poder processar uma grande quantidade de amostras nos dados da imagem de maneira semi automática, para o processo de quantificação e modelagem, o que permite o auxílio para a compreensão de fenômenos ligados a elas.

A continuação na secção 2 ilustra-se os conceptos relacionados ás imagens na microscopia de força atômica, a motivação da ferramenta visual é explicada na secção 3 e seus detalhes são exibidos na secção 4, destaca-se as medidas extraídas na secção 5, as amostras SEBS [6][15] são utilizadas pra mostras os resultados da ferramenta na secção 6, e finalmente na secção 7 temos as conclusões.

2. Imagens AFM

2.1 Técnica AFM

A técnica de microscopia de força atômica (AFM – Atomic Force Microscopy) [1] pertence uma família de instrumentos denominada microscopia de ponta de prova (SPM – Scanning Probe Microscopy) [2]. A SPM contém as técnicas mais atuais [3] para medir as propriedades das superfícies nas dimensões do espaço interatômico, tendo como característica principal a proximidade de uma ponta de prova para a superfície analisada, usando a interação elétrica, magnética ou efeitos térmicos. Na Figura 1 podemos observar os principais componentes do microscópio de ponta de prova SPM: o escâner, o sensor de movimento da ponta, a ponta de prova, a amostra e a interface eletrônica.

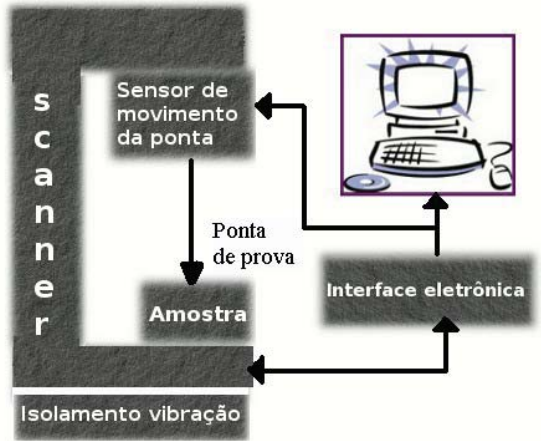


Figura 1. Esquema dos componentes do microscópio de ponta de prova SPM

A ponta de prova faz uma varredura sobre a superfície da amostra (direções x e y), controlando a distância (direção z) entre a superfície e a ponta de prova, em cada ponto da amostra envia as sinais de sua interação com o relevo, força elétrica, força magnética, gradiente térmico, etc para o sensor de movimento da ponta. O sensor transforma os sinais em voltagem o qual é aplicado ao escâner, este é contraído ou dilatado de acordo à quantidade da voltagem. O escâner é um dispositivo composto por algumas cerâmicas piezoelétricas. A interface eletrônica provê uma interface entre o computador e os dados; o escâner e o sensor de movimento da ponta, também controlam o sistema de retro alimentação para manter a força entre a amostra e a ponta de prova. O computador manipula os dados produzidos pelo sistema para seu processamento, visualização e análise.

2.2 Formação da Imagem

A técnica AFM constrói um arquivo o qual contem um cabeçalho ASCII e os dados das imagens. Estes dados contem a informação da fase e relevo, estes são armazenados em dois bytes (precisão de 16 bits). O cabeçalho segue o formato dos instrumentos digitais CIAO (Control Input and Output), e contem vários parâmetros de configuração usados durante a captura da imagem original. Se destacam os parâmetros de soft-scale e hard-scale para poder realizar a respectiva calibração e escala dos dados da imagem, outros parâmetros capturados incluem o nome de arquivo, tipo do instrumento, altura, comprimento, tamanho de varredura (scanner length) o, etc.

O máximo tamanho de varredura das imagens em resolução (bit) e tamanho (pixel) depende da eleição do escâner, podendo escanear até 200µm lateralmente (em x e y) e 8µm verticalmente (em z) veja a Tabela 1, os scanners disponíveis de fabricante Digital Instruments, Veeco Metrology Group de acordo as especificações de MultiMode SPM [4], tendo a maioria de vezes dimensões quadradas, tem três tamanhos de imagens por defeito: 128 x 128 pixeis, 256 x 256 pixeis, 512 x 512 pixeis, e com uma resolução de 16 bits nos três eixos. Depende do número das amostras tomadas pelas linhas ES escaneadas (128,256,512), é que o tamanho do arquivo varia desde 41K até 532K por imagem.

Modelo	Tamanho do escâner	Rango vertical
AS-0.5 (“A”)	0.4µm x 0.4µm	0.4µm
AS-12 (“E”)	10µm x 10µm	2.5µm
AS-12V (“E” vertical)	10µm x 10µm	2.5µm
AS-130 (“J”)	125µm x 125µm	5.0µm
AS-130V (“J” vertical)	125µm x 125µm	5.0µm
AS-200	200µm x 200µm	8.0µm

Tabela 1. Modelos de scanners de Digital Instruments Veeco Metrology Group, USA [4]

A calibração é feita pela informação obtida no cabeçalho, reflexa os dados da ponta de prova sobre a amostra, a qual envia sinais em forma de voltagem ao sensor de movimento da ponta, a relação entre a voltagem (volts) e o deslocamento (nm) em cada ponto é obtida através da calibração do instrumento [5], a precisão dos dados depende do valor correto na sensitividade do escaneado.

A informação obtida pelo aparato digital é numérica e chamada LSB (Least-significant-bit), é a representação digital do vóltio ou frequência. Os dados LSB são convertidos á respectiva escala do relevo e fase, mediantes as equações de calibração respectivas [5]:

$$\text{Re}\,levo = \left(\frac{LSB}{65536}\right) \times ZscaleHeight$$

$$Fase = \left(\frac{LSB}{65536}\right) \times ZscalePhase$$

3. Problema

Sendo a técnica AFM muito recente, existem poucas ferramentas de software e de uso geral. As ferramentas AFM normalmente produzem imagens em duas dimensões (x e y), e um indicador na terceira dimensão. Em consequência, é difícil visualizar e executar uma análise quantitativa para esta dimensão. A ferramenta desenvolvida analisa as microestruturas morfológicas com instruções mais específicas e considera a visualização na terceira dimensão.

Uma das principais motivações foi fazer uma ferramenta para auxiliar o modelamento da dinâmica formação dos blocos dos polímeros [6][15], para quantizar as estruturas internas e estudar os padrões regulares apresentados em sua morfologia. Assim foram implementadas diversas técnicas para fazer uma análise quantitativa nas imagens.

4. Sistema Proposto

A ferramenta foi desenvolvida utilizando a tecnologia Java e as bibliotecas imageJ [20], Jrenderer3D [19] e processing [21] . Tendo como principais fases o processo de leitura, dos filtros nas imagens e a visualização.

4.1 Processo de leitura e geração das imagens

Os arquivos gerados pelo software do instrumento AFM, contem a informação necessária para gerar as imagens da fase e relevo para qualquer linguagem de programação, observa-se na Figura 2 a geração das imagens através da leitura do arquivo, na primeira parte do arquivo de dados, o cabeçalho contem informação relevante: calibração, versão do instrumento, escala, sensitividade do escaneado, números de linhas, quantidade de amostras por linha, tamanho do escaneado. E um dado muito importante chamado offset, o qual indica a posição do início de dados para gerar as imagens do relevo e a fase. A segunda parte contem os dados da imagem, dados com uma precisão de 16 bits e utilizando o tipo numérico short sim precisão para sua representação. Com os dados capturados na primeira parte do arquivo podemos fazer a calibração respectiva para na escala nanométrica [5].

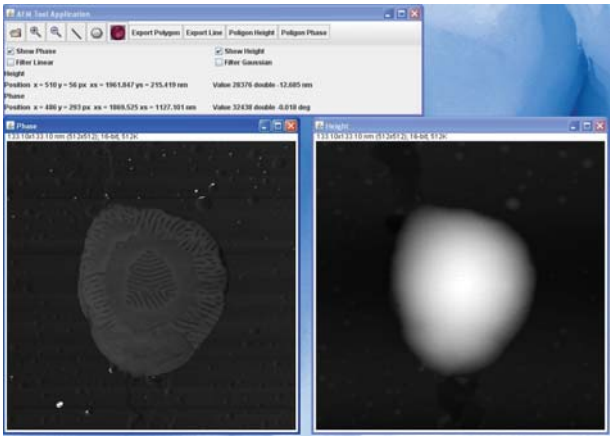


Figura 2. Ferramenta desenvolvida mostrando as imagens de fase e relevo, com sua respectiva calibração

4.2 Processo dos filtros nas imagens

Nesta fase foram utilizados filtros para remover os ruídos, suavizar e realçar a imagem, observa-se na Figura 3 os filtros usados. O filtro linear substitui o valor de cada pixel pelo valor médio de sua vizinhança, tendo como parâmetros o tamanho do filtro (quantidade de pixeis contribuem para o calculo do novo valor do pixel), a forma do filtro (quadrático, circular, etc) e os pesos para os valores da vizinhança (por geral maiores valores perto do centro da região). O filtro matricial é um filtro linear permite ao usuário gerar vários resultados, através da modificação dos parâmetros. O filtro gaussiano também é um filtro linear, o que corresponde à função gaussiana em duas dimensões. O comportamento ideal de um filtro para suavizar a imagem é ter um comportamento isotrópico (uniforme em cada direção) [8].

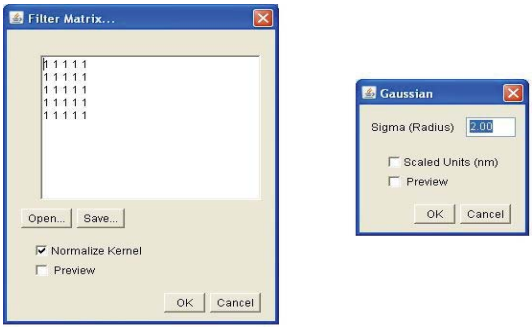


Figura 3. Parâmetros da ferramenta com respeito ao filtro matricial e o filtro gaussiano.

4.3 Processo de visualização das imagens

A projeção das imagens 2D é feita mediante a biblioteca Jrenderer3D [19], usando o algoritmo Z-buffer, chamado também buffer de profundidade, o qual faz uma renderização das imagens de acordo sua profundidade e sua luminância para cada pixel, em base a elo, calcula a projeção para o eixo z.

O algoritmo Z-buffer surgiu como ideia para resolver o problema de projeção de superfícies escondidas. O buffer salva todas as intensidades da imagem, para comparar com os valores dos pontos já existentes e assim calcular o valor z. Se o novo ponto esta atrás, este é descartado e se o novo ponto esta na frente, este substitui o anterior valor [9].

5. Extração das Medidas

Um importante ponto da ferramenta é a extração e análise das medidas na imagem. Para a extração das medidas usamos as figuras geométricas, e que foi feito selecionando diversos pontos em diferentes localizações sobre a imagem, produzindo automaticamente uma conexão entre os pontos. Foram utilizados principalmente linhas e polígonos para medir as estatísticas relevantes na imagem, em especial para realizar comparações de diversas áreas, mediante a seleção de polígonos, usando um marco de referência radial e sendo o centro radial o ponto mais alto da imagem do relevo. Esta seleção também se pode apreciar em sua projeção 3D. A informação obtida pela análise dos polígonos nos permite quantificar a distâncias entre eles, obter área, volume, momentos estatísticos (máximo, mínimo, desvio padrão, etc) e outros dados relevantes.

6. Resultados

A formação dos polímeros e sua auto-organização é muito conhecida pela simetria nos blocos dos copolímeros, e suas formações morfológicas das estruturas em lamelas, espirais, listras, gotas, etc dependendo dos parâmetros para a preparação das amostras [13] [14]. A aplicação destas estruturas de SEBS está sendo consideradas como potencial alternativa para a litografia macia na escala nanométrica, em especial na área de electroelectrônicos e outros ramos de tecnologia como a microfluídica e a biotecnologia [7] [11] [12].

Ultimamente as soluções obtidas pelo desmolhamento em filmes finos de SEBS, sob a influência da evaporação do solvente tolueno, apresenta uma auto-organização de padrões em forma de gotas e listras sub micrométrica [10]. O tolueno é um solvente apropriado para os polímeros, já que não influencia as interações moleculares iniciais, as ações da dinâmica molecular e a formação final da estrutura dependem apenas das interações específicas entre cada bloco do copolímero e o subtrato. Estudos de copolímeros com propriedades similares aos SEBS formam sobre o substrato estruturas cilíndricas orientadas perpendicularmente e paralelamente [16], a orientação dos cilindros de poliestireno depende da espessura do filme e taxa de evaporação do solvente, assim o equilíbrio termodinâmico o que envolve as interações entre o filme e o substrato, define a orientação. [17] [18].

Os resultados da eficácia da ferramenta desenvolvida, é feita pelas imagens das amostras das estruturas SEBS [10] da Figura 4, a ferramenta pretende contribuir ao estudo de padrões das nanoestruturas, centrando-se na análise das formações geométricas e morfológicas dos terraços. A ferramenta obtêm importantes dados quantitativos o que ajuda ao estudo da dinâmica da evolução das estruturas, substituindo outros métodos tradicionais providos por outros softwares comerciais e de código aberto. Além de poder processar uma grande quantidade de amostras nos dados da imagem, de maneira semi automática, o qual permite o estudo do comportamento estatístico dos dados coletados.

Na Figura 4 é observada a formação de terraços, mediante a ferramenta desenvolvida é selecionado estas áreas do interesse, com a finalidade de quantificar e extrair medidas. No lado da esquerda temos a fase e do lado direito temos ao relevo, a fase exibe quatro terraços os quais são marcados pelos polígonos, e ao mesmo tempo é marcado no relevo.

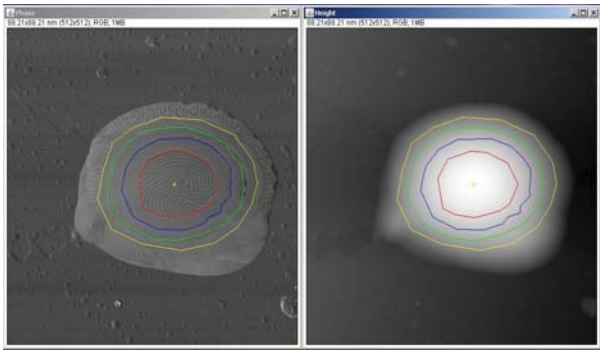


Figura 4. Imagem AFM das estruturas SEBS, fase e relevo dos 4 terraços selecionados respectivamente.

Na Figura 5 temos a visualização em 3D do relevo das estruturas SEBS, onde também é marcado a seleção feita na fase da imagem, apreciando-se a auto organização e auto estratificação dos terraços e uma proporção na altura entre cada terraço.

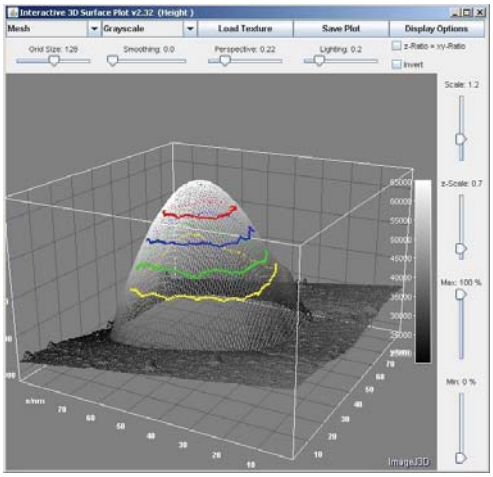


Figura 5. Visualização 3D de um imagem com seleção de polígonos.

A ferramenta desenvolvida faz uma análise com o conjunto de polígonos da imagem, na Figura 6 temos gráficos das alturas dos pontos dos polígonos, na esquerda o gráfico de linhas, e na direita o gráfico de barras das alturas respectivamente, onde os terraços reportam uma formação com respeito a altura, com uma diferença entre eles no intervalo de 20 a 35 nm.

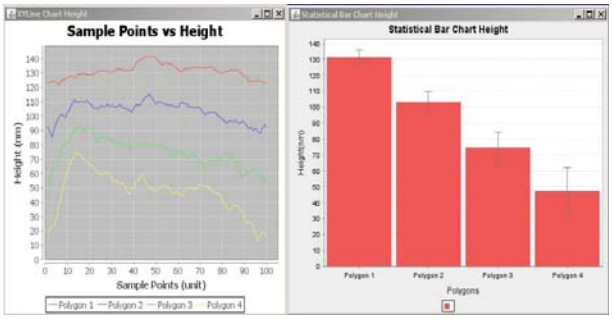


Figura 6. Gráfico de linhas e de barra da altura dos polígonos respectivamente.

Outros softwares extraem medidas utilizando apenas unicamente pontos específicos tal como o software comercial Nanoscope, ou outras ferramentas usam linhas tal como o software de código aberto, Gwyddion, tendo por consequência poucas amostras para a análise, desta maneira, a ferramenta desenvolvida permite ter estatísticas mais importantes para consolidar os resultados da análise.

Um análises feita para consolidar os resultados da pesquisa dos blocos SEBS [6][7][15], foi feita com dez imagens, este amostra a distancia media entre os terraços, em cada terraço obtemos uma amostra de cem pontos no polígono, a media desta diferença se mantém no intervalo de 20 a 35nm. Nestas imagens no centro tem linhas bem marcadas, na distancia media entre elas também esta dentro do intervalo de 20 a 35 nm, os resultado são exibidos na Tabela 2 e na Figura 7 temos os gráficos destes resultados. Os blocos SEBS tem uma formação estratificada e concêntrica de cilindros, os quais tem um diâmetro no intervalo de 20 a 35 nm, e um comprimento variado.

Nro.	Nro. Terraço	Nro. pontos polig.	Δμ altura polig.	Nro. Linhas	Δμ altura linhas
1	4	100	28.22	11	35.72
2	4	100	28.18	10	33.32
3	4	100	27.31	10	35.78
4	4	100	27.46	11	37.84
5	4	100	27.35	10	34.43
6	4	100	26.97	10	37.38
7	4	100	27.53	11	33.5
8	4	100	28.05	14	35.63
9	4	100	28.76	9	29.23
10	4	100	27.91	8	29.45

Tabela 2. Resultado da extração de medidas nas imagens

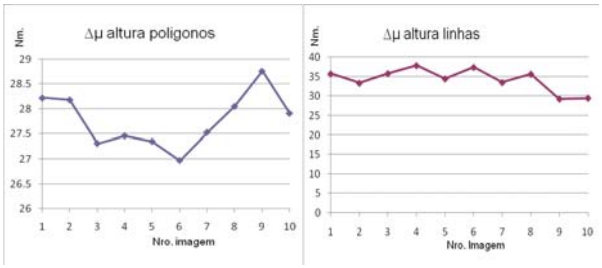


Figura 7. Gráfico de linhas de Δμ altura polígonos e de Δμ altura linhas respectivamente.

7. Conclusões

È muito conhecida na literatura dos polímeros, a presença da estrutura do cilindro em sua auto organização e auto estratificação, já seja alinhado em forma paralela ou perpendicular à superfície do solido [12] [16]. A ferramenta desenvolvida permite o processamento das imagens para obter uma grande quantidade dos dados, para a análise quantitativa, assim o estudo dos materiais seja mais preciso, já que identifica e analisa as estruturas das imagens. Às técnicas de seleção da ferramenta contribuem em grande medida, para especificar o dinamismo relacionado sua organização estratificada [10].

8. Referencias

[1] Binning, G; Quate, C.; Gerber, Ch., Atomic Force Microscope, Phys. Rev. Lett., v.56, p.930, 1986.

[2] Wiesendanger, R.; Guntherodth.-J., Scanning Tunneling Microscopy II, Berlin, Springer, 1995.

[3] Veeco, Scanning Probe Microscopy: Technology Overview, Digital Instruments, 2003.
http://nano.nd.edu/SC190/scanningprobe.pdf

[4] Veeco, MultiMode Scanning Probe Microscope Instruction Manual – Digital Instruments, 1996.
https://turing.phas.ubc.ca/mediawiki/upload/7/7f/MltiMode.pdf

[5] Veeco, NanoScope Command Reference Manual – Version 5.12, 2000.
http://www.ncf.uic.edu/manuals/MultimodeSPM_CRM_512C.pdf

[6] Salvador, MA; Bianchi, AGC; Pereira-da-Silva, MA. Dynamic formation of SEBS copolymer submicrometric structures. Polymer Vol. 51 Iss. 18 Pag. 4145-4151, 2010.

[7] M.A. Pereira-da-Silva, Análise de Superfície de Filmes Finos de Polianilina depositados em diferentes substratos por microscopia de força atômica. Teses para Doutor em Ciência e Engenharia de Materiais, IFSC 2001.

[8] Wilhelm Burger, Mark James Burge, Digital Image Processing – An Algorithmic Introduction using Java, Springer 2008, ISBN 1846283795.

[9] Edwin Catmull, Subdivision Algorithm for the Display of Curved Surfaces, PhD Thesis, University of Utah, 1974.
http://www.pixartouchbook.com/storage/catmull_thesis.pdf

[10] Michele A. Salvador; Andrea G. C. Bianchi; Marcelo A. Pereira-da-Silva; Antonio J.F. Carvalho; Roberto M. Faria. Dynamic formation of SEBS copolymer submicrometric structures. Polymer Vol. 51 Iss. 18 Pag. 4145-4151, 2010.

[11] Xu C, Fu XF, Fryd M, Xu S, Wayland BB, Winey KI, et al. Nano Lett 2006;6: 282-7.

[12] Karthaus O, Koito T, Shimomura M. Mater Sci Eng C 1999; 8-9:523-6.

[13] Liu Y, Li M, Bansil R, Steinhart M. Surface Effects on the Thin Film Morphology of Block Copolymers with Bulk Order−Order Transitions. Macromolecules 2007, Vol.40, pag. 9482-9490. DOI: 10.1021/ma1001194

[14] Sierra CA, Galán C, Fatou JG, Parellada MD, Barrio JA. Thermal and mechanical properties of poly-(styrene-b-ethylene-co-butylene-b-styrene) triblock copolymers. Polymer 1997, Vol.38, Num.17, pag. 4325-4335.

[15] Carvalho AJF, Pereira-da-Silva MA, Faria RM. Self-organization of triblock copolymer patterns obtained by drying and dewetting. Eur. Phys. J. E 20, 309{315 (2006).
DOI 10.1140/epje/i2006-10021-5.

[16] M. A. van Dijk, R. van den Berg. Ordering Phenomena in Thin Block Copolymer Films Studied Using Atomic Force Microscopy. Macromolecules 1995, 28 (20), pag 6773-678. DOI: 10.1021/ma00124a011.

[17] Kim G, Libera M.; Morphological Development in Solvent-Cast Polystyrene−Polybutadiene−Polystyrene (SBS) Triblock Copolymer Thin Films. Macromolecules, 1998 31 (8) pp. 2569-2577. DOI: 10.1021/ma971349i.

[18] M. Motomatsu, T. Takahashi, H.-Y. Nie, W. Mizutani and H. Tokumoto. Microstructure study of acrylic polymer-silica nanocomposite surface by scanning force microscopy. Polymer 38, 177-182 (1997).

[19] Kai Uwe Barthel, 3D-Data Representation with ImageJ. ImageJ User and Developer Conference, Mai 2006.

[20]Wayne Rasband, ImageJ, NIH Image software, 2008.
http://rsbweb.nih.gov/ij/

[21] Ben Fry and Casey Reas, Processing 2001.
http://www.processing.org/

Efficient Retina-like Resampling from Cartesian Images

Hugo Vieira Neto, Diogo Rosa Kuiaski and Gustavo Benvenutti Borba
Graduate School of Electrical Engineering and Applied Computer Science
Federal University of Technology - Paraná, Brazil
hvieir@utfpr.edu.br, diogo.kuiaski@gmail.com, gustavobborba@gmail.com

Abstract

This paper describes a novel method to resample cartesian images to a retina-like hexagonal tessellation, consisting of foveal and peripheral regions. We introduce the inflated-hexagons model for the foveal region resampling, which has an efficient data structure for storage and mapping of cortical coordinates. The integral image representation is used as intermediate step for the peripheral region log-polar resampling, which is less computationally demanding than the simulation of Gaussian receptive fields with varying widths. An important feature of the resulting model is that a gap-free transition is obtained between the near-uniformly sampled fovea and the space-variant periphery. Experiments with Principal Component Analysis (PCA) compression of the ORL Database of Faces show that the proposed retina-like resampling, in spite of massively reducing the amount of image data, still retains most of the information contents of conventional cartesian image sampling.

1. Introduction

The human retina can be subdivided topologically into foveal and peripheral regions. The fovea is the region at the centre of the retina, where the highest visual acuity is obtained. Surrounding the fovea there is a peripheral region whose visual acuity is increasingly lower with eccentricity [9]. In biological vision, retinal photoreceptors are arranged in a nearly hexagonal tessellation [3], while in computer vision image representation is often made using an uniformly sampled 2D cartesian coordinate system.

The use of the log-polar mapping for retina-like peripheral resampling is well-known for its advantages in achieving rotation- and scale-invariant representations [4, 8]. A review of the methods available in the literature can be found in [3]. This paper presents a new method to resample cartesian images into a full retina-like representation that differs from known models, such as the ones presented in [2] and [5], in the sense that it has an efficient data structure for storage and also that there is no sampling gap in the transition between fovea and periphery.

The proposed model is completely described by simple equations and does not resort to more complex methods, such as self-organisation [1], in order to generate the tessellation. The resulting data structure allows accurate and fast mapping between retinal and cortical coordinates. Cartesian image reconstructions from the retina-like resampling and quantitative assessment using incremental PCA [11] compression of the ORL Database of Faces [10] demonstrate the model's ability to preserve visual information contents.

2. The retinal model

To remap a conventional cartesian image that uses a uniform square tessellation to a log-polar image representation that uses a space-variant hexagonal tessellation, a data structure consisting of two regions – the fovea and the periphery – is needed. This happens because in the foveal region there is a large amount of evenly distributed photoreceptors, while in the peripheral region the amount of photoreceptors decreases with eccentricity. So, the model needs distinct data structures for the fovea, which is a small area with high photoreceptor density, and for the peripheral region, which is a larger area but with sparser photoreceptor density.

2.1. Foveal region

In the foveal region of our model, resampling is made with a series of concentric rings of photoreceptors, as shown in figure 1. The first internal ring has the shape of an hexagon and contains one photoreceptor in each of its vertexes. The second ring, which surrounds the first one, mostly maintains the hexagonal shape, but contains an additional photoreceptor between vertexes, along each side of the hexagon. As the rings grow farther from the centre of the fovea, the number of additional photoreceptors increases between the vertexes of the hexagon, while the hexagonal

rings themselves start to lose more and more their original shape and tend to become more and more circular in shape. We have called this "the inflated-hexagons model of the fovea".

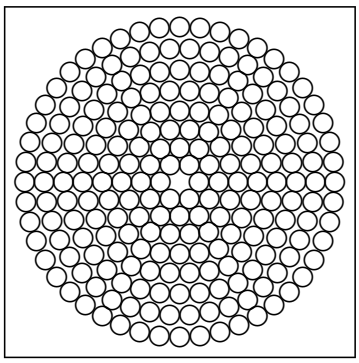


Figure 1. Spatial distribution of photoreceptors in the inflated-hexagons model of the fovea. Outer rings are more circular in shape, while inner rings are more hexagonal.

The high density of sampling in the foveal region leads to each photoreceptor having the same size of a cartesian pixel in every ring. The foveal region of the model is specified by the desired number of photoreceptor rings (M_f) given by the user. Ring radii increase in arithmetic progression and are given by $R_i = i + 1$, for $0 \leq i < M_f$. In each ring i there are $6 \times R_i$ photoreceptors.

The centre of each inflated-hexagons model photoreceptor is given in (x, y) cartesian coordinates, which are computed according to equation 1:

$$\begin{aligned} x &= x_a + (x_b - x_a)\,\tfrac{i}{M_f} \\ y &= y_a + (y_b - y_a)\,\tfrac{i}{M_f} \end{aligned}, \tag{1}$$

where (x_a, y_a) and (x_b, y_b) are given by

$$\begin{aligned} x_a &= x_c + R_i \left[\cos\theta_j + (\cos\theta_{j+1} - \cos\theta_j)\,\tfrac{k}{R_i}\right] \\ y_a &= y_c + R_i \left[\sin\theta_j + (\sin\theta_{j+1} - \sin\theta_j)\,\tfrac{k}{R_i}\right] \end{aligned} \tag{2}$$

and

$$\begin{aligned} x_b &= x_c + R_i \left\{\cos\left[\theta_j + (\theta_{j+1} - \theta_j)\,\tfrac{k}{R_i}\right]\right\} \\ y_b &= y_c + R_i \left\{\sin\left[\theta_j + (\theta_{j+1} - \theta_j)\,\tfrac{k}{R_i}\right]\right\} \end{aligned}, \tag{3}$$

for $0 \leq k < R_i$.

Equations 2 and 3 compute intermediate coordinate pairs for hexagonal (x_a, y_a) and circular (x_b, y_b) ring shapes, respectively, which are later composed into a single coordinate pair (x, y) by the weighted sums in equation 1. Equation 2 yields coordinates for a perfect hexagonal ring, while equation 3 yields coordinates for a perfect circular ring. When these coordinates are composed in equation 1, rings near to the centre of the fovea appear more like hexagons and rings far from the centre appear more like circles. In these equations, (x_c, y_c) represents the centre of the fovea, and θ_j and θ_{j+1} are angles between the x axis and lines connecting neighbour vertexes of the base hexagon to the centre, which are given by:

$$\theta_j = \frac{\pi}{3} j, \tag{4}$$

with $0 \leq j < 6$.

Cartesian visual data is remapped through bilinear interpolation and stored in memory as a 2D array – the cortical image – with the first column corresponding to the first photoreceptor ring, the next two columns corresponding to the second ring, the next three to the third and so forth, following an arithmetic progression, as indicated in figure 2. This topological arrangement also results in each row of the storage array representing wedges of $\pi/3$ rad of the foveal region.

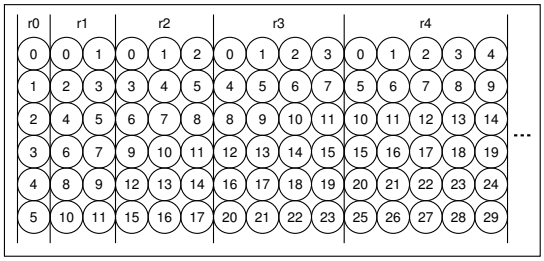


Figure 2. Corresponding cortical image for the inflated-hexagons model of the fovea. Column sections r_i represent photoreceptor rings increasingly far from the foveal centre and rows represent photoreceptor wedges of $\pi/3$ rad.

Coordinates in the foveal image are given in terms of ring and sector (r, s) and have corresponding coordinates (m, n) in the cortical image, where m is the cortical column coordinate and n is the cortical row coordinate. Retrieving visual data from the cortical array while searching the foveal image is straightforward. Equation 5 shows how foveal coordinates (i, s) can be computed from cortical coordinates (r, c):

$$\begin{matrix} i = \left\lceil \frac{\sqrt{8c+1}-1}{2} \right\rceil \\ s = i\left(r - \frac{1}{2}\right) - \frac{i^2}{2} + c \end{matrix}, \quad (5)$$

where $(\lceil . \rceil)$ denotes the ceiling function for rounding.

Equation 5 is derived from the arithmetic progression resulting from the cortical image arrangement, as shown in figure 2. The value of i is given by the largest value of c that satisfies the condition $i - \sum_{j=0}^{c} j \geq 0$. Therefore, solving Gauss's equation for the sum of the elements of an arithmetic progression yields the relationship between i and c.

Equation 6 shows how to compute the inverse mapping, obtaining (r, c) as a function of (i, s):

$$\begin{matrix} c = \frac{i^2-i}{2} + \mathrm{mod}\{s-1, i\} \\ r = \mathrm{mod}\{3(i^2-i) + \left\lfloor \frac{n-1}{m} \right\rfloor, 6\} \end{matrix}, \quad (6)$$

where $\mathrm{mod}\{a, b\}$ is the remainder of the integer division a/b and $\lfloor . \rfloor$ denotes the floor function for rounding.

2.2. Peripheral region

In the periphery, where a log-polar model is used, photoreceptor rings have a perfect circular shape and a constant number of photoreceptors per ring, but photoreceptor size starts to increase as a function of the distance to the foveal centre. The total size of the peripheral region is provided by the user in terms of the desired number of rings (M_p).

In order to keep coherence at the border between the fovea and the periphery, the number of sectors in the peripheral region is the same as in the last ring of the foveal region. Therefore, the number of sectors in the periphery is given by $N = 6 \times M_f$ and the half angular resolution can be computed as:

$$\theta = \frac{\pi}{N}. \quad (7)$$

The photoreceptor radius r_i is related to the ring radius R_i according to:

$$r_i = R_i \sin\theta. \quad (8)$$

As the minimum desired size for a photoreceptor is equivalent to one cartesian pixel ($r_0 = 1/2$), the innermost peripheral ring radius R_0 is given by:

$$R_0 = \frac{1}{2\sin\theta} \quad (9)$$

Ring radii increase in a geometric progression given by:

$$R_i = bR_{i-1}, \quad (10)$$

where b is computed as follows:

$$b = \frac{\sin\theta(\sin\theta + \sqrt{2\cos\theta + 1}) + \cos\theta}{(\cos\theta)^2}. \quad (11)$$

Finally, the cartesian coordinates of each photoreceptor centre is given by:

$$\begin{matrix} x_{i,j} = x_c + R_i \cos[\theta(i+2j)] \\ y_{i,j} = y_c + R_i \sin[\theta(i+2j)] \end{matrix}, \quad (12)$$

for $0 \leq i < M_p$ and $0 \leq j < N$.

Figure 3 shows the resulting space-variant tessellation with $M = 18$ rings and $N = 48$ sectors. Notice that the final organisation consists of spiral sectors and rings.

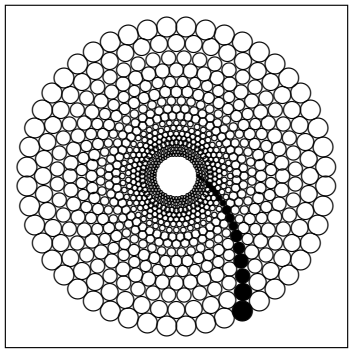


Figure 3. Peripheral log-polar hexagonal tessellation with $M = 18$ rings and $N = 48$ sectors. The first spiral sector is indicated by black circles.

One of the most important features of the proposed retina-like model is that there is no sampling gap between fovea and periphery, unlike other models available in the literature [5]. Notice that the centre of the image is lost if a purely log-polar structure is used, as shown in figure 3.

2.3. Efficient implementation

Photoreceptor values in the peripheral region can be approximated by the average of cartesian pixel values within the square-shaped areas shown in figure 4.

An efficient implementation to compute the peripheral log-polar mapping can be made using the integral image representation introduced in [12]. The integral image at location (x, y) contains the sum of the pixel values above and to the left of (x, y), as follows:

$$ii(x, y) = \sum_{m \leq x, n \leq y} i(m, n), \quad (13)$$

where $ii(x, y)$ is a pixel of the integral image and $i(m, n)$ is a pixel of the original image.

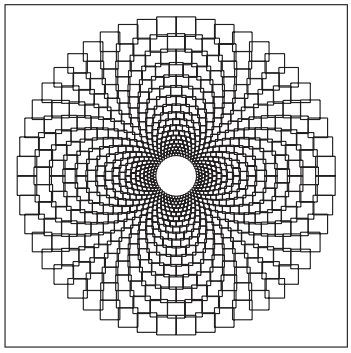


Figure 4. Square photoreceptor approximation for the peripheral log-polar tessellation in figure 3.

The integral image can be computed in a single pass over the original image with the following recurrences [12]:

$$\begin{aligned} s(x,y) &= s(x,y-1) + i(x,y) \\ ii(x,y) &= ii(x-1,y) + s(x,y) \end{aligned}, \tag{14}$$

where $s(x,-1) = 0$ and $ii(-1,y) = 0$.

The most relevant property of the integral image representation is that the sum of pixel values within a rectangle of any size in the original image can be computed in constant time using the values of the corresponding four corners in the integral image, as shown in figure 5.

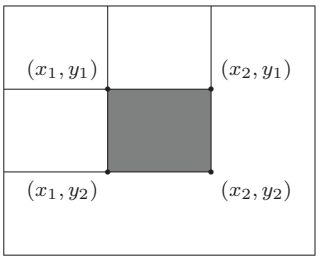


Figure 5. Integral image: the sum of the pixel values within the grey rectangle can be computed as $ii(x_1,y_1) - ii(x_2,y_1) - ii(x_1,y_2) + ii(x_2,y_2)$.

The sum of pixel values in the square areas shown in figure 4 can be very efficiently computed with the use of look-up tables containing their corner coordinates:

$$\begin{aligned} x_1 &= x_{i,j} - r_i \\ y_1 &= y_{i,j} - r_i \\ x_2 &= x_{i,j} + r_i \\ y_2 &= y_{i,j} + r_i \end{aligned} \tag{15}$$

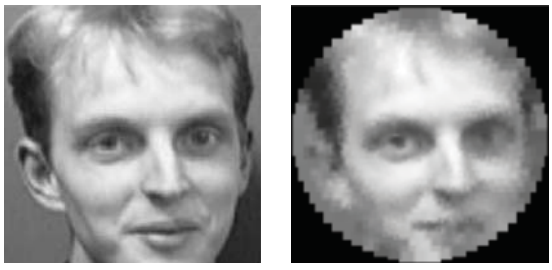
Therefore, average photoreceptor values are obtained by:

$$p = \frac{ii(x_1,y_1) - ii(x_2,y_1) - ii(x_1,y_2) + ii(x_2,y_2)}{(x_2 - x_1)(y_2 - y_1)} \tag{16}$$

3. Experiments

In our initial experiments, we have resampled cartesian images of 100×100 pixels from the ORL Database of Faces [10] with our retina-like model. Ten photoreceptor rings in the fovea and 16 in the periphery were used in the structure of the model, resulting in a total of 1290 photoreceptors. When compared to the original cartesian representation, this number of photoreceptors represents a reduction in storage space in excess of 87%.

Figure 6 shows the example of an original image and its respective retina-like resampled output reconstruction for an image from the ORL Database of Faces. The data reduction seems extremely lossy when a comparison is to be made between the original and reconstructed images. However, the reconstructed image obtained is still reasonably intelligible, demonstrating that the "essential information" of the original image was preserved. The main advantage of this fact is that image processing methods can be executed faster in the cortical map domain because the amount of data to be processed is reduced.



(a) (b)

Figure 6. ORL Database of Faces example: (a) original image; (b) retina-like resampled image reconstruction.

It is important to keep in mind that such retina-like representation is useful in applications that involve robotic vision, in which cameras are able to move with respect to the environment. In these applications, it is necessary to centre the high-resolution fovea of the model at the current object of interest that needs to be analysed. The low-resolution periphery in this kind of application is used mostly to shift the focus of attention of the system [7].

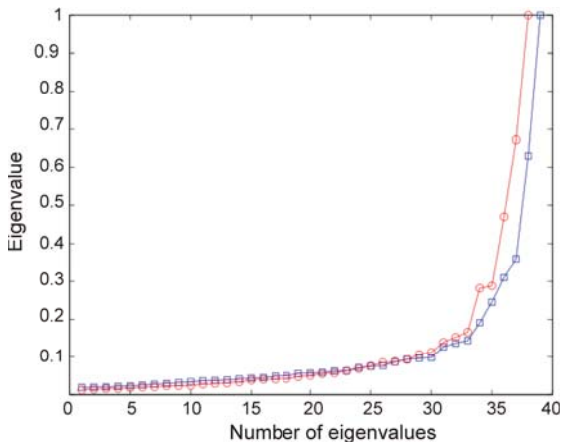


Figure 7. PCA of the ORL Database of Faces: (a) eigenvalue plot; (b) cumulative eigenvalue plot. The blue plots correspond to the cartesian image domain and the red plots correspond to the retinal image domain.

In order to verify if the main information contents was preserved after the retina-like resampling, we performed incremental PCA [11] on the complete ORL Database of Faces. PCA – also known as the Hotteling Transform or discrete Karhunen-Loeve Transform in the context of image processing – provides optimal image reconstruction in the least-squares sense [6]. Therefore, we refer here to "amount of information" or "main information" as the eigenvalues obtained through PCA.

In figure 7a, it is possible to visualise that the amount of information for the retina-like representation is similar to that of the cartesian representation, as the number of eigenvalues obtained is almost the same. Figure 7b shows a graph for the cumulative eigenvalue plot, where one can observe that information contents grows even faster for the retina-like mapping.

An additional experiment was conducted in order to demonstrate some of the topological properties of the proposed retina-like resampling. Figure 8 shows the cortical data structure and the reconstructed image in contrast with the original cartesian image. In this example, it is possible to notice that circular regions are remapped to horizontal lines in the cortical periphery and that corners with at least $\pi/3$ rad are also remapped to horizontal lines in the cortical fovea of the data structure.

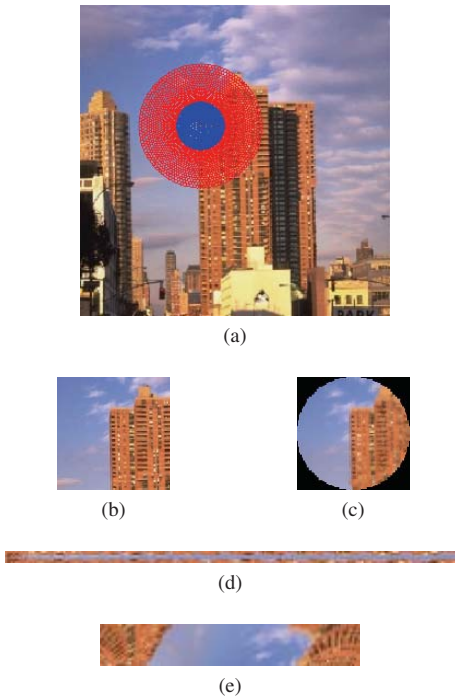


Figure 8. Topological properties: (a) retinal model superimposed to an image; (b) original image patch; (c) retina-like resampled image patch reconstruction; (d) cortical fovea; (e) cortical periphery.

4. Conclusions

We have presented a model for retina-like resampling of cartesian images composed of a high-resolution fovea and a space-variant log-polar periphery. Important features of the model are the gap-free transition between fovea and periphery, and the use of simple mathematical equations to define the mapping between cartesian and retina-like domains. The model provides a coherent cortical map representation for both foveal and peripheral region, which can use an efficient implementation based on the integral image representation.

Experiments using incremental PCA demonstrate that our model, despite massively reducing the amount of image data, still preserves most of the information contents of the ORL Database of Faces. This is also observed through visual comparison of original cartesian images and the images resampled using the model.

Future work includes investigation of specific characteristics of the model, in special its topological properties and behaviour in the frequency domain. It should be noted that our model has the advantage of presenting periodicity in one of the dimensions of both cortical images – foveal and peripheral – which presents potential advantages for computing its Fourier Transform.

We also intend to investigate the well-know properties of log-polar representations to achieve rotation- and scale-invariance for feature extraction in content-based image retrieval and robotic vision applications. The model seems to be particularly interesting to be used in conjunction with a visual attention mechanism [7], where the salient visual locations should be centred on the fovea.

References

[1] L. S. Balasuriya and J. P. Siebert. An artificial retina with a self-organised retinal receptive field tessellation. In *Proceedings of the AISB'03 Symposium on Biologically-inspired Machine Vision, Theory and Application*, pages 34–42, Aberystwyth, UK, 2003.

[2] A. Bernardino, J. Santos-Victor, and G. Sandini. Model-based attention fixation using log-polar images. In V. Cantoni, A. Petrosino, and M. Marinaro, editors, *Visual Attention Mechanisms*. Plenum Press, New York, NY, 2002.

[3] M. Bolduc and M. D. Levine. A review of biologically motivated space-variant data reduction models for robotic vision. *Computer Vision and Image Understanding*, 69(2):170–184, 1998.

[4] P. Cavanagh. Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance. In D. Rose and V. G. Dobson, editors, *Models of the Visual Cortex*, pages 85–95. Wiley, New York, NY, 1985.

[5] H. M. Gomes and R. B. Fisher. Learning-based versus model-based log-polar feature extraction operators: A comparative study. In *Proceedings of the Sixteenth Brazilian Symposium on Computer Graphics and Image Processing*, pages 299–306, São Carlos, Brazil, 2003.

[6] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Pearson, 3rd edition, 2008.

[7] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, 2000.

[8] T. Kurita, K. Hotta, and T. Mishima. Scale and rotation invariant recognition method using higher-order local autocorrelation features of log-polar images. In *Proceedings of the 3rd Asian Conference on Computer Vision*, volume 2, pages 89–96, Hong Kong, China, 1998.

[9] S. E. Palmer. *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA, 1999.

[10] F. Saramaia and A. Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pages 138–142, Sarasota, FL, 1994.

[11] H. Vieira Neto and U. Nehmzow. Incremental PCA: An alternative approach for novelty detection. In *Proceedings of TAROS 2005: Towards Autonomous Robotic Systems*, pages 227–233, London, UK, 2005.

[12] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, Kauai, Hawaii, 2001.

Reconhecimento Multibiométrico de Pessoas em Imagens de Vídeo para Monitoramento Visual de Ambientes

Gustavo Botelho de Souza, Ícaro Ribeiro, Aparecido Nilceu Marana
Universidade Estadual Paulista (UNESP), Faculdade de Ciências - Bauru
Departamento de Computação
gustavo.botelho@gmail.com, icaroribeiro@fc.unesp.br, nilceu@fc.unesp.br

Abstract

Biometrics has been widely employed for fast, automatic, and robust people identification. Recently, in order to increase the population coverage and improve the identification rates of automatic identification systems, the multibiometric identification has been proposed. Good results have been reported in the literature regarding multibiometric systems based on the fusion of biometric features like fingerprint, palm vein, iris, etc. However, these biometric features require user collaboration and do not allow identification at a distance, which is a necessity in applications like surveillance. For this purpose, "soft" biometric traits, such as height, weight, gender, hair and skin color, among others, seems to be more suitable since they can be extracted at a distance with a acceptable accuracy, in a covert and non-intrusive way. In this work we present a prototype of a multibiometric system designed for people identification in a surveillance application. This system is based on two "soft" biometric traits, height and skin color, and one traditional biometric trait, face. The proposed multibiometric system was implemented and assessed in an indoor and controlled environment. Results obtained during the experiments indicate that the fusion of traditional "strong" features, like face, with new "soft" features make possible reach a high accuracy in the process of people identification and can be proper for covert surveillance applications.

1. Introdução

A biometria ("bio": vida, "metria": medida) pode ser definida como um ramo da biologia que emprega a estatística e a análise quantitativa no estudo de seus "dados" [1]. Nos últimos anos ela vem sendo associada ao reconhecimento automático de pessoas (por sistemas computacionais) baseado em suas características físicas, fisiológicas e comportamentais.

O homem sempre teve necessidade de identificar-se para obter acesso a determinado recurso, lugar ou privilégio. Ao longo da história, equipamentos, tais como cadeados e cofres foram desenvolvidos e aprimorados para impedir o acesso ou uso de recursos por pessoas não autorizadas. Com o advento dos computadores, as senhas e cartões magnéticos tornaram-se mecanismos de segurança comuns no cotidiano das pessoas.

Os cadeados e cofres, justamente por serem dispositivos mecânicos, podem ser "facilmente" burlados por usuários mal intencionados.

Por sua vez, as senhas e cartões também apresentam falhas. Os *passwords* podem ser esquecidos, inferidos ou compartilhados indevidamente. Os cartões por sua vez podem ser roubados, perdidos ou emprestados para uso indevido.

Nos últimos anos, com o aumento do poder computacional e o barateamento de sensores, técnicas mais eficientes começaram a ser empregadas no reconhecimento de pessoas para diversos fins, entre eles a segurança [2]. Estas técnicas valem-se da biometria e da alta capacidade de processamento dos computadores atuais para, por exemplo, permitir acesso de um indivíduo a lugares restritos, a sistemas computacionais, a informações sigilosas, ou para outros fins como a identificação criminal e controle de ponto.

Baseado nas características físicas, fisiológicas e comportamentais dos indivíduos, os sistemas são capazes de identificá-los. Com isto, é possível reconhecer o indivíduo "pelo que ele é" e não "pelo que ele possui" (no caso dos cartões) ou "pelo que sabe" (no caso das senhas) [3].

Na realidade, os sistemas biométricos podem se valer de uma ou mais evidências biométricas (neste caso são chamados de multibiométricos). Em geral, os sistemas multibiométricos apresentam melhores desempenhos que os sistemas que se baseiam em apenas uma evidência biométrica. Além disso, propiciam uma maior cobertura da população e dificultam os ataques fraudulentos [4].

O processo de reconhecimento biométrico que se vale apenas de características como a impressão digital, a íris e as veias das mãos, por exemplo, não permite a identificação à distância e, portanto, não pode ser usado para o monitoramento discreto de ambientes. Além disso, nesse tipo de aplicação, o sistema biométrico precisa operar no modo de identificação. Isso exige uma busca (dadas as características biométricas extraídas da pessoa sob análise no momento) por todos os registros da base de

dados, o que pode causar uma demora demasiada dependendo do número de usuários cadastrados no sistema.

Para a identificação discreta de pessoas em aplicações de vigilância, algumas características físicas tais como sexo, peso, altura, raça, cor do cabelo, cor da pele, cor dos olhos, dentre outras, conhecidas como *"soft" biometric traits* [4, 5, 6], são mais apropriadas e podem tornar o sistema de reconhecimento biométrico mais eficiente e preciso, quando utilizadas em conjunto com características biométricas tradicionais, como as faces.

Uma das grandes vantagens do uso das características *"soft"* em sistemas multibiométricos é que em geral elas podem ser extraídas de forma não intrusiva, silenciosa e à distância, com razoável precisão.

Dentre as aplicações que exigem identificação de pessoas que mais podem se beneficiar pelo reconhecimento biométrico por meio das características *"soft"* encontra-se o monitoramento sigiloso de ambientes.

O monitoramento de ambientes pode ser realizado para diversos fins como, por exemplo, para a análise da movimentação de pessoas em certo local a fim de adequá-lo e facilitar a locomoção dos pedestres. Entretanto, em geral, ele se volta à segurança do local sendo monitorado.

O monitoramento destinado à segurança é uma tendência, principalmente nas grandes metrópoles. Na Inglaterra, estima-se que o número de câmeras CCTV (*Closed-Circuit Television*) espalhadas por todo o país tenha ultrapassado um milhão. Segundo Xiang [7], um cidadão inglês, em média, aparece nas filmagens destas câmeras cerca de 300 vezes ao dia. Em geral as câmeras gravam imagens de locais onde é necessário garantir-se segurança como bancos, estacionamentos, etc. a fim de identificar criminosos ou comportamentos suspeitos.

Este artigo tem como objetivo apresentar o protótipo de um sistema de visão computacional projetado para ser utilizado em aplicações de monitoramento discreto de pessoas, utilizando a face como característica biométrica principal, associada a duas características biométricas *"soft"*: a altura e a cor da pele das pessoas. Esse sistema biométrico foi projetado para operar no modo de identificação de pessoas.

2. Material e Metodologia

2.1. Material

Para desenvolvimento e avaliação do protótipo do sistema de visão computacional biométrico projetado foram utilizados os seguintes recursos:

- Vídeos capturados por uma *webcam* com resolução de 1.3 megapixels (a uma taxa de 30 quadros por segundo), posicionada sobre um suporte a uma altura de aproximadamente 1.70 metros, em um ambiente interno, com iluminação artificial constante, homogênea e não muito intensa;
- Computador com sistema operacional Windows 7, linguagem C++ com ambiente de desenvolvimento Visual Studio 2008 (Visual C++);
- Biblioteca OpenCV (versão 1.0) [8].

2.2. Metodologia

Para atingir os objetivos desse trabalho, inicialmente foram definidas que duas características biométricas *"soft",* altura e cor da pele das pessoas, seriam extraídas das imagens de vídeo capturadas pela *webcam* para serem utilizadas pelo sistema proposto em conjunto com a análise facial.

Para o cálculo da altura foi utilizado o algoritmo proposto por Jain et al. [6] com algumas adaptações que, em síntese, consiste em armazenar uma imagem com o *background*, ou seja, o local onde foram capturados os vídeos, sem a presença de pessoas no *foreground*, e realizar comparação quadro a quadro do vídeo detectando assim a posição da pessoa se movendo no local. Esta comparação só se inicia quando um classificador baseado nas características de Haar, similar ao proposto por [9], detectar que há uma pessoa no quadro atual do vídeo. Com isto, ao perceber que a pessoa passa por uma posição pré-determinada no ambiente (para a qual se sabe a altura) o sistema consegue estimar a altura da pessoa utilizando-se regra de proporção.

Com relação à análise da cor da pele, foi usada uma técnica simples e de baixo esforço computacional proposta por diversos autores, dentre eles, Kovac, Peer, e Solina [10]. Esta técnica consiste basicamente em selecionar os pixels da imagem (no formato RGB) cujos componentes R, G e B atendam às seguintes restrições:

1ª: $R > 95$;
2ª: $G > 40$;
3ª: $B > 20$;
4ª: $\max(R,G,B) - \min(R,G,B) > 15$;
5ª: $|R - G| > 15$;
6ª: $R > G$;
7ª: $R > B$.

Essas restrições definem o subespaço de cores (dentro do espaço RGB) onde se encontram as cores de pele humana conhecidas. Com isto, uma cor média para a pele da pessoa presente na imagem pode ser calculada e usada no processo de reconhecimento, tendo como base os pixels selecionados.

Na realidade, estas restrições servem apenas para detectar pele em situações onde a iluminação é uniforme, constante e não muito intensa. Conforme mencionado por

pesquisadores que propuseram ou utilizaram esta técnica, as regras acima foram definidas e são justificadas da seguinte forma: as três primeiras são responsáveis por modelar a tonalidade da cor de pele, a quarta evita que os valores de R, G e B do pixel estejam muitos próximos diminuindo assim a influência de tons de cinza, a quinta restrição é responsável por garantir a atenção à real complexidade dos tons de pele, mantendo os valores de R e G distantes e, por fim, as duas últimas regras asseguram que para o pixel ser um pixel de pele humana seu componente R deve ser maior que os outros (G e B), uma vez que, visivelmente, as cores de pele são mais próximas do vermelho (marrom, etc.) que do azul ou do verde.

No sistema desenvolvido, para evitar que se considerem pixels de objetos com cores próximas às cores de pele no cálculo da cor média da pele da pessoa presente no vídeo, restringiu-se a possibilidade de classificação como pixel de pele aos pixels da região da face da pessoa no quadro do vídeo sendo tratado.

Para se realizar o reconhecimento facial foi empregado um algoritmo que opera em três fases: primeiro realiza-se a detecção facial no quadro do vídeo baseado nas características de Haar (algoritmo similar ao proposto em [9] e implementado na biblioteca OpenCV), depois a face detectada é recortada do quadro do vídeo em um retângulo de tamanho padrão e transformada para tons de cinza e, por fim, a face é projetada no subespaço das autofaces (técnica PCA - *Principal Component Analysis*) de forma similar ao apresentado em [11]. Vale observar que o subespaço citado é gerado a partir das faces já cadastradas na base de dados do sistema desenvolvido.

Para avaliar o desempenho do sistema multibiométrico de identificação à distância proposto que se vale das três análises explanadas (face, altura e cor da pele) foram realizados experimentos (gravados vídeos) em um ambiente interno, com iluminação controlada e no qual apenas uma pessoa por vez adentra a cena.

A Figura 1 mostra um esboço do ambiente onde o sistema de monitoramento foi avaliado. Nesse ambiente, a pessoa caminha por um corredor (posição 1), cruza uma porta (posição 2), entra em uma sala (local onde há uma câmera filmando o ambiente), caminha para frente (posição 3) e finalmente desvia para a direita, conforme indicado pelas setas na própria Figura 1. Ao passar pela porta (posição 2) sua altura é estimada com base na altura do batente. Ao chegar na posição 3, o sistema detecta e analisa a face da pessoa, bem como extrai a cor média da pele da sua face (conforme os algoritmos descritos anteriormente).

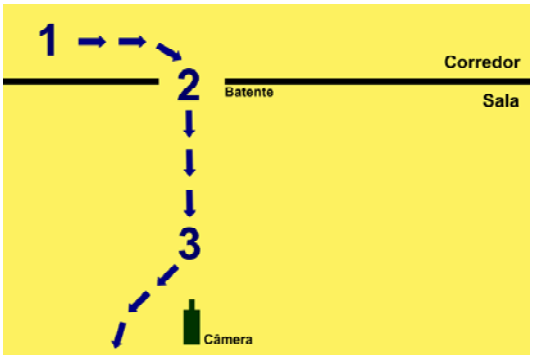


Figura 1. Esboço do ambiente utilizado para avaliar o sistema multibiométrico de identificação à distância.

A partir do momento em que a face da pessoa começa a ser detectada (próximo à posição 3) repetem-se os seguintes passos até o momento em que não é não é mais possível detectar faces nos quadros do vídeo:

- O sistema calcula (para cada registro da base de dados) a soma ponderada das distâncias normalizadas obtidas com cada característica biométrica individualmente (altura, cor da pele e face), sendo que cada distância individual indica o grau de similaridade entre a característica extraída da pessoa detectada no vídeo e a característica do registro armazenado na base de dados. Como observação, a distância facial entre uma face sob análise e uma face de um registro do banco de dados é obtida pela soma das diferenças ao quadrado dos coeficientes relacionados aos autovetores do subespaço gerado pela PCA da face sob análise e da face do registro, a distância da altura é a distância euclidiana entre a altura estimada no momento e a do registro na base de dados e para a cor da pele a distância gerada é a soma das distâncias euclidianas entre os valores dos componentes R, G e B da cor da pele extraída e estes mesmos componentes da cor da pele do registro do banco de dados. A soma ponderada mencionada gera a distância final de cada registro da base de dados e é calculada atribuindo-se peso 0.8 à distância facial (pois esta é a característica biométrica utilizada no sistema que apresenta maior unicidade), peso 0.15 para a distância da cor da pele e peso 0.05 para a distância da altura. Nos casos em que o sistema não consegue detectar a pessoa passando pela porta (o que não raramente ocorre visto que o padrão corporal das pessoas é muito variável dificultando a detecção da pessoa no vídeo pelo classificador usado), a característica altura é desprezada no cálculo das distâncias finais e os pesos para as características

face e cor da pele são alterados para 0.9 e 0.1, respectivamente;

- Caso a menor distância final calculada esteja abaixo de um limiar definido com base na distribuição de casamento genuínos e impostores entre as faces já cadastradas na base de dados ("casadas" duas a duas), a pessoa é reconhecida como sendo aquela relacionada com o *template* que proporcionou a menor distância. Caso contrário, a pessoa é rejeitada. Se a decisão de rejeição se mantiver para os quadros subsequentes a pessoa é considerada desconhecida pelo sistema e sua imagem é armazenada no banco de dados para verificação posterior.

Quando a face da pessoa deixa de ser detectada nos quadros do vídeo, a identidade que mais vezes foi encontrada nos passos anteriores é associada à pessoa. Nesse momento, o sistema registra no banco de dados que aquele usuário esteve presente naquele local e em determinado dia e horário.

3. Sistema de Monitoramento Desenvolvido

No protótipo do sistema biométrico desenvolvido, apenas os administradores podem cadastrar novos usuários ou iniciar o módulo de reconhecimento de pessoas. Ao acessar o sistema, justamente por motivos de segurança, registra-se também na base de dados o horário e a data da entrada e saída de cada administrador. Na Figura 2 apresenta-se a interface de acesso ao sistema de monitoramento desenvolvido.

Figura 2. Interface de acesso ao sistema.

O sistema possui um módulo de cadastramento de pessoas. Neste módulo é possível cadastrar um indivíduo com seus dados pessoais bem como suas características biométricas (altura, face e cor de pele). O sistema detecta nas imagens de vídeo capturadas pela *webcam* a face da pessoa sendo cadastrada em tempo real e segmenta as regiões de pele da face, encontrando a cor média da pele da pessoa. No momento do cadastro, o administrador informa altura da pessoa no campo apropriado.

O sistema permite o cadastro de mais de uma face com sua respectiva cor de pele para cada pessoa. A Figura 3 mostra a interface do módulo de cadastro de pessoas.

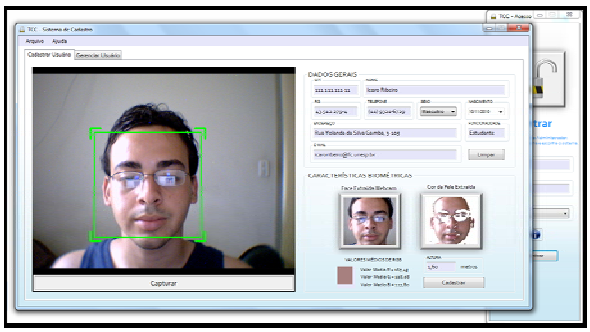


Figura 3. Módulo de cadastro do sistema desenvolvido.

A Figura 4 mostra o gerenciamento de usuários, também feito no módulo de cadastro do sistema onde é possível alterar dados pessoais de um usuário, consultar por um usuário, excluir certo indivíduo do sistema bem como cadastrar novas informações biométricas para o mesmo ou excluir alguma informação indesejada (uma imagem facial com pouca qualidade, por exemplo).

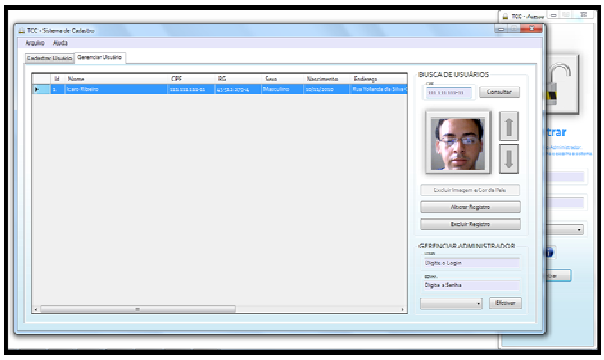


Figura 4. Módulo de cadastro (aba "Gerenciar Usuário") do sistema desenvolvido.

O módulo de reconhecimento do sistema por sua vez, é responsável por realizar a função principal do sistema: a identificação da pessoa presente no vídeo. Neste módulo são exibidos os nomes das pessoas associadas aos *templates* mais "votados" durante o intervalo de tempo decorrido desde o momento que a pessoa surge no vídeo até o momento em que ela desaparece da cena. Ao término deste intervalo, uma janela exibe a identificação final da pessoa. Essa identificação é armazenada no banco de dados, juntamente com a data e horário da ocorrência

desse evento. Nas Figuras 5, 6, 7 e 8 apresenta-se o sistema em funcionamento, estimando a altura e a cor média da pele de uma pessoa, executando a detecção da face da mesma e seu reconhecimento.

A Figura 5 apresenta o módulo de reconhecimento (aba "Vídeo da Altura"). Nesta aba são exibidos os resultados do algoritmo de detecção da pessoa e estimativa da altura da mesma. O segmento de reta na cor azul indica o rastreamento da pessoa detectada quadro a quadro no vídeo, enquanto o segmento de reta na cor verde permanece sobre a lateral da porta (utilizada nesse ambiente como referência para o cálculo da altura das pessoas). Após passar pela porta a altura da pessoa é calculada e exibida no campo "Altura Extraída do Vídeo". No exemplo apresentado na Figura 5, a altura estimada para a pessoa no vídeo foi de 1.75 metros, valor bem próximo da altura real da mesma (1.72 metros).

Figura 5. Aba "Vídeo da Altura.

Na Figura 6 mostra-se o resultado obtido com a técnica de segmentação da cor da pele. Nesse caso, a pessoa já está mais próxima da câmera. Na aba "Vídeo da Pele" é exibida a imagem obtida do quadro atual do vídeo com as regiões de pele segmentadas. Nota-se que várias partes da imagem foram consideradas erroneamente como sendo regiões de pele. Porém, o algoritmo de reconhecimento do sistema usa apenas as regiões da face da pessoa para estimar a cor média da sua pele.

A Figura 7 por sua vez mostra o sistema operando sobre um quadro posterior do vídeo onde a pessoa está ainda mais próxima da câmera. Na aba "Vídeo da Face" é exibida a imagem obtida do quadro atual do vídeo com a face detectada (retângulo verde).

Nota-se que o algoritmo de detecção facial foi muito preciso e pode-se perceber que o algoritmo de reconhecimento (que opera combinando informações da face, pele e altura extraídas da pessoa) acertou, pois a pessoa da imagem no campo "Face Extraída" é a mesma da imagem do campo "Face da Base".

Na tabela do canto inferior direito da tela apresentada na Figura 7 são registradas as identificações parciais realizadas pelo sistema no decorrer do vídeo.

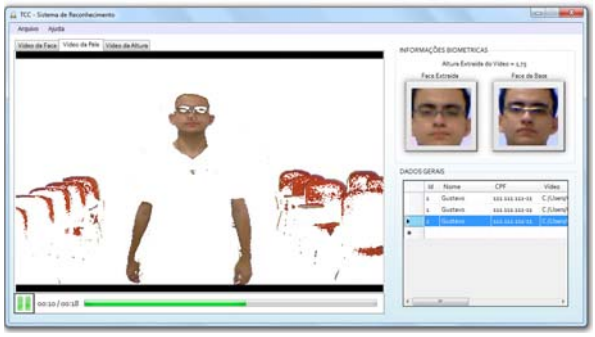


Figura 6. Módulo de reconhecimento (aba "Vídeo da Pele").

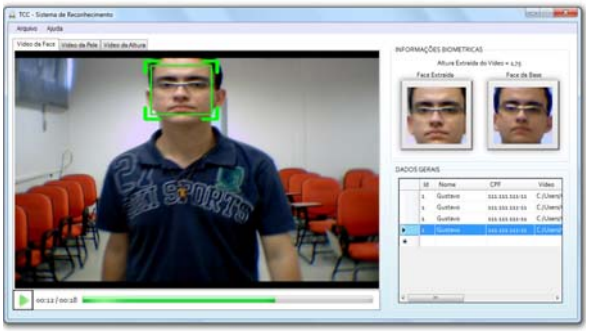


Figura 7. Módulo de reconhecimento (aba "Vídeo da Face").

Na Figura 8 mostra-se a mensagem informando a identidade final da pessoa que apareceu no vídeo, tendo em vista os reconhecimentos intermediários realizados durante o decorrer do mesmo. No exemplo mostrado, o sistema determinou corretamente a identidade da pessoa. Essa identidade final é registrada no banco de dados.

Figura 8. Após a pessoa sair da cena, o sistema exibe uma mensagem informando a identidade "final" da mesma.

4. Resultados Experimentais

Para avaliar o desempenho do protótipo do sistema multibiométrico proposto para monitoramento à distância de ambientes, como descrito na subseção 2.2, foram realizados experimentos utilizando-se vídeos previamente gravados de 7 pessoas distintas caminhando naturalmente pela área sob monitoração, conforme mostrado na Figura 1. Nesses experimentos obteve-se 100% de acerto na identificação das pessoas.

5. Discussão e Conclusão

Nesse artigo apresentamos o protótipo implementado de um sistema de visão computacional projetado para auxiliar no monitoramento de ambientes, utilizando a fusão de características biométricas tradicionais, como a face, e de características *"soft"*, como a altura e a cor da pele da pessoa.

Uma das características mais importantes do sistema proposto é a possibilidade de identificar pessoas de forma discreta e à distância, sem exigir muita colaboração durante o processo de monitoramento do ambiente.

As características biométricas *"soft"*, apesar de apresentarem baixo grau de unicidade, são importantes para aumentar as taxas de reconhecimento correto das pessoas, quando fundidas com outras características biométricas. Além disso, elas podem ser úteis quando utilizadas como índices em grandes bancos de dados, para otimizar as consultas, uma vez que apenas os *templates* associados a pessoas que possuem as características biométricas *"soft"* parecidas com as da pessoa sob análise teriam que ser analisados.

Embora os resultados obtidos com os experimentos realizados durante esse trabalho não sejam conclusivos, eles mostraram a viabilidade da proposta, ou seja, permitiram concluir que é possível utilizar as características biométricas *"soft"* altura e cor da pele da pessoa em conjunto com o reconhecimento facial para o monitoramento de ambientes à distância, de forma discreta e sem muita colaboração das pessoas sob análise.

6. Trabalhos Futuros

Pretende-se, como extensão desse trabalho, realizar novos experimentos utilizando uma base de dados maior e mais heterogênea bem como fazer experimentos de monitoramento de pessoas em outros ambientes, com diferentes condições de iluminação. Pretende-se também desenvolver e incorporar ao sistema técnicas para a identificação de várias pessoas presentes simultaneamente nos quadros do vídeo.

7. Referências

[1] Webster's New World Dictionary of the American Language, College Edition, World Publishing Co., New York, 1966.

[2] A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, and J. L. Wayman, "Biometrics: A Grand Challenge", Proc. International Conference on Pattern Recognition (ICPR), vol. II, pp. 935-942, Cambridge, UK, August 2004.

[3] A. K. Jain, A. Ross and S. Prabhakar, "An Introduction to Biometric Recognition", IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol. 14, No. 1, pp. 4-20, January 2004.

[4] A. K. Jain, S. C. Dass, K. Nandakumar, "Soft Biometric Traits for Personal Recognition Systems", Proceedings of the Int. Conf on Biometric Authentication, LNCS 3072, pp. 731-738, Hong Kong, July 2004.

[5] A. K. Jain, K. Nandakumar, X. Lu, and U. Park, "Integrating Faces, Fingerprints, and Soft Biometric Traits for User Recognition", in Proceedings of Biometric Authentication Workshop, in conjunction with ECCV2004, LNCS 3087, pp. 259-269, Prague, May 2004.

[6] A. K. Jain, S. C. Dass and K. Nandakumar, "Can soft biometric traits assist user recognition?", in Proceedings of SPIE Vol. 5404, Biometric Technology for Human Identification, pp. 561-572, Orlando, FL, April 2004.

[7] L. Xiang, "Visual Surveillance – Scene Inventories", Master Thesis. Cranfield University, September 2007.

[8] Intel – OpenCV (http://www.intel.com/) (último acesso em 28 de maio de 2008).

[9] M. J. Jones, P. Viola, "Robust Real-Time Face Detection", Int. J. Comput. Vision, 2004.

[10] J. Kovac, P. Peer, F. Solina, "Human Skin Colour Clustering for Face Detection", in International Conference on Computer as a Tool EUROCON 2003, Ljubljana, Slovenia, September 2003.

Cálculo de Histograma em FPGA para processamento de imagens em tempo real

Jovander da Silva Freitas, Maximiliam Luppe
Depto. Eng. Elétrica – EESC – USP
{jovander,maxluppe}@sc.usp.br

Abstract

Histogram is a key element in many digital image processing applications. Histogram can provide useful global statistics data about an image, and it can be used in image compression and image segmentation algorithms. Histogram can be easily obtained through familiar software tools, but little is made on its implementation in hardware. This paper presents a system implemented in FPGA that obtain the histogram of digital images in real time. This work also presents its use in calculation of the fractal dimension. The same system can be used in future projects for real time image processing, such as histogram equalization and binarization.

1. Introdução

O Histograma é a base para numerosas técnicas de processamento no domínio espacial [1]. Ele representa a distribuição dos níveis de cinza de uma dada imagem digital, ou seja, fornece a informação sobre a quantidade de pixels que uma determinada imagem possui com relação aos possíveis níveis de intensidade. Ao analisar um histograma, é possível analisar se uma imagem foi capturada em uma condição de maior ou menor luminosidade e analisar a relação entre partes da imagem, como fundo e objetos. Um problema encontrado com freqüência em pesquisas que envolvem uso de histograma para processamento de imagens em tempo real é a variação de intensidade de luz sofrida pelo dispositivo de captura. Em alguns casos, existe a necessidade de realizar a captura e já de imediato obter o histograma.

Apesar de ser facilmente implementado em linguagens de programação, há poucas referências [2-4] que relatam qual a melhor forma de sua implementação e quais os melhores métodos para realizar tal tarefa em hardware, mais especificamente em dispositivos reconfiguráveis. Por isso o interesse de se desenvolver um sistema que obtenha o histograma de imagens obtidas a partir de uma câmera de vídeo comum de forma eficiente e funcional.

Neste trabalho é apresentada a implementação de um sistema desenvolvido em FPGA da Altera para realizar em tempo real os cálculos que envolvem a obtenção de um histograma a partir de uma imagem obtida por meio de uma câmera de vídeo. Essa implementação foi aplicada no cálculo da dimensão fractal em tempo real utilizando o método de Minkowsky [5], e pode ser aplicada em arquiteturas para binarização e tratamento de imagens.

2. Histograma

O histograma de uma imagem digital representada pela matriz IM, de NxM pixels com L níveis de cinza, é uma função discreta representada por $h(r_k) = n_k$, onde r_k é o k-ésimo nível de cinza e n_k é o número de pixels na imagem com nível de cinza r_k [1]. O histograma pode ser obtido pela seguinte algoritmo:

```
for (i=0; i<N; i++)
   for (j=0; j<M; j++)
      h[IM[i,j]]++;
```

O gráfico de um histograma apresenta a distribuição dos níveis de intensidade dos pixels e retrata problemas encontrados no processamento de imagens em geral, como as condições de luminosidade sobre uma imagem. A figura 1 apresenta este problema.

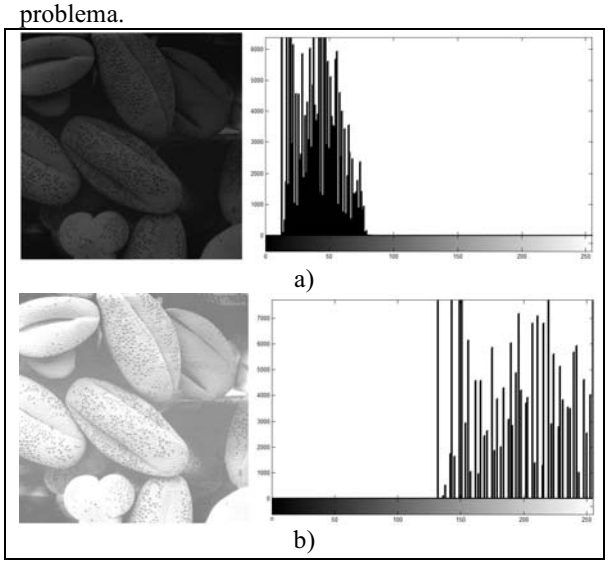


Figura 1 – Exemplos de Histograma: a) imagem com grande quantidade de pixels de tons escuros, b) imagem com grande quantidade de pixels claros. [1]

A figura (1a) mostra uma imagem escura, onde é possível notar que as barras que representam os pixels mais escuros estão concentradas no canto esquerdo do histograma. Na figura (1b), que possui uma

intensidade de luz mais clara, nota-se que as barras que representam os pixels mais próximos do branco têm maior concentração, à direita do gráfico. Além disso, a partir do histograma é possível realizar diversos processamentos no domínio espacial, como a binarização, utilizada na segmentação de imagens, a equalização, para o tratamento visual da imagem, e no cálculo da dimensão fractal pelo método de Minkowski [5].

Uma forma básica de se implementar o cálculo do histograma é utilizar L contadores, um para cada possível nível de cinza, com $\log_2(NxM)$ bits cada. Contudo, esta técnica demanda um consumo muito grande de elementos lógicos. Como apenas um contador é acessado por vez, é possível utilizar memória operando num esquema *read-modify-write*. Neste esquema, uma determinada posição de memória é acessada utilizando o valor do nível de cinza do pixel. O valor armazenado nesta posição é lido, incrementado e armazenado na mesma posição. A memória mais adequada é a memória síncrona com no mínimo duas portas. Por uma porta é feita a leitura do valor atual de $h(r_k)$, n_k, e por outra é feita a sua atualização, n_k+1, utilizando r_k, o nível de cinza do pixel atual, tanto como endereço de leitura como de escrita. Normalmente existe uma latência de um clock entre a leitura e a escrita da memória. Um problema que pode ocorrer durante a atualização é o conflito de leitura. Três técnicas básicas podem ser aplicadas para evitar este conflito [6]: Contadores, Handshaking e Clock-2x. Estas técnicas são descritas a seguir.

3. Técnicas para obtenção de histograma em *hardware*

Conforme dito anteriormente, três técnicas básicas podem ser aplicadas para evitar o conflito na obtenção do histograma utilizando memórias com no mínimo duas portas. Na técnica com contadores, um contador é mantido atualizado com o valor atual de $h(r_k)$. Se o próximo dado a ser atualizado (INPUT VALUE) é igual ao atual (ADDRB), o contador é novamente incrementado antes da atualização de $h(r_k)$. Desta forma a memória só é atualizada quando o valor do endereço de escrita (ADDRB) da memória é diferente do estado atual do dado de entrada (INPUT VALUE), conforme pode ser observado na figura 2.

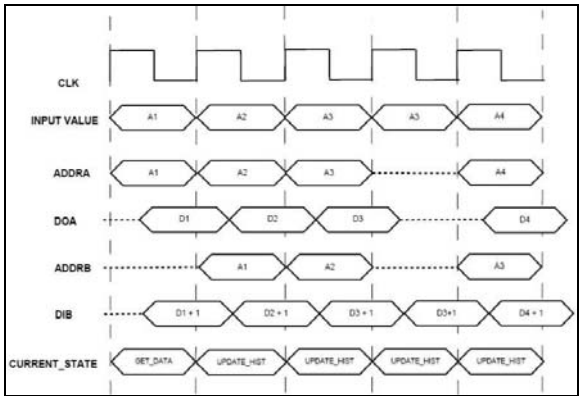


Figura 2 – Diagrama de estado do método de contadores [6]

Na técnica de Handshaking, somente quando as portas de leitura e de escrita estiverem sem conflitos, ou seja, quando o endereço de escrita for diferente do endereço de leitura, é que será realizada a atualização do histograma. Caso ocorra um conflito entre o dado de entrada (INPUT VALUE) e o endereço da segunda porta da memória (ADDRB), é invocado um estado de espera (DELAY) para que se possa realizar a atualização do histograma, conforme pode ser visto na figura 3. Observamos que a entrada de dados também deve ser atrasada.

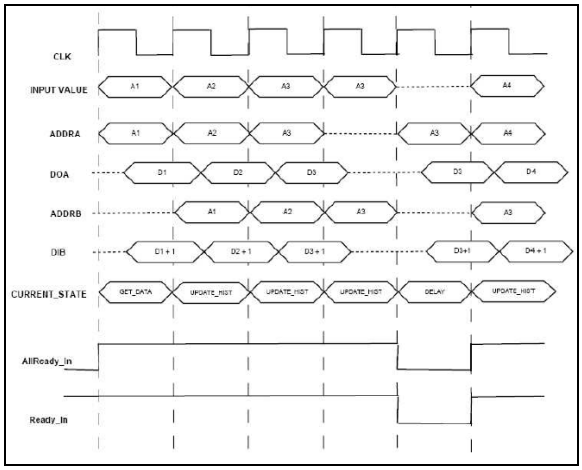


Figura 3 – Diagrama de estado do método de Handshaking [6]

Na técnica Clock-2x, o clock de entrada é dobrado para que a operação de escrita e leitura na memória seja realizado em um único ciclo de clock, evitando assim o conflito de endereços ente as duas portas da memória, conforme pode ser observado na figura 4.

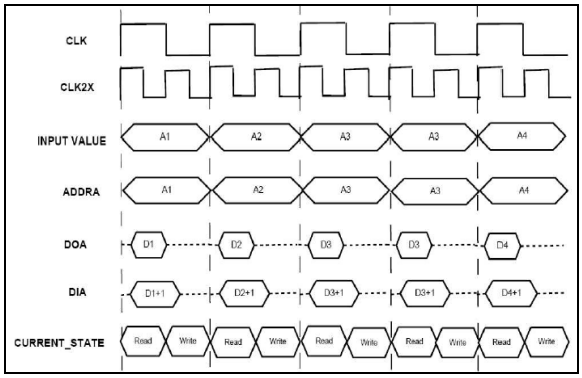


Figura 4 – Diagrama de estado do método utilizando memória de Clock-2x [6]

Na primeira parte do ciclo de clock é feita a leitura da memória. O valor lido é incrementado por meio de um contador e o resultado é armazenado no mesmo endereço na parte final do ciclo de clock. O próprio sinal de clock é utilizado para controlar a leitura/escrita da memória. Esta técnica é utilizada neste trabalho, pois tem um menor consumo de células lógicas, de acordo com Matha[6], não exige a implementação de componentes seqüenciais extras (contadores), nem força o atraso na entrada de dados (DELAY), permitindo o processamento em tempo real.

4. Implementação

Para a implementação da técnica Clock-2x foi utilizada uma memória de 3 portas: por uma porta é feita a leitura do valor atual do histograma (rdaddress_a), por outra é feita a sua atualização (waddress), e uma terceira porta (rdaddress_b) é utilizada tanto para descarregar o histograma da memória (qb) como para zerar (0) seu conteúdo para gerar um novo histograma. Na figura 5 temos o esquema da implementação do sistema.

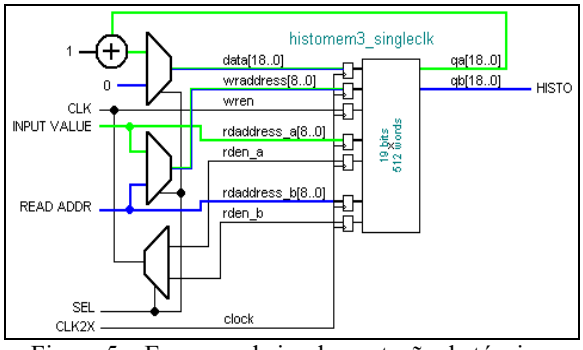


Figura 5 – Esquema da implementação da técnica Clock-2x

Durante o processo de geração do histograma, o valor do nível de cinza do pixel atual r_k (INPUT VALUE) é utilizado tanto como endereço de leitura (rdaddress_a) como de escrita (waddres), por meio do segundo multiplexador. O dado (qa) é lido durante o ciclo inicial do clock e é incrementado por meio de um somador e escrito na memória (data) no ciclo final do clock, por meio do primeiro multiplexador. Para a descarga do histograma, um contador externo é utilizado para a geração dos endereços para a leitura dos dados (READ_ADDR), por meio do segundo multiplexador, e consequentemente realiza a limpeza da memória (armazenamento do valor 0), por meio do primeiro multiplexadores. O controle destas operações é realizado por meio de multiplexador e do sinal SEL.

Este gerador de histograma foi utilizado numa arquitetura para o cálculo de dimensão fractal em tempo real[5]. Nesta arquitetura, o cálculo da dimensão fractal é obtido por meio da regressão linear de um histograma acumulativo dos níveis de cinza da imagem, sendo que estes níveis de cinza representam os valores das possíveis distâncias euclidianas relativas aos pontos de contorno dos objetos na imagem. Para se obter o histograma acumulativo, basta descarregar o histograma num circuito somador-acumulador. Cada dado do histograma acumulativo representa a área $A(\varepsilon)$ coberta pelos pixels para cada valor de distância ε representável na imagem pelas distâncias.

Uma vez calculado o histograma acumulativo, é necessário calcular o logaritmo dos dados deste para obter os dados necessários para gerar o gráfico log $A(\varepsilon)$ vs log ε e, a partir deste, obter o coeficiente angular da reta que melhor se ajusta aos dados pelo método LSM (Least Square Method). O cálculo do logaritmo é realizado pelo componente `i2048log_log`, criado a partir da megafunção `ALTFP_LOG`. Para o cálculo do coeficiente angular é utilizado o Método do Mínimo Quadrático (MMQ). Considerando um conjunto de pares ordenados (x_i, y_i) que descrevem uma reta $y = a + bx$, podemos encontrar os coeficientes a e b por meio das seguintes operações:

$$a = \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\, x_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad (1a)$$

$$b = \frac{n\left(\sum_{i=1}^{n} y_i\, x_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad (1b)$$

Como estamos interessados apenas no coeficiente angular b, uma vez que a dimensão fractal é igual a 2 – coeficiente angular do gráfico log $A(\varepsilon)$ vs log ε, e que estamos trabalhando com um conjunto fixo de distâncias, já que as distâncias representáveis são limitadas pelo tamanho da imagem, podemos observar que alguns parâmetros da equação do coeficiente angular são constantes (dependem apenas da variável x_i), além do próprio n:

$$c_0 = \left(\sum_{i=1}^{n} x_i \right) \quad (2a)$$

$$c_1 = n\left(\sum_{i=1}^{n} {x_i}^2 \right) - \left(\sum_{i=1}^{n} x_i \right)^2 \quad (2b)$$

Sendo assim, a equação para o coeficiente angular fica:

$$b = \frac{n\left(\sum_{i=1}^{n} y_i x_i \right) - c_0 \left(\sum_{i=1}^{n} y_i \right)}{c_1} \quad (3)$$

O que reduz o cálculo do coeficiente angular para apenas duas multiplicações, uma subtração e uma divisão. Todas as operações são realizadas com números inteiros. Tanto a somatória de y_i, quanto a somatória do produto de x_i com y_i, são obtidas da mesma forma que o histograma acumulativo, por meio de um componente acumulativo: um acumulador (`ACC`) para a somatória de y_i e um multiplicador acumulador (`MAC`) para a somatória do produto de x_i com y_i. O resultado destes somatórios é enviado ao componente `MMQ_cte`. Na figura 6 é apresentado o esquema geral do módulo para o cálculo da dimensão fractal em tempo real. Em azul está representada a largura dos barramentos de dados e, entre parênteses, o atraso causado por cada módulo, em ciclos de clock.

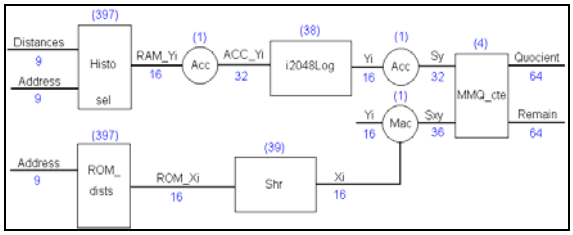


Figura 6 – Esquema geral do Módulo de Dimensão Fractal

A seguir apresenta-se a Tabela 1 com os valores médios obtidos com o processamento de alguns fractais e os valores teóricos para os mesmos. Valores menores que os teóricos eram esperados, pois os valores da dimensão fractal obtidos pelo método da salsicha de Minkowski são menores que os valores teóricos.

Curva	Teórico	Obtido
Line	1,0000	0,974
Contour of the Gosper island	1,0686	1,040
Vicksel Fractal	1,4650	1,260
Quadratic von Koch curve type 1	1,4650	1,180
Peano curve	2,0000	1,754

Tabela 1 – Comparação entre valores teóricos com os valores médios obtidos pela arquitetura desenvolvida.

5. Conclusão

Este artigo apresentou a implementação em FPGA de um sistema que obtém o histograma de uma imagem em tempo real, por meio da técnica Clock-2x, que requer a utilização de memória de pelo menos duas portas (leitura e gravação), para arquitetura apresentada foi necessário à utilização de memória de três portas, sendo uma para obtenção do histograma acumulado. Esta técnica permite realizar a leitura e a escrita da memória do histograma, num esquema *ready-modify-write*, num mesmo ciclo de clock, evitando conflitos no acesso à memória. Para demonstrar o funcionamento do sistema, o mesmo foi utilizado para realizar o cálculo da dimensão fractal em tempo real pelo método de Minkowski. Futuras aplicações do sistema, como binarização de imagens em tempo real, serão desenvolvidas.

6. Referências

[1] Gonzalez, R. C., Woods, R. E, *Digital Image Processing,* 2nd ed., Prentice Hall, 2002

[2] Garcia, E., *Implementing a Histogram for Image Processing Applications*, Xilinx, Xcell Magazine, (38), pp. 40-46, 2000

[3] Jamro, E., Wielgosz, M., Wiatr, K., *FPGA Implementation of the Strongly Parallel Histogram Equalization*, In Proc. 10th IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems, April 2007

[4] Shahbahrami, A., Hur, J. Y., Juurlink, B. H. H., Wong, S., *FPGA Implementation of Parallel Histogram Computation*, 2nd HiPEAC Workshop on Reconfigurable Computing, pp. 63-72, Gothenburg, Sweden, January 2008

[5] Luppe, M., *Implementação em Hardware do método de Minkowsky para o cálculo da dimensão Fractal,* VI Southern Conference on Programmable Logic Conference (SPL'2010), Designer Forum, pp. 47-51, Porto de Galinhas, PE, Brazil, March 2010

[6] Matha, M. H., *Integration of reduction operators in a compiler for FPGAs*, Dept. of Electrical and Computer Engineering, Colorado State University, http://www.radar.colostate.edu/ece/current_students/grad_exams/matha.pdf

Localização e Reconhecimento de Caracteres em Placas de Automóveis

Leonardo Augusto de Oliveira, Adilson Gonzaga
Laboratório de Visão Computacional (LAVI), Universidade de São Paulo (USP)
Av. Trabalhador São-Carlense, 400
São Carlos - SP - Brasil, CEP 13566-590
leoadeoliveira@gmail.com, agonzaga@sc.usp.br

Abstract

This paper discuss techniques of computer vision applied to the process of automatic recognition of license plates, and provides a simple and efficient solution for localization, segmentation and character recognition. The Localization stage is based on the search of image regions of highest concentration of abrupt changes in intensity. The Segmentation step consists in processing the image by applying spatial filtering and morphological operations in order to obtain a binary image without noise and the objects well defined. The stage of Recognition consists in examining each object in comparison with a set of masks that indicate his identification between the 36 alphanumeric characters. The present work aimed at balancing simplicity and efficiency. Each single module has an accuracy rate around 90%.

Keywords: Computer vision, Plate recognition, LPR.

1. Introdução

O crescente interesse pela automatização de processos e introdução de novas tecnologias nos campos de trabalho são fatores motivadores para diversas pesquisas no ramo de reconhecimento de padrões em uma imagem. Visão computacional e reconhecimento de caracteres para sistemas de reconhecimento de placas de automóveis (LPR, do inglês *License Plate Recognition*) são módulos principais para os sistemas inteligentes de infraestrutura, como os de controle de tráfego e de cobrança de pedágio [1].

As técnicas utilizadas em reconhecimento de caracteres são variadas: reconhecimento ótico, por sistemas digitais, utilizando redes neurais e algoritmos genéticos. Na década de 70, quando iniciaram-se os trabalhos no assunto, fatores computacionais eram limitantes, não permitindo aplicabilidade direta das técnicas desenvolvidas. Com o avanço da microeletrônica e das arquiteturas de computadores, técnicas digitais passaram a ser mais empregadas, sendo fortemente difundidas e ainda estudadas nos dias de hoje [2]. Atualmente, métodos computacionais inteligentes permitem a abordagem do problema utilizando Lógica Fuzzy, Redes Neurais Artificiais (RNA) e Algoritmos Genéticos [3].

O objetivo de um sistema de reconhecimento automático de placas de automóveis é fornecer uma saída simples para uma entrada complexa. A entrada é uma imagem estática, contendo diversos elementos desnecessários e o elemento essencial: a placa do automóvel. Computacionalmente, essa entrada é uma matriz numérica que contém informações dos píxeis da imagem. A saída pode ser, por exemplo, uma sequência de caracteres, uma *string*, com os números e letras que compõe a placa.

Este trabalho descreve um método de localização de regiões de interesse (ROI, do inglês *Region of Interest*) e reconhecimento de padrões aplicado à identificação de placas de automóveis. O ojetivo é obter um algoritmo de LPR simples e eficiente que possa ser usado com imagens de placas em carros brasileiros.

O texto está organizado da seguinte maneira: na seção 2 são apresentadas técnicas de LPR difundidas, usando-as como base para a formulação da nossa abordagem, descrita na seção 3; na seção 4 são apresentados os resultados obtidos com o algoritmo desenvolvido; a seção 5 conclui sobre a metotodologia proposta.

2. Técnicas de LPR

São três etapas básicas no processo de reconhecimento de uma placa de automóvel: **(a)** localização e extração da placa, **(b)** segmentação de caracteres e **(c)** reconhecimento dos caracteres.

2.1. Localização

A Localização consiste em realizar um rastreamento em toda a imagem a fim de identificar a exata região que contém somente a placa do automóvel e extraí-la.

Grande parte dos trabalhos, no que diz respeito à Localização, baseia-se em alguma característica das placas que a destacam na imagem de um carro, como a alta frequência de mudança de brilho na região da placa. Brandão, Sequeira e Albuquerque [4] propõe um algoritmo "baseado em operações morfológicas multi-estágio". A busca por candidatos à placa em uma imagem considera três características básicas, para as quais foram desenvolvidos estágios independentes: **(a)** são regiões de alto contraste, **(b)** possuem sequência alinhada de objetos (os caracteres) e **(c)** são retangulares. O algoritmo percorre os três estágios em busca de regiões de interesse que obedeçam aos critérios. Os resultados indicam taxa de sucesso de 89,15% com o uso do estágio *(a)*, 94,21% com o uso dos estágios *(a)* e *(b)* e 97,18% de sucesso com o uso dos três estágios.

C. Anagnostopoulos, I. Anagnostopoulos, Loumos e Kayafas [5] apresentam um estudo acerca das diversas abordagens do tema. Os autores reiteram a observação de que métodos baseados em detecção de bordas são altamente susceptíveis a ruídos indesejados. Eles analisam que os ruídos considerados não são somente devido à má qualidade da imagem ou má iluminação, mas a objetos que têm características semelhantes às placas. Os autores analisam que o uso de técnicas morfológicas para eliminar bordas indesejadas previamente à aplicação do método de detecção de bordas resulta em índices de sucesso relativamente altos e computacionalmente rápidos se comparados a outros métodos.

Os autores propõe um método de análise estatística. A técnica "Janelas Concêntricas Deslizantes" (do inglês *Sliding Concentric Windows* - SCW) foi desenvolvida para identificar irregularidades locais na imagem utilizando medidas estatísticas como desvio padrão e média. Duas janelas retangulares concêntricas A e B (de tamanhos diferentes) deslizam pela figura, obtendo valores estatísticos que denunciam a característica irregular (alta frequência de mudança de brilho) na localidade: quando a razão das medidas estatísticas das janelas A e B for maior que um valor definido T, a janela retangular A é considerada uma ROI. Para definir os tamanhos das janelas A e B e o valor de T os autores realizaram testes, buscando os melhores resultados. Espera-se que as janelas A e B com proporção igual à das placas de automóveis sejam as melhores, como os testes comprovaram. Em relação ao parâmetro T, os autores assumem que não há evidência de como obtê-lo e portanto a melhor maneira é por tentativa-e-erro.

Martinsky [6] propõe um algoritmo em que se busca uma faixa horizontal na qual a ROI deve estar contida. Dentro desta faixa, calcula-se a projeção vertical e análise semelhante é feita para identificar a posição da placa dentro daquela faixa. Com algumas candidatas selecionadas, o autor faz uma análise heurística de características que validarão a melhor candidata. São consideradas 4 características: **(a)** a altura da região, sendo preferidas as de menor altura, **(b)** a altura do pico da projeção horizontal, sendo preferidas as regiões onde forem identificados maior quantidade de bordas verticais, **(c)** valor da área sob o gráfico deste pico e **(d)** proporção do retângulo. O autor atribui pesos a cada característica, de acordo com critérios empíricos, e realiza uma seleção das melhores candidatas a partir dos valores obtidos.

2.2. Segmentação

A segmentação de caracteres consiste em separar cada objeto correspondente a um caractere para análise e reconhecimento em etapa posterior.

O uso de métodos mais simples ou mais complexos, bem como a taxa de sucesso deles, depende muito de como foi feita a localização da placa e da qualidade dos algoritmos de pré-processamento das imagens, principalmente de binarização.

Chang, L. Chen, Chung e S. Chen [7] realizam o agrupamento de pixels pela rotulação daqueles que estejam de alguma forma conectados. Para definir quais objetos interessam e quais são descartáveis, o algoritmo, primeiramente, exclui os objetos que tenham proporção fora de uma faixa pré-estabelecida, que é típica dos caracteres das placas analisadas. Em seguida, os objetos restantes devem estar alinhados. Para este cálculo, utiliza-se a Transformada Hough com os centróides para determinar o alinhamento dos objetos. Finalmente, se o número de objetos for maior que um número pré-estabelecido, elimina-se um a um a partir do de menor tamanho. Os autores, porém, consideram diversas possibilidades de erros, como um caractere ser composto de dois objetos ou dois caracteres formarem um único objeto. Estes fatores prejudicam a segmentação dos caracteres. Considerá-los torna o projeto extremamente robusto. Os autores propõe um algoritmo que realiza sequencialmente operações de exclusão, junção e divisão de objetos. A cada sequência, faz-se as análises já citadas: proporção, alinhamento e quantidade de objetos. O processo é repetido até que se chegue a uma solução satisfatória. É um algoritmo complexo, mas com taxa de 95,6% de sucesso para a identificação de caracteres com uso de imagens de entrada de alta complexidade.

Conci e Monteiro [8] fazem uma análise mais simples: após a rotulação dos píxeis e obtenção dos objetos contidos na imagem, estimam-se os limites superior e inferior dos caracteres na placa. Dessa forma, os objetos que não obedeçam a esse critério são eliminados e apenas os 7 caracteres interessantes restarão.

2.3. Reconhecimento

O reconhecimento de caracteres é a etapa em que um objeto é analisado e associado a um único caractere alfanumérico.

Diversos autores utilizam redes neurais para o processo de Reconhecimento. O objetivo básico de sistemas baseados em redes neurais é realizar um treinamento prévio do programa com um banco de dados específico para o treinamento, para então fazer a análise das placas de outro banco de dados.

A análise das projeções vertical e horizontal dos objetos é utilizada por Belvisi, Freitas, Figueiredo e Marcovitch [9]. A comparação é feita pelas projeções verticais e horizontais das funções $f(x,y)$ e $w(x,y)$ do objeto e da máscara. A medida de igualdade entre as projeções do objeto e da máscara é dada através de um limite máximo percentual de pontos diferentes. Os autores não especificam um valor, mas assinalam uma alta taxa de acerto do sistema.

Conci e Monteiro [8] aplicam o método dos Momentos Invariantes de Hu. Espera-se que cada caractere tenha um valor definido para os momentos que o diferem dos demais caracteres. Definidos os valores de momentos para cada caractere de **0** a **9** e de **A** a **Z**, e calculados os momentos do objeto desconhecido, basta estabelecer com qual deles há maior similaridade. Os autores obtiveram 99% de acerto utilizando um banco de imagens simples, porém com um longo tempo de processamento, já que os cálculos dos Momentos de Hu são computacionalmente dispendiosos.

Nosso trabalho baseia-se na morfologia e estatísticas das imagens para realizar o reconhecimento dos caracteres. A análise de histogramas e de características morfológicas exige, porém, boa segmentação e processamento prévio da imagem de modo a aproximar ao máximo o caractere desconhecido de sua máscara comparativa previamente armazenada, que serve como gabarito para o sistema.

3. Material e método

Para a implementação do algoritmo proposto neste trabalho, foram utilizadas imagens próprias e outras disponíveis na Internet. São três bancos de imagens: BD1, BD2 e BD3. O BD1 foi obtido junto ao Laboratório de Processamento Digital de Sinais e Imagens (LPDSI) e é composto de 75 imagens em níveis de cinza, obtidas em cancelas de pedágio no estado do Rio de Janeiro [10]. As imagens são consideradas de má qualidade. Os bancos BD2 e BD3 são compostos por, respectivamente, 79 e 127 imagens, capturadas no estacionamento do *campus* da Universidade. As imagens têm boa qualidade. O BD3 diferencia-se do BD2 por este ter sido utilizado como conjunto de treinamento para o desenvolvimento do algoritmo, enquanto aquele é utilizado apenas como conjunto de teste. Para obtenção das imagens dos bancos de imagens BD2 e BD3 foi utilizada uma câmera do modelo Sony Cyber-shot DSC-H10, em modo automático, sem flash e com resolução mínima (640×480 píxeis).

O algoritmo foi desenvolvido utilizando MATLAB versão R2008b com o *Image Processing Toolbox* (IPT). O *software* foi desenvolvido em um *notebook* modelo HP Pavillion dv4000, com processador AMD Turion X2 de 2,1GHz, 4G de memória RAM e sistema operacional Windows 7 Professional.

3.1. Localização da região da placa

O método proposto para Localização da placa é baseado na busca por regiões da imagem onde o gradiente horizontal tem valor mais significativo. O vetor gradiente $\nabla f_y = \frac{\delta f}{\delta y}$ terá os valores correspondentes às diferenças entre os píxeis vizinhos na direção horizontal. Uma imagem resultante dessa operação (considerando o módulo do gradiente) acusa regiões de maior mudança brusca de intensidade onde o gradiente tiver maior valor.

O método proposto baseia-se na varredura da imagem gradiente feita por janelas retangulares de tamanho proporcional à placa. Em cada ponto da varredura é calculada a média dos píxeis daquela região, comparando sempre com a maior média já encontrada. Ao fim da varredura, a região com a maior média entre todas será a placa do automóvel. A Figura 1 ilustra o procedimento.

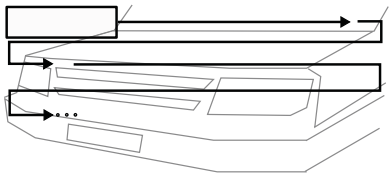


Figura 1: Varredura da imagem por janelas

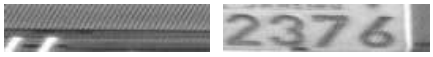
O processo desenvolvido exige uma filtragem de suavização para retirar ruído e eliminar a possibilidade de detecção de elementos indesejados. A grade frontal presente em alguns modelos de carros é um exemplo de elemento indesejado. Estes, porém, costumam apresentar frequência de mudança de intensidade maior do que os caracteres da placa. O filtro de suavização aliado à operação morfológica de abertura promovem a eliminação dessas altas frequências, fazendo com que a região de maior densidade de magnitude do gradiente seja a região da placa.

Aplica-se suavização linear de média e abertura morfológica em escala de cinza, de ordem 5×5. A abertura é uma operação de erosão seguida de uma dilatação, usando um elemento estruturante quadrado 5×5. Basicamente,

elementos da imagem de largura menor do que 5 píxeis serão eliminados enquanto elementos do plano de fundo serão preenchidos. A Figura 2 exemplifica o processamento da imagem e a Figura 3 os resultados da Localização.

(a) Original (b) Suavização (c) Abertura

Figura 2: Filtragem da imagem para Localização



(a) Sem filtragem (b) Com filtragem

Figura 3: Resultados da Localização

3.2. Processamento de imagem da placa

Com a extração da região da placa, o histograma equivalente apresenta dois picos bem definidos (caracteres e fundo). A próxima etapa, visando à segmentação dos caracteres, é a binarização da região da placa segmentada. Foi utilizado o método *Basic Global Thresholding* (BGT) para definir o valor de limiar [2]. O algoritmo é composto por 5 passos: **(i)** Definir um valor estimado inicial T_0. **(ii)** Dividir os píxeis em dois grupos: G_1, com todos aqueles com intensidade $\leq T$, e G_2, com os restantes. **(iii)** Calcular as médias m_1 e m_2 dos píxeis de G_1 e G_2. **(iv)** Computar o novo valor de limiar por $T = \frac{m_1+m_2}{2}$. **(v)** Caso a diferença entre T e T_0 seja menor que um valor pré-definido ΔT, finalizar. Caso contrário, T_0 assume o valor de T e retorna-se ao passo *(ii)*. O método funciona bem quando há clara definição entre picos e vale no histograma da imagem.

3.3. Segmentação de caracteres

A entrada para o processo de Segmentação dos caracteres é uma imagem binária, com poucos agrupamentos de píxeis além dos caracteres.

Todos os agrupamentos a serem segmentados são analisados por heurísticas. O objetivo é eliminar todos os objetos que excedam ao grupo de 3 letras e 4 números. São atribuidos valores a uma variável em cada teste. Quanto mais distante do valor esperado o objeto estiver, maior será o valor dessa variável. Ao final dos testes, as 4 variáveis (H_A, H_B, H_C e H_D) serão somadas e os 7 agrupamentos que apresentarem os menores valores totais H_T serão os caracteres.

Os 4 testes heurísticos propostos consideram as seguintes características: posicionamento do centróide, altura do objeto, *aspect ratio* ($AR = Largura/Altura$) e área do objeto.

- **Teste A** - Espera-se que o centróide de um objeto que corresponda a um caractere esteja na porção central da imagem, na direção vertical. Portanto, neste teste obtém-se a diferença entre a posição do centróide e a metade da altura da imagem. Foi atribuído peso $X = 50$ a essa diferença, resultando em H_A.

- **Teste B** - Obtendo a altura de todos os objetos da imagem, sabe-se que o valor da mediana da sequência de valores de altura corresponderá necessariamente à altura de um caractere. Dessa forma, essa altura será considerada como gabarito para cada caractere. Portanto, calcula-se H_B como o módulo da diferença entra a altura do objeto e a altura mediana, normalizada pela altura da imagem e com peso X igual a 80.

- **Teste C** - Entre os caracteres de uma placa de automóvel, a letra **I** ou o número **1** podem ser considerados aqueles de menor *aspect ratio* e letras como **M** ou **G** podem ser consideradas as de maior *aspect ratio*. Deve-se estabelecer um valor mínimo e um valor máximo para *aspect ratio* e testar os objetos quanto a essa característica. Para o banco BD1, foram obtidos $AR_{min} = 0,50$ e $AR_{max} = 1,75$. Para os demais, foram obtidos $AR_{min} = 0,15$ e $AR_{max} = 1,00$. O valor de H_C é dado por:

$$H_C = \begin{cases} X\frac{|AR_{min}-AR|}{AR} & \text{para } AR < AR_{min} \\ X\frac{|AR_{max}-AR|}{AR} & \text{para } AR > AR_{max} \\ 0 & \text{caso contrário} \end{cases}$$

O valor do peso X é utilizado para maximizar os valores daqueles objetos que apresentem *aspect ratio* muito distante do esperado. Estabeleceu-se:

$$X = \begin{cases} 60 & \text{para } AR \leq \frac{AR_{min}}{1,5} \\ 20 & \text{para } \frac{AR_{min}}{1,5} < AR \leq \frac{AR_{min}}{1,2} \\ 10 & \text{para } \frac{AR_{min}}{1,2} < AR \leq AR_{min} \\ 90 & \text{para } AR \geq 1,5AR_{max} \\ 30 & \text{para } 1,2AR_{max} \geq AR < 1,5AR_{max} \\ 15 & \text{para } AR_{max} \geq AR < 1,2AR_{max} \end{cases}$$

- **Teste D** - O último teste diz respeito à área dos objetos. Com o *aspect ratio* mínimo e máximo esperados, definidos anteriormente, e a altura esperada, definida

como a mediana das alturas dos objetos contidos na imagem, tem-se:

$$Área = AspectRatio \cdot AlturaMediana^2$$

Portando, pode-se considerar $A_{min} = AR_{min} \cdot AM^2$ e $A_{max} = AR_{max} \cdot AM^2$. Assim, H_D é calculada por:

$$H_D = \begin{cases} X\frac{|A_{min}-A|}{A} & \text{para } A < A_{min} \\ X\frac{|A_{max}-A|}{A} & \text{para } A > A_{max} \\ 0 & \text{caso contrário} \end{cases}$$

Neste teste, estabeleceu-se, empiricamente:

$$X = \begin{cases} 70 & \text{para } A \leq \frac{A_{min}}{2,0} \\ 10 & \text{para } \frac{A_{min}}{2,0} < A \leq \frac{A_{min}}{1,5} \\ 5 & \text{para } \frac{A_{min}}{1,5} < A \leq A_{min} \\ \\ 70 & \text{para } A \geq 2,0A_{max} \\ 10 & \text{para } 2,0A_{max} > A \geq 1,5A_{max} \\ 5 & \text{para } 1,5A_{max} > A \geq A_{max} \end{cases}$$

A altura do objeto é fator mais determinante do que o posicionamento do centróide, pois este tem uma valor esperado menos constante do que a altura: alguns caracteres tem o centróide deslocado em relação ao centro geométrico, como a letra **L**, enquanto a altura esperada de todas os caracteres é a mesma. Portanto, o Teste B deve ter maior peso no resultado final do que o Teste A. Os pesos foram definidos empiricamente por tentativa-e-erro.

3.4. Reconhecimento

O Reconhecimento é realizado por correlação da imagem obtida de um caractere com diversas imagens pertencentes a uma base de dados padrão: as máscaras.

As máscaras foram divididas em dois bancos: o primeiro é formado por imagens de todos os 10 caracteres numéricos e 26 letras do alfabeto escritos com a fonte *Mandatory*, definida oficialmente pelo Conselho Nacional de Trânsito (CONTRAN) pela Resolução 231, de 2007; o segundo é formado por imagens dos 36 caracteres alfunuméricos utilizando a fonte *DIN Mittelschrift*, usual nos carros emplacados antes de 2007.

A comparação entre o caractere e as máscaras é feita por Correlação 2-D, por meio do Coeficiente de Correlação de Pearson (CCP), uma medida estatística para comparação linear entre dois conjuntos de dados. Seu valor, em módulo, varia de 0 a 1. Quanto mais perto de 1 maior é o grau de dependência estatística linear entre as variáveis. No outro oposto, quanto mais próximo de zero, menor é a força dessa relação [11]. A Equação 1 representa o CCP, denotado por ρ, entre dois vetores.

$$\rho = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2} \qquad (1)$$

A imagem do caractere desconhecido é comparada, por meio do cálculo do CCP, com as 72 máscaras. A máscara mais compatível com o caractere, ou seja, com o maior valor do CCP entre ambos, indicará qual é a letra ou número correspondente.

4. Resultados

A Tabela 1 apresenta os resultados de desempenho do algoritmo de Localização da placa dentro das imagens em cada banco.

Tabela 1: Resultados do processo de Localização

Banco	Total	Acertos	Taxa de acerto
BD1	75	73	97,3%
BD2	79	77	97,5%
BD3	127	116	91,3%

Apesar da má qualidade das imagens do BD1, foram obtidos ótimos resultados. A taxa de 97,3% alcançada neste trabalho é extremamente positiva, visto que foi utilizado um método simples e rápido.

O desempenho do algoritmo de Segmentação dos caracteres é apresentado na Tabela 2. O total de caracteres refere-se ao número de placas corretamente localizadas vezes 7.

Tabela 2: Resultados (por caractere) da Segmentação

Banco	Total	Acertos	Taxa de acerto
BD1	511	455	89,0%
BD2	539	493	91,5%
BD3	812	774	95,3%

Os resultados do BD3 são os mais animadores. É o maior banco de imagens utilizado no trabalho e apresentou taxa de acerto de 95,3%. As imagens que o compõe não foram utilizadas no conjunto de treinamento, mas somente no conjunto de testes, o que confere maior significância aos resultados obtidos.

A taxa de acerto, ou taxa de reconhecimento final, pode ser vista na Tabela 3.

Tabela 3: Resultados (por caractere) do processo de Reconhecimento considerando os acertos nas etapas anteriores

Banco	Total	Acertos	Taxa de acerto
BD1	455	428	94,1%
BD2	493	460	93,3%
BD3	774	724	93,5%

Os resultados combinados das três etapas são apresentados na Tabela 4. Aqui são consideradas todas as imagens dos bancos vezes 7 caracteres a serem reconhecidos por placa.

Tabela 4: Resultado geral da abordagem proposta

Banco	Total	Acertos	Taxa de acerto
BD1	525	428	81,5%
BD2	553	460	83,2%
BD3	889	724	81,4%
Todos	1967	1612	82,0%

Para se analisar os tempos de processamento do programa desenvolvido com a metodologia proposta, realizou-se cinco execuções do mesmo código e foram obtidos os melhores e piores tempos de execução. O tempo médio total para localizar, segmentar e reconhecer os caracteres de uma placa em uma imagem foi de aproximadamente 2,2 segundos.

5. Conclusão

A taxa de acerto acima de 80% em tempo de processamento de 2,2 segundos foi considerada razoável. Ressalva-se, porém, que são exigidas melhores taxa de acerto e tempo de processamento para aplicação prática do algoritmo em tempo real. Gonzalez e Woods [2] afirmam que o reconhecimento com sucesso será fortemente influenciado por uma segmentação precisa. Acreditamos que trabalhos futuros devem focar em melhorias no processo de segmentação para buscar incremento na taxa de acerto e diminuição do tempo de processamento.

Percebe-se que as taxas de acerto do BD1 são próximas às taxas dos outros bancos, mesmo com a utilização de imagens visivelmente de menor qualidade. As imagens do BD1 não apresentam, no entanto, os problemas de iluminação e sombra que as imagens dos outros bancos apresentam. A simples combinação de qualidade de fotografia e planejamento de como obter as imagens, eliminando efeitos de sombra, deve produzir resultados extremamente satisfatórios, próximos à taxa de 95%. Esta afirmação é fundamentada devido a observações dos resultados acima da média com imagens de boa qualidade sem problemas de iluminação ao longo do desenvolvimento do sistema.

O resultado final deste projeto é encorajador, visto que o desenvolvimento de um sistema prático e totalmente aplicável no mundo atual pode ser realizado com algoritmos relativamente simples.

Agradecimentos

Os autores agradecem à FAPESP pelo suporte financeiro à esta pesquisa.

Referências

[1] C. N. Anagnostopoulos, I. Anagnostopoulos, I. D. Psoroulas, V. Loumos, and E. Kayafas, "License plate recognition from still images and video sequences: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, pp. 377–391, September 2008.

[2] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. New Jersey: Pearson Prentice Hall, third ed., 2008.

[3] T. J. de Campos, "Reconhecimento de caracteres alfanuméricas de placas em imagens de veículos," Mestrado, Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2001.

[4] T. Brandão, M. M. Sequeira, and M. Albuquerque, "Multi-stage morphology-based license-plate location algorithm," *5th International Workshop on Image Analysis for Multimedia Interactive Services*, 1999.

[5] C. N. Anagnostopoulos, I. Anagnostopoulos, V. Loumos, and E. Kayafas, "A license plate recognition algorithm for intelligent transportation system applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, pp. 377–392, September 2006.

[6] O. Martinsky, "Algorithmic and mathematical principles of automatic number plate recognition systems," B.Sc. Thesis, Faculty of Information Technology, Brno University of Technology, Brno, 2007.

[7] S.-L. Chang, L.-S. Chen, Y.-C. Chung, and S.-W. Chen, "Automatic license plate recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, pp. 42–53, March 2004.

[8] A. Conci and L. H. Monteiro, "Reconhecimento de placas de veículos por imagens," Mestrado, Universidade Federal Fluminense, Niterói, 2004.

[9] R. Belvisi, R. Freitas, R. Figueiredo, and D. Marcovitch, "Um sistema de reconhecimento automático de placas de automóveis," *Anais do XIX Congresso Nacional da Sociedade Brasileira de Computação - Encontro Nacional de Inteligência Artificial*, vol. 4, pp. 537–539, Julho 1999.

[10] LPSDI, "LPR_ImageDataBase_Sample." Disponível em <http://www.cbpf.br/cat/pdsi/downloads/LPR_ImageDataBase_Sample.zip>, Acesso em 10/11/2010.

[11] D. B. Figueiredo Filho and J. A. Silva Júnior, "Desvendando os mistérios do Coeficiente de Correlação de Pearson," *Revista Política Hoje*, vol. 19, no. 1, pp. 115–146, 1999.

Detecção de Tomadas em Vídeos de Futebol Utilizando Teoria da Informação Não-Extensiva

Guilherme A. Wachs-Lopes Werner Fukuma
Paulo S. Rodrigues

Centro Universitário da FEI São Paulo, Brazil

Resumo

Currently, Digital TV systems present several challenges in the area of video and image analysis. Among them, there are the detection of logo time exposure in sports events such as soccer. An important early step is the shot (camera) type classification due to the complexity of scene analysis. A well known example is the classification of these shots as main camera shot and secondary camera shot. The main camera shot detection is a step that all the remaining processes depend of. Recent works show that the histogram analysis based on HSV system have generated promising results. Furthermore, researches have shown that the video and image analysis using non-extensive entropy as a classification tool is a new approach of investigation. In this paper we propose the use of non-extensive entropy for a binary classification of main camera shot. The results confirm the performance found in the literature.

1. Introdução

A análise de vídeos de futebol é uma área que tem ganhado muita atenção principalmente de emissoras de TV. Seu principal interesse remete à necessidade de se calcular a quantidade e o tempo de propagandas que estão presentes em cada evento.

Muitos trabalhos têm sido propostos para resolver este problema. Em [Yeh et al., 2005], os autores consideram duas características para detectar comerciais. A primeira consta de características específicas de comerciais e a segunda está relacionada com a detecção de cenas. Outros trabalhos como [Hsu et al., 2004, Kuhmunch, 1997] utilizam *template matching* para encontrar a ocorrência de propagandas.

Entretanto, sabe-se que há diversos tipos de tomadas presentes em uma transmissão completa de futebol tais como: tomadas de curto e longo alcance. Uma tomada de curto alcance é aquela que é caracterizada por imagens detalhadas, com foco em um determinado jogador ou objeto. Nesses tipos de tomadas, as propagandas são raramente encontradas nas camisetas dos jogadores. Por outro lado, uma tomada de longo alcance mostra a visão geral sobre o evento. As propagandas normalmente são encontradas em letreiros, placares ao redor do campo. O reconhecimento dessas tomadas pode ser útil para eliminar partes do vídeo onde não há propagandas como, por exemplo, tomadas com zoom em jogadores. Além disso, pode-se utilizar detectores específicos para cada tipo de câmera, resultando em uma melhoria na qualidade do reconhecimento.

Com o objetivo de utilizar a informação de tomada, o trabalho de [Watve and Sural, 2008] propôs um método que, inicialmente, detecta o tipo de cada cena e utiliza segmentação para encontrar possíveis regiões de interesse, que são regiões que podem potencialmente conter *outdoors*. Finalmente, utilizando *template matching*, os *outdoors* são reconhecidos individualmente.

Trabalhos recentes na área de análise de imagens [Rodrigues and Giraldi, 2009], principalmente baseados em mecânica estatística e informação mútua [Esqueff, 2002], sugerem que imagens naturais podem ser melhor estudadas, caso sejam consideradas sistemas não-extensivos. Assim, o presente trabalho propõe analisar uma imagem de evento esportivo, como futebol, não como um sistema físico tradicional (como tendo distribuição de probabilidades de características com igual importância no cálculo da informação), mas como um sistema físico cujos elementos correlacionados entre si possuem importância ponderada. Desta forma, tais elementos, como características de cor, podem ser modelados como um sistema físico não-extensivo Tsalliano [Tsallis, 1988]. Resultados experimentais mostram que, considerar a informação mútua entre *frames* consecutivos melhora a detecção de transições em eventos distintos como câmera 1 e não-câmera 1, tratando-se essa a conclusão e contribuição principal desse trabalho. Assim, é proposto neste artigo a medida da informação mútua entre *frames* consecutivos, calculadas sob a distribuição de probabilidade da mantissa em um sis-

tema *HSV*, utilizando uma base supervisionada de 45.000 *frames*. A medida do limiar de corte de separação entre as classes é calculada como aquela que maximiza a área sob a curva *ROC*.

Este artigo é organizado da seguinte forma. As Seções 2 e 3 apresentam os conceitos presentes na teoria da informação. Na Seção 4 descrevemos o modelo e os experimentos. Finalmente, na Seção 5 apresentamos e discutimos os resultados obtidos.

2. Sistemas não-extensivos

Muitos sistemas estudados em áreas clássicas como a mecânica estatística e, até mesmo, a termodinâmica, apresentam características macroscópicas que podem ser investigadas estatisticamente a partir de características microscópicas. Esses sistemas também são conhecidos como sistemas extensivos e possuem uma das propriedades físicas mais conhecidas chamada entropia, cujo tipo mais estudado é a entropia de *Shannon*, dada pela Eq. (1):

$$S = \sum_{i}^{k} p_i \log p_i \tag{1}$$

onde k é a quantidade de estados do sistema e p_i é a probabilidade do estado i ocorrer no sistema.

Contudo, esse tipo de entropia pode não gerar os mesmos resultados esperados para muitos sistemas que apresentam características específicas, tais como: interações de longo alcance, tanto espacial quanto temporal, e comportamento fractal nas periferias.

Partindo desse princípio, Tsallis propôs uma generalização da entropia tradicional, criando o conceito de entropia não-extensiva, definida por [Tsallis, 2001, Tsallis, 1988, Tsallis, 1999]:

$$S_q = k\frac{1 - \sum_{i=1}^{n} p_i^q}{q-1} \tag{2}$$

onde q é chamado de parâmetro de não-extensividade. É importante notar que, quando q tende à 1, a Eq. (2) resume-se à entropia tradicional de *Shannon*, sendo portanto uma generalização da mesma.

3. Entropia Relativa

Definida em 1951 por Kullback e Leibler para sistemas tradicionais, a Entropia Relativa é uma medida de divergência estatística entre duas distribuições probabilísticas. Alguns trabalhos referem-se à Entropia Relativa também como "Distância de *Kullback-Leibler*", "Divergência I" e "Ganho de Informação de *Kullback-Leibler*". A Entropia Relativa é definida como sendo:

$$D_{KL}(P, P') = \sum_{i=1}^{k} p_i \cdot \log \frac{p_i}{p'_i} \tag{3}$$

onde P e P' são as distribuições e k o número de estados do sistema físico considerado. É importante destacar que, para aplicar a Eq. (3), o alfabeto das distribuições deve ser o mesmo.

A Entropia Relativa isoladamente não deve ser considerada como uma medida de distância métrica, uma vez que não atende à propriedade da desigualdade triangular. Então,

$$D_{KL}(p, p') \neq D_{KL}(p', p) \tag{4}$$

Dessa forma, em [Jeffreys, 1939] foi proposta uma versão simétrica para entropia relativa:

$$D(p, p') = D_{KL}(p, p') + D_{KL}(p', p) \tag{5}$$

Em 1998, [Borland et al., 1998] propôs a generalização da entropia relativa para sistemas não-extensivos, adicionando o parâmetro entrópico q à comparação estatística entre duas distribuições, conforme a Eq. (6):

$$D_{KL_q}(p, p') = \sum_{i=1}^{k} \frac{p_i^q}{1-q} \cdot (p_i^{1-q} - p_i'^{1-q}) \tag{6}$$

A vantagem da utilização da divergência de *Kullback-Leibler* extendida é a adição do parâmetro q como um ajuste fino na equação. Dessa forma, podemos obter um q que maximiza os resultados.

4. Metodologia

Neste trabalho, propomos a classificação dos *frames* como: tomadas de câmera principal (classe 1) e tomadas de câmera secundária(classe 2). Então, podemos considerar este sistema como um classificador binário.

Genericamente falando, há duas maneiras de solucionar este tipo de problema. A primeira delas é através da análise direta da imagem, utilizando técnicas de reconhecimento de objetos inseridos na cena. Esse processo requer geralmente o uso de algoritmos computacionalmente pesados, muitas vezes inviáveis para tempo real, uma vez que demandam heurísticas com alto nível de abstração. A segunda maneira utiliza descritores estatísticos para extrair informações relacionadas à quantidade de informação.

A maneira mais tradicional para se medir a quantidade de informação em uma distribuição é através da entropia clássica de *Shannon*. Porém, o surgimento da entropia não-extensiva de *Tsallis* permitiu a inclusão de um novo parâmetro que permitiu o uso da entropia para sistemas onde a teoria clássica não era válida. Um exemplo é o trabalho de [Rodrigues and Giraldi, 2009] onde foi proposto um método de segmentação de imagens com cálculo automático do parâmetro q. Os resultados mostraram que essa é uma técnica promissora para o tratamento de imagens naturais.

Figura 1. Exemplo de alguns quadros da base. Os quadros 2, 3, 4, 6 e 8 são considerados câmera 1, e os demais não câmera 1

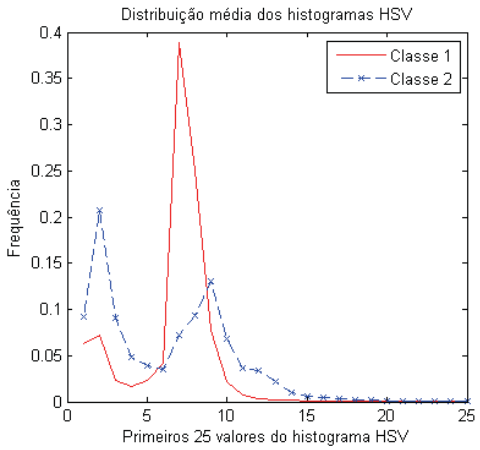


Figura 2. Histograma HSV médio para ambas as classes, c_1 e c_2

Este trabalho utiliza uma base de dados composta por 3 vídeos de aproximadamente meia hora cada um, com um total de 45.000 quadros, obtidos através de gravação de uma partida de futebol televisionada. A Fig. 1 mostra alguns quadros característicos da base.

Com o objetivo de estudar as características mais discriminantes entre as classes, os vídeos foram supervisionados e cada *frame* foi manualmente classificado.

Durante a classificação, percebeu-se que os *frames* que foram classificados como tomada principal (classe 1) continham alta concentração da cor verde, porém, em diferentes níveis de intensidade.

Tendo isso em vista, propomos o uso do histograma *HSV* como característica discriminante entre ambas as classes, uma vez que diversos tons de verde são representados em um intervalo contínuo e reduzido neste tipo de histograma.

Conforme o trabalho de [Bimbo, 1999], a representação de cores do sistema *HSV* baseia-se na percepção humana. Neste modelo, propomos a discretização das cores *HSV* da seguinte forma. O componente H é discretizado em 18 valores de mantissa , o componente S em 3 valores de saturação e, finalmente, o componente V em 3 valores de intensidade. Essa discretização gera um histograma de 162 posições possíveis. Para efeitos de comparação, foram calculados os histogramas médios das classes 1 e 2. Esses histogramas são calculados a partir da Eq. (7).

$$Hist_m = \frac{1}{|C_1|} \sum_{i \in C1} Hist(i) \qquad (7)$$

Onde C_1 é o conjunto de *frames* que correspondem às tomadas da câmera principal, $|C_1|$ é o número de elementos de C_1 e $Hist(i)$ é a função para o cálculo do histograma do *frame* i. Os histogramas médios são mostrados na Fig. 2.

A Fig. 2 mostra que as médias das classes são diferentes, indicando que a cor pode ser uma característica discriminante. Dessa forma, consideramos a média da classe 1 como a distribuição padrão e utilizamos a divergência de *Kullback-Leibler* extendida para medir o quanto o histograma de cada *frame* diferencia desse padrão.

Um experimento foi realizado comparando, através da divergência de *Kullback-Leibler* extendido, todos os *frames* da classe 1 e da classe 2 com a média da classe 1. A Fig. 3 mostra as divergências de cada *frame*. A linha tracejada separa a distribuição de duas formas: o lado esquerdo representa os *frames* que pertencem à classe 1 e, do lado direito, os *frames* que pertencem à classe 2. Dessa forma, nota-se que a maioria das distâncias do lado esquerdo são baixas (próximas ao padrão) e, as distâncias do lado direito, são altas (fora do padrão). A linha contínua representa o limiar que melhor separa ambas as classes.

Uma das contribuições desse trabalho é um método automático para o cálculo desse limiar.

5. Experimentos e Discussão

Para o modelo proposto nesse artigo, há dois parâmetros que devem ser ajustados com o objetivo de maximizar os resultados. O primeiro é o limiar t das distâncias entre as distribuições para separar as classes. O segundo está diretamente relacionado à função de distância (Divergência de *Kullback-Leibler* extendida) e é conhecido como parâmetro entrópico q.

Propomos então o uso da curva *ROC*. Dessa forma, variando os parâmetros, é possível observar os efeitos na qualidade. De acordo com [Fawcett, 2006], a curva

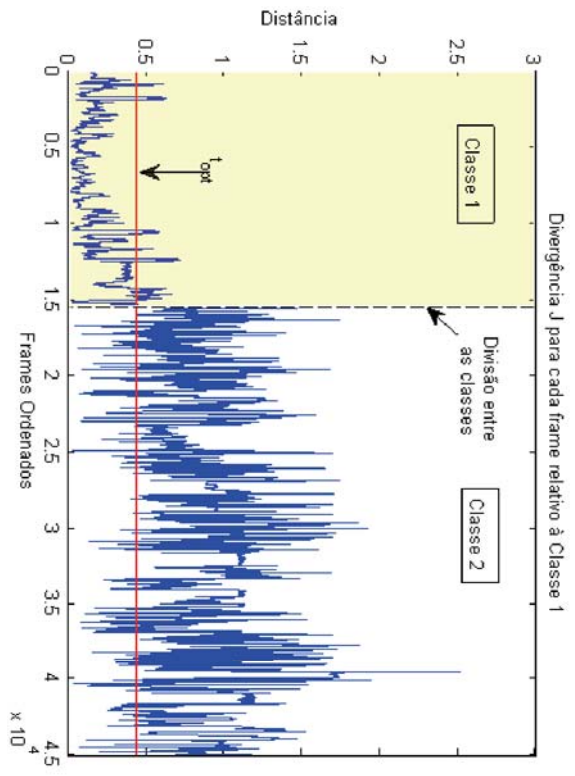


Figura 3. Distâncias entre os histogramas dos frames e o histograma médio da classe 1

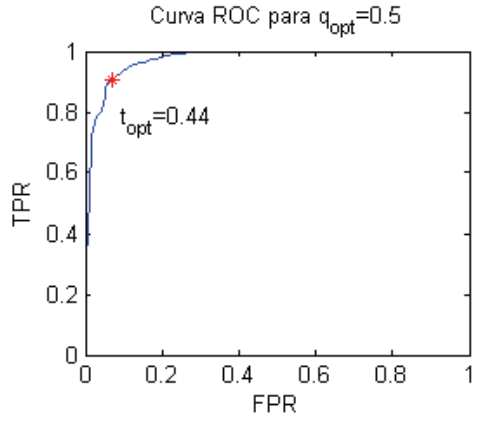


Figura 4. Curva ROC para q_{opt}

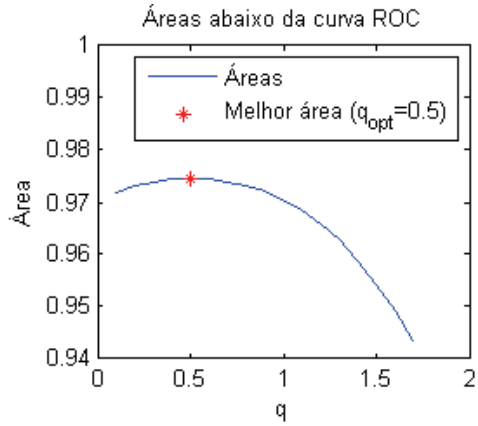


Figura 5. Áreas abaixo da curva ROC para diferentes valores de q

ROC é uma técnica que relaciona a quantidade de falsos positivos e verdadeiros positivos. Essa curva é gerada a partir de um parâmetro t que varia sobre uma distribuição, separando-a em dois grupos. Quanto mais separáveis forem as classes (menor sobreposição de elementos das classes), maior será a área sob a curva *ROC* (Az). Portanto, ajustando os parâmetros do classificador, podemos maximizar a área da curva *ROC* (Az).

Para este experimento, definimos que TP (*True Positive*) são os *frames* que foram supervisionados como pertencentes à classe 1 e corretamente classificados pelo sistema como pertencentes à esta classe e, FP (*False Positive*) são os *frames* que foram supervisionados como pertencentes à classe 2 e classificados de forma incorreta como pertencentes à classe 1.

Em relação ao tempo computacional, este experimento é $O(h \times w \times n)$, onde h é a altura do *frame*, w é a largura e n é quantidade de *frames* analisados. A divergência de *Kullback-Leibler* faz a comparação dos histogramas par-a-par em tempo $O(l)$, onde $l = 162$ é o número de entradas do histograma, ou seja, não é afetado pela quantidade de *frames*. A geração dos histogramas é feita em tempo $O(h \times w \times n)$, uma vez que é necessário a passagem por todos os *pixels* de todas as imagens. Como a comparação de todos os frames com a média da classe é da ordem $O(n)$, a complexidade do algoritmo é limitada superiormente por $O(h \times w \times n)$.

A Fig. 4 representa a melhor curva *ROC* gerada nos experimentos. A Fig. 5 mostra as áreas das curvas *ROC* para diferentes valores de q. Pode-se notar que a maior área, $Az = 0.974$, está relacionada ao $q = 0.5$.

Na Fig. 4 o ponto asterisco representa a melhor relação entre FPR e TPR, ou seja, a melhor classificação das classes. O valor de t é o parâmetro variante para obtenção dessa curva, variando de 0 ao máximo da distribuição. Com isso, o valor de t_{opt} (que melhor classifica os dois grupos) é $t_{opt} = 0.44$. Na Fig. 3, a linha contínua representa t_{opt}.

Os resultados indicam que o histograma *HSV* é

uma característica que pode classificar com precisão de até 97% dos *frames* (de acordo com a curva *ROC*). Isso significa que, para se alcançar 100% de acerto, deve-se eleger uma nova característica discriminante. Além disso, pode-se notar que a variação do valor do parâmetro entrópico q não alterou os resultados de forma significativa (menor que 4%), como pode ser observado na Fig. 5. Os mesmos testes foram efetuados para diferentes discretizações do histograma *HSV*. Os resultados ficaram próximos aos apresentados nesse artigo.

5.1. Conclusão

Neste artigo propomos um método para classificação da tomada de câmeras em jogos de futebol através da divergência de *Kullback-Leibler* extendida. Foram analisados uma hora e meia de vídeo em cores num total de 45.000 *frames* de três vídeos diferentes. Os tipos de tomadas que classificamos foram câmera 1 (câmera principal) e não câmera 1. A classificação como câmera 1 é fundamental para outras finalidades posteriores, tal como a detecção de placas de propagandas, demandada por emissoras de TV.

A metodologia proposta usa o histograma *HSV-162* para reduzir o espaço de busca com 18 valores de mantissa, 3 de Saturação e 3 de Intensidade. Os experimentos mostram que a maior predominância da mantissa é na faixa da cor verde,o que está de acordo com o valor de mantissa predominante nos *frames* que representam a câmera 1, e também é a característica predominante na classificação, levando a 97% da área máxima possível da curva *ROC*, o que representa uma perda pouco significativa em relação ao total. Esse resultado também reforça os dados apresentados em trabalhos prévios da literatura que não usam entropia não-extensiva, mas indicam também que a mantissa verde é a mais discriminante.

No uso da divergência de *Kulback-Leibler* extendida, o valor de q foi variado para uma faixa de 0 a 2, e mostrou influência pouco significante no cálculo automático do limiar de separação entre as classes, indicando que a escolha desse valor não é uma tarefa crítica do processo proposto, no entanto, verificando que a variação do valor ótimo de q encontra-se abaixo de 1.0, pela literatura da entropia não-extensiva sugere-se que o sistema estudado aqui pode ser não-extensivo. Uma consequência imediata dessa conclusão é que o sistema físico estudado pode então ser melhor avaliado caso sejam consideradas interações de longo alcance espaciais e temporais entre os seus estados. A câmera 1, uma vez encontrada, pode ser subdividida em subclasses, onde finalmente podem ser feitas análises mais precisas para detecção de objetos e pessoas em cenas.

As conclusões tomadas aqui podem ser estendidas, como trabalhos futuros para outros tipos de eventos televisivos que envolvem análise de vídeo, tais como: detecção de movimento e análise de cena para indexação, outros tipos de eventos como volei, basquete e corridas automobilísticas.

Agradecimentos

Os autores gostariam de agradecer ao CNPq, CAPES e FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, projeto Nº 2010/04917-8) - agências de financiamento científico, bem como à FEI (Fundação Educacional Inaciana), pelo suporte a este trabalho.

Referências

[Bimbo, 1999] Bimbo, A. D. (1999). *Visual Information Retrival*. Morgan Kaufmann.

[Borland et al., 1998] Borland, L., Plastino, A. R., and Tsallis, C. (1998). Information gain within nonextensive thermostatistics. *Journal Of Mathematical Physics*, 39:6490–6501.

[Esqueff, 2002] Esqueff, I. A. (2002). Tecnicas de entropia em processamento digital de images. Master's thesis.

[Fawcett, 2006] Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recogn. Lett.*, 27:861–874.

[Hsu et al., 2004] Hsu, W., fu Chang, S., wei Huang, C., Kennedy, L., yung Lin, C., and Iyengar, G. (2004). Discovery and fusion of salient multi-modal features towards news story segmentation. In *in IS&T/SPIE Electronic Imaging*.

[Jeffreys, 1939] Jeffreys, H. (1939). *Theory Of Probability*. Oxford University Press, second edition.

[Kuhmunch, 1997] Kuhmunch, C. (1997). On the detection and recognition of television commercials. In *Proceedings of the 1997 International Conference on Multimedia Computing and Systems*, pages 509–, Washington, DC, USA. IEEE Computer Society.

[Rodrigues and Giraldi, 2009] Rodrigues, P. S. and Giraldi, G. A. (2009). Computing the q-index for tsallis nonextensive image segmentation. *Graphics, Patterns and Images, SIBGRAPI Conference on*, 0:232–237.

[Tsallis, 1988] Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Physics*, 52.

[Tsallis, 1999] Tsallis, C. (1999). Nonextensive statistics: Theoretical, experimental and computational evidences and connections. *Brazilian Journal Of Physics*, 29(1).

[Tsallis, 2001] Tsallis, C. (Springer, Berlin, 2001). Nonextensive statistical mechanics and its applications. *Series Lecture Notes In Physics*. The Bibliography Of The Subjects Is Regularly Updated At Http://Tsallis.Cat.Cbpf.Br/Biblio.Htm.

[Watve and Sural, 2008] Watve, A. and Sural, S. (2008). Soccer video processing for the detection of advertisement billboards. *Pattern Recogn. Lett.*, 29:994–1006.

[Yeh et al., 2005] Yeh, J.-H., Chen, J.-C., Kuo, J.-H., and Wu, J.-L. (2005). Tv commercial detection in news program videos. In *ISCAS (5)*, pages 4594–4597.

Influência do Uso de Sensores Coloridos em Retinógrafos Digitais

Flávio Pascoal Vieira
Opto Eletrônica S.A.
Departamento de Pesquisa e Desenvolvimento
Rua Joaquim A. R. de Souza, 1071
CEP: 13563-330 São Carlos, SP, Brasil
flavio.vieira@opto.com.br

Evandro L. L. Rodrigues
Universidade de São Paulo
Escola de Engenharia de São Carlos
Av. Trabalhador São-carlense, 400
CEP: 13566-590 São Carlos, SP, Brasil
evandro@sc.usp.br

Abstract

The observation and imaging of retina is a procedure that allows the diagnosis of several pathologies not only related to the eye fundus, but also many different structures of the human eye. The increase of resolution of fundus camera is always a target in the development process of these equipments. This paper analyses the influence of a Bayer color image sensor in a digital fundus camera. The main explored point is the image quality degradation of angiography and red free retinography, which generates grayscale images, when captured with a color CCD. A simulation shows the sensor MTF degradation and the spurious response from the linear interpolation process, applied in a Bayer pattern sensor. Knowing the pixel size and the fundus camera magnification is possible to apply the results and calculate the minimum structure size that can be resolved.

1. Introdução

O aumento da longevidade está mudando a estrutura etária do Brasil e provocando o surgimento de preocupações de ordem social, em especial as relacionadas com a saúde da população, com destaque para a oftalmologia que vem recebendo frequentes investimentos do governo [1]. Segundo Yorston [2], no Brasil, Argentina e Chile, o glaucoma é considerado a grande causa de perda da visão, ao lado da Degeneração Macular Relacionada à Idade (DMRI) que se apresenta como principal retinopatia dentre os idosos.

A observação e registro de imagens da retina são de fundamental importância no diagnóstico de diversas doenças oculares, tais como retinopatia diabética, glaucoma, retinose pigmentar e DMRI. Além disso, a técnica de registro de imagens da retina permite que sejam vistos diretamente vasos sanguíneos (artérias e veias retinianas), possibilitando ao médico detectar algumas outras enfermidades, como por exemplo, toxoplasmose, diabetes, câncer e leucemia [3].

Com o aumento da demanda destes exames, destacam-se os equipamentos para observação e registro das imagens do fundo do olho, em especial os retinógrafos digitais, principal objeto de estudo neste trabalho.

Um desafio permanente entre os desenvolvedores desses equipamentos é o aumento da resolução das imagens captadas, que está relacionado principalmente à qualidade óptica, e o tipo de sensor utilizado.

No mercado encontram-se alguns retinógrafos com apenas um sensor de imagem, geralmente CCDs (*Charge-Coupled Device*) coloridos, com filtro de Bayer [1]. Neste tipo de sensor um filtro espectral é depositado, geralmente por processo de evaporação, de modo que apenas uma estreita faixa do espectro visível atingirá o elemento sensível [4]. Desta forma, para compor a imagem final, é necessário um processo matemático de interpolação [5].

O principal objetivo deste trabalho é avaliar a degradação inerente ao processo de interpolação de um sensor de imagem colorido utilizado na captura de uma cena monocromática. A avaliação pode ser estendida ao universo da retinografia visto que alguns equipamentos utilizam sensores coloridos nos exames de retinografia anerítra e angiografias, os quais resultam em imagens monocromáticas.

2. Funcionamento do Retinógrafo

Seguindo as técnicas básicas da fotografia é necessário iluminar a cena que se deseja fotografar e capturar a imagem de maneira conveniente. Para efeito de projeto, estudo e análise é comum dividir o sistema óptico de retinógrafos em dois subsistemas, o de iluminação e o de captação.

O sistema de iluminação como o nome sugere, tem a função de iluminar o fundo do olho do paciente de forma homogênea. Já o sistema de captação projeta no sensor a imagem proveniente da retina [1].

Como a energia refletida e espalhada pela retina é diminuta, visto que o olho é uma estrutura biológica com a função de absorver luz, quase a totalidade dos retinógrafos digitais utiliza sensores de imagem com tecnologia CCD. Estes conferem ao equipamento uma melhor relação sinal-ruído, contribuindo positivamente para a qualidade final da imagem.

A combinação entre o espectro de iluminação e filtros espectrais inseridos no sistema de captação define os tipos de exames que podem ser realizados.

2.1. Tipos de Exame

As imagens do fundo do olho podem ser divididas em dois grupos: as retinografias e as angiografias. Os principais tipos de retinografia são a colorida (Figura 1) e a anerítra, (Figura 2) também conhecida como *red free*. Já as angiografias envolvem a administração de contraste na corrente sanguínea do paciente para destaque da circulação. Os tipos mais comuns são com Fluoreceína (Figura 3) e com Indocianina Verde (Figura 4).

Os princípios de funcionamento destes exames são descritos com detalhes por Modugno [1] e Carvalho [6].

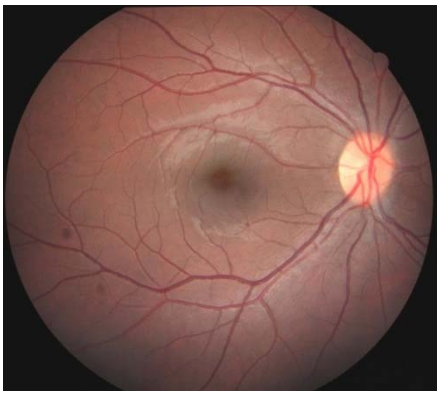


Figura 1 - Retinografia colorida

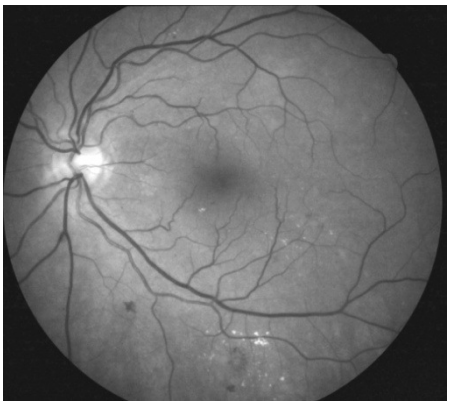


Figura 2 - Retinografia anerítra (*red free)*

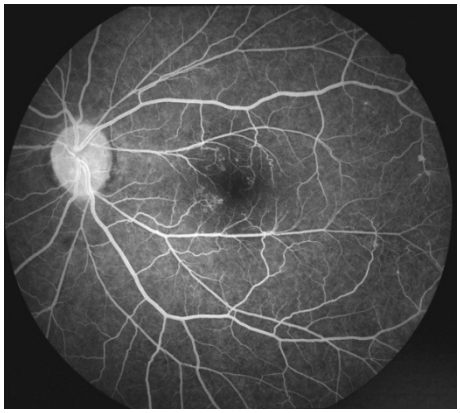


Figura 3 - Angiografia com fluoresceína (FA)

Figura 4 - angiografia com Indocianina verde (ICG)

2.2. Arquitetura do sistema de captação

A resolução da imagem final está ligada diretamente às funções de transferência de cada conjunto que compõe o sistema de captura. Desta forma pode-se representar a resolução em termos da função de transferência de modulação (MTF) [7]. Ou seja:

$$\mathrm{MTF}_{\mathrm{total}} = \mathrm{MTF}_{\mathrm{óptica}} \cdot \mathrm{MTF}_{\mathrm{sensor}} \cdot \mathrm{MTF}_{\mathrm{eletrônica}}$$

Explorando a equação anterior pode-se expressar a influência do sensor como:

$$\mathrm{MTF}_{\mathrm{sensor}} = \mathrm{MTF}_{\mathrm{geométrica}} \cdot \mathrm{MTF}_{\mathrm{difusão\ de\ carga}} \cdot \mathrm{MTF}_{\mathrm{CTE}}$$

A $\mathrm{MTF}_{\mathrm{difusão\ de\ cargas}}$ depende de parâmetros construtivos do componente, e a $\mathrm{MTF}_{\mathrm{CTE}}$ está relacionada à interferência entre pixels vizinhos. Ambos os valores, dependem da tecnologia de fabricação do componente.

A $\mathrm{MTF}_{\mathrm{geométrica}}$ está relacionada ao teorema da amostragem e ligada diretamente ao tamanho do pixel.

Esta provoca na imagem uma degradação fixa no sinal, expressa pela função *sinc*:

$$\mathrm{MTF}_{\text{geométrica}}(f_s) = \left|\frac{\text{sen}\,(\pi \cdot f_s \cdot \Delta S)}{(\pi \cdot f_s \cdot \Delta S)}\right| = |\text{sinc}(f_s \cdot \Delta S)|$$

onde:
f_s = frequência espacial
ΔS = tamanho do pixel

A frequência de Nyquist f_N, que limita a amostragem e em consequência a resolução espacial, é expressa por:

$$f_N = \frac{1}{2 \cdot \Delta S}$$

Esta análise é aplicável diretamente para sensores monocromáticos. No caso de sensores coloridos é necessária a consideração de outros elementos [8].

Um ponto notável dos quatro exames apresentados é que apenas um deles gera imagens coloridas. Alguns fabricantes de retinógrafos utilizam apenas sensores coloridos, outros optam por sistemas que utilizam dois tipos de sensores, um monocromático e outro colorido. Entretanto, o sensor de imagem compõe uma considerável parcela no custo do equipamento.

A análise deste trabalho é modelada tendo como base o imageamento de uma cena essencialmente monocromática, que é o caso dos três tipos de exames anteriormente citados. Busca-se neste caso avaliar a influência de um sensor colorido em comparação com um monocromático. Ambos os sistemas foram modelados conforme a Figura 5.

Considera-se que a óptica é perfeita e o tamanho do pixel é igual para os dois sistemas, visto que neste trabalho será avaliada apenas a influência da arquitetura dos sensores.

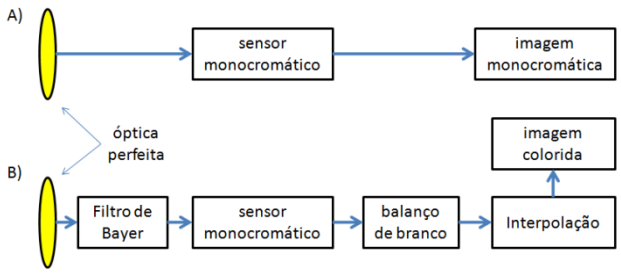


Figura 5 - Modelo para o sistema de aquisição de imagens (A) monocromático e (B) colorido.

O sistema B na Figura 5, além do filtro de Bayer, possui dois elementos adicionais [8]. O sistema de balanço de branco, necessário porque a atenuação do filtro é diferente para cada canal de cor, e o sistema de interpolação, que será abordado em detalhes.

3. Filtro de Bayer e interpolação

Considerando um filtro de Bayer clássico com configuração BGGR, mostrado esquematicamente na Figura 6, é possível observar que cada pixel é responsável por apenas um canal de cor [9]. Neste caso para a composição da imagem colorida final é necessário um processo de interpolação. Este processo em geral é implementado na câmera ou na placa de captura e fica transparente ao usuário [5].

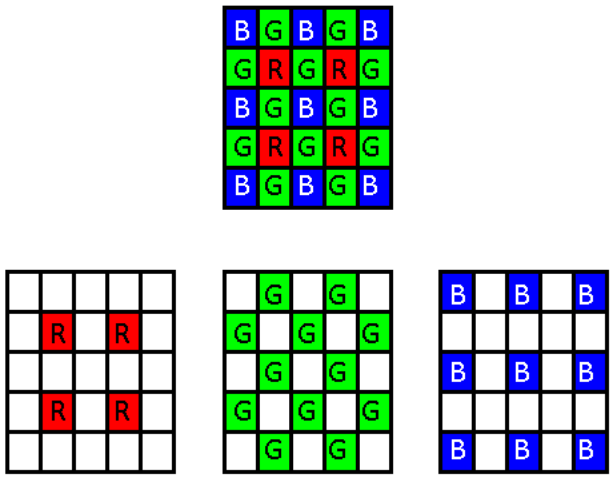


Figura 6 - Sensor com filtro de Bayer. Cada pixel recebe apenas um canal de cor.

Neste trabalho será considerado o processo de interpolação linear. Mesmo não sendo o algoritmo mais eficiente, ainda é amplamente utilizado devido a sua facilidade de implementação. Neste processo o valor de cada pixel após a interpolação é uma função linear de sua vizinhança na imagem original [8].

O algoritmo de interpolação linear é expresso por:

$$\mathrm{Vout}_c(x,y) = \sum_{x'}\sum_{y'} \mathrm{Vin}_c(x',y') \cdot f_c(x,y,x',y')$$

onde:
Vout_c é o resultado do valor interpolado para o canal c,
Vin_c é o valor do canal c na imagem original, e
f_c é a função peso de interpolação.

Para o valor de cada canal deve ser observado que:

$$\mathrm{Vin}_c(x,y) = \begin{cases} V(x,y) & \text{caso pixel}(x,y) \in \text{canal de cor c} \\ 0 & \text{caso contrário} \end{cases}$$

onde:
$V(x,y)$ é o valor amostrado do pixel de coordenadas (x,y).

Neste trabalho, a exemplo de outros algoritmos de interpolação linear [8], utilizaram-se como funções peso as f_c^k mostradas na Tabela 1 onde c representa o canal de cor da imagem interpolada (RGB) e k representa o plano do filtro de Bayer (BGGR).

Tabela 1 - Funções peso para interpolação linear

$$f_R^B = \begin{bmatrix} 0{,}25 & 0 & 0{,}25 \\ 0 & 0 & 0 \\ 0{,}25 & 0 & 0{,}25 \end{bmatrix} \quad f_R^G = \begin{bmatrix} 0 & 0 & 0 \\ 0{,}5 & 0 & 0{,}5 \\ 0 & 0 & 0 \end{bmatrix} \quad f_R^G = \begin{bmatrix} 0 & 0{,}5 & 0 \\ 0 & 0 & 0 \\ 0 & 0{,}5 & 0 \end{bmatrix} \quad f_R^R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$f_G^B = \begin{bmatrix} 0 & 0{,}25 & 0 \\ 0{,}25 & 0 & 0{,}25 \\ 0 & 0{,}25 & 0 \end{bmatrix} \quad f_G^G = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad f_G^G = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad f_G^R = \begin{bmatrix} 0 & 0{,}25 & 0 \\ 0{,}25 & 0 & 0{,}25 \\ 0 & 0{,}25 & 0 \end{bmatrix}$$

$$f_B^B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad f_B^G = \begin{bmatrix} 0 & 0{,}5 & 0 \\ 0 & 0 & 0 \\ 0 & 0{,}5 & 0 \end{bmatrix} \quad f_B^G = \begin{bmatrix} 0 & 0 & 0 \\ 0{,}5 & 0 & 0{,}5 \\ 0 & 0 & 0 \end{bmatrix} \quad f_B^R = \begin{bmatrix} 0{,}25 & 0 & 0{,}25 \\ 0 & 0 & 0 \\ 0{,}25 & 0 & 0{,}25 \end{bmatrix}$$

Com base nos resultados obtidos por Yotam [8], a degradação da resolução espacial de um sensor de imagem com filtro de Bayer é devida principalmente a dois fatores decorrentes do processo de interpolação: o primeiro é a atenuação do sinal luminoso que chega ao elemento sensível, e o segundo é decorrente do surgimento de freqüências espúrias no sistema.

Ainda seguindo os resultados de Yotam [8] tem-se na forma analítica:

$$MTF_{interpolação} = 0{,}65 + 0{,}35\cos(2\pi f)$$
$$Efeito\ espúrio = 0{,}1 - 0{,}1\cos(2\pi f)$$

Aplicando-se uma imagem da forma $\phi(f) = \sqrt{sinc\,(f)}$ como entrada do sistema e considerando:

$$\mathrm{MTF_{óptica}} = \mathrm{MTF_{eletrônica}} = \mathrm{MTF_{ideal}}$$
$$\mathrm{MTF_{ideal}} = \begin{cases} 1 & \text{para } f < f_N \\ 0 & \text{cc} \end{cases}$$

Pode-se traçar o gráfico apresentado na Figura 7 que mostra, em função da freqüência espacial, a função de transferência do sensor de imagem, para o sistema monocromático e colorido, além curva da resposta espúria do processo de interpolação.

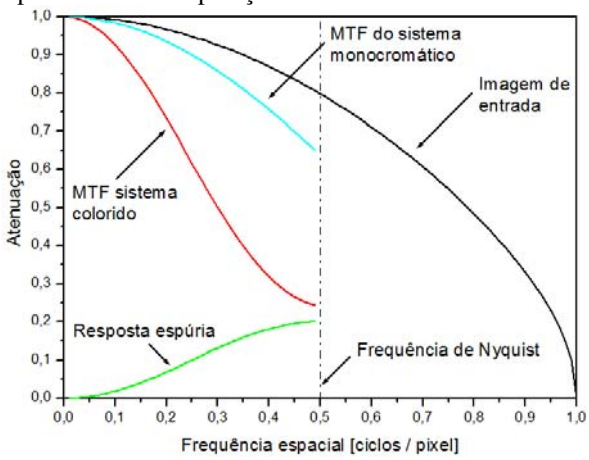


Figura 7 – Resultados analíticos da MTF para um sistema colorido com filtro de Bayer e processo de interpolação linear, e um sistema monocromático.

Para validar os resultados analíticos e gerar um padrão de comparação para o caso específico de um retinógrafo digital, do qual se tem todas as especificações necessárias, foi feita uma simulação numérica.

4. Resultados

A simulação utiliza um princípio comum na estimação da função de transferência, o algoritmo de *Knife Edge*, fundamentado na resposta ao degrau [1]. Nesta técnica, e com as considerações feitas acerca da idealidade da óptica e eletrônica, supõe-se no CCD uma imagem composta de apenas dois tons. De um lado a energia mínima (preto) e do outro a máximo (branco).

A transição entre estes dois tons representa matematicamente um degrau de excitação. Analisando-se, no domínio da frequência, por meio da transformada de Fourier, a resposta captada pelo sensor obtém-se a função de transferência do sistema.

A simulação consiste na aplicação das funções de interpolação linear apresentadas na borda de transição entre o campo claro e o escuro. Desta forma é obtido o valor para cada canal de cor.

Os valores de entrada, resultados da interpolação e valores finais estão dispostos na Figura 8, que representa uma parcela de 20 pixels do CCD colorido, na qual a transição ocorre entre a segunda e terceira coluna. A Figura 9 identifica os itens na representação de cada pixel.

B	0	G	0,5	B	0,5	G	1	B	1
	0		0		0,75		1		1
0	0	0	0	1	1	1	1	1	1
0		0,15		0,7		1		1	
G	0	R	0	G	1	R	1	G	1
	0		0,25		1		1		1
0	0	0	0,25	1	0,5	1	1	1	1
0		0,2		0,95		1		1	
B	0	G	0,5	B	0,5	G	1	B	1
	0		0		0,75		1		1
0	0	0	0	1	1	1	1	1	1
0		0,15		0,7		1		1	
G	0	R	0	G	1	R	1	G	1
	0		0,25		1		1		1
0	0	0	0,25	1	0,5	1	1	1	1
0		0,2		0,95		1		1	

Figura 8 – Representação simulada da imagem de um degrau atingindo um sensor com filtro de Bayer.

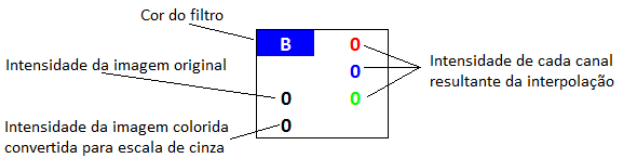


Figura 9 – Identificação dos valores para cada pixel apresentados na simulação.

Para a imagem capturada com o sensor colorido é necessária a conversão para a escala de cinza, já que a visualização e reprodução, assim como a cena capturada são monocromáticas. Esta conversão pode ser realizada por meio da equação de luminância recomendada pela CIE [10]:

$$V_{PB} = [0{,}2989 \quad 0{,}5870 \quad 0{,}1140] \cdot \begin{bmatrix} V_R \\ V_G \\ V_B \end{bmatrix}$$

Observando-se a Figura 10 (B), na segunda coluna os pixels que originalmente possuíam valor 0 (preto) passaram para os valores 0,15 e 0,2 (conforme Figura 8). De forma semelhante, na terceira coluna onde originalmente os valores eram 1 (branco), após o processo de interpolação os pixels assumiram valores 0,7 ou 0,95.

O resultado se traduz no surgimento de valores espúrios na proximidade da borda de transição, conforme representado na Figura 10. Em A tem-se representado o degrau de excitação, que devido às suposições de óptica e eletrônica perfeitas, é coincidente com a imagem obtida com o sensor monocromático. Já em B observa-se a imagem resultante do processo de interpolação, inerente ao sistema colorido.

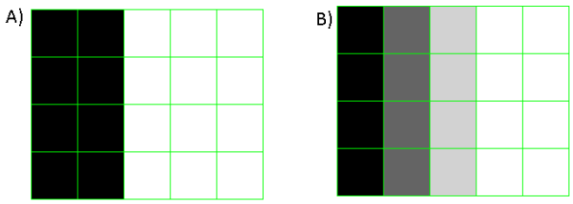


Figura 10 - Resultado do imageamento de uma cena monocromática com um sensor monocromático (A) e um sensor colorido (B)

Por meio da transformada de Fourier destas duas imagens resultantes, é possível determinar a função apresentada na Figura 11. Para tal foram consideradas especificações do tamanho do pixel igual a 9µm, o que implica no valor da freqüência de Nyquist próximo de 55 ciclos/mm.

Quanto maior a taxa de contraste, melhor será a nitidez da imagem captada, facilitando a identificação das estruturas da retina. Avaliando a função de transferência para um valor de 75% (apenas na MTF_{sensor}) pode-se calcular o tamanho do menor elemento que se consegue resolver. Com base no gráfico apresentado tem-se em 75% do contraste uma resolução de 10 ciclos/mm para o sistema colorido e 18 ciclos/mm no sistema monocromático.

Um retinógrafo comercial que possui uma ampliação de 1,85 pode-se obter que a menor estrutura possível de ser observada tem o tamanho de 54µm para o modelo colorido e 30µm para o modelo monocromático.

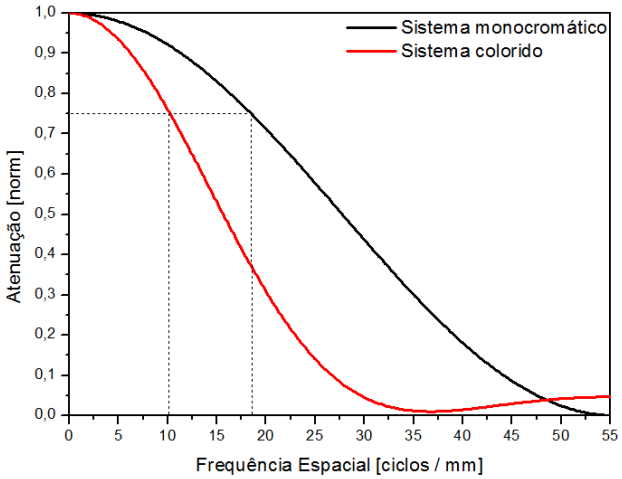


Figura 11 - Resultado da simulação para os dois sistemas propostos.

5. Conclusão

Com os resultados apresentados conclui-se que para cenas monocromáticas, que é o caso das angiografias e retinografia anerítra, o sistema com sensores coloridos apresenta um desempenho inferior quando comparado ao sistema monocromático. O efeito espúrio decorrente do processo de interpolação acarreta em uma diminuição significativa da resolução do sistema.

Com a eliminação do filtro de Bayer no sistema de captação de retinógrafos digitais, obtêm-se um aumento da qualidade da imagem da retina possibilitando ao médico um diagnóstico mais seguro e preciso. O uso de sensores monocromáticos para os exames citados é vantajoso, entretanto geram complicações no processo de aquisição de uma imagem de retinografia colorida.

Este trabalho apresentou uma série de estudos com os quais se pretende embasar uma proposta de modificação no sistema de iluminação de retinógrafos digitais, de modo a manter a resolução espacial na melhor condição, e ainda possibilitar a obtenção de imagens coloridas, foco de trabalho futuro.

Com a continuidade do trabalho de pesquisa, do qual os resultados aqui apresentados são parte integrante da fundamentação teórica, objetiva-se propor um sistema de iluminação cromático para retinógrafos digitais. Na proposta a obtenção de imagens coloridas, será realizada a partir de três capturas com um sensor monocromático. Espera-se que o desempenho do sistema, em termos da resolução espacial, seja superior quando comparado ao tradicional com sensores coloridos. Desta forma será possível o diagnóstico tanto baseado em informações radiométricas quanto espectrais.

Agradecimentos

Agradeço ao Dr Mário Stefani pelo suporte e incentivo à produção acadêmica nas dependências da Opto Eletrônica SA. À todos os colegas que contribuíram com este trabalho em especial ao físico Diego Lencione pelos inestimáveis ensinamentos.

Referências Bibliográficas

[1] – Modugno, R. G. *Uma Contribuição ao Projeto de Retinógrafos Digitais*. Dissertação (Mestrado em Engenharia Elétrica) – Departamento de Engenharia Elétrica, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2008.

[2] – Yorston, D. "Retinal Diseases and Vision 2020", *Community Eye Health Journal*. v. 16, no. 46, 2003, p. 19-20.

[3] – Oliveira, T. B. "Desenvolvimento e Resultados Preliminares de um Sistema Cromático de Iluminação para Oftalmoscópios Indiretos" – *Arq Bras Oftalmologia*; 2009, p. 146-151.

[4] – Martins, A. L., Rodrigues, E. L. L; "Sistema de Visão para Monitoramento Óptico Banda Larga, em Tempo Real, para Fabricação de Filtros com Filmes Finos Multicamadas". *VI Congresso de Visão Computacional*, VI Workshop de Visão Computacional, Presidente Prudente, SP., 2010, p. 168-174.

[5] – Rajeev R., "Demosaicking methods for Bayer color arrays", *Journal of Electronic Imaging*, North Carolina State University, 2002, p. 306-315

[6] – Carvalho, E. G. *Desenvolvimento de um Sistema Óptico para Retinografia e Angiografia Digital*. Dissertação (Mestrado em Física) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2006.

[7] – Scaduto, L. C. N., *Desenvolvimento e avaliação do desempenho de sistema óptico aplicado a sensoriamento remoto orbital.* Dissertação (Mestrado em Física Aplicada) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2008

[8] – Yotam, E., Ephi, P., and Ami, Y. "MTF for Bayer pattern color detector", *Proceedings SPIE 6567*, doi:10.1117/12.723140, 2007.

[9] – Bayer, B. E. "Color imaging array", *U.S. Patent No 3,971,065,* 1976.

[10] – CIE (1932). *Commission internationale de l'Eclairage proceedings*, Cambridge University Press, Cambridge, 1931.

Detecção de Expressões Faciais: Uma Abordagem Baseada em Análise do Fluxo Óptico

Leonardo Panta Leão
Departamento de Computação
Universidade Federal de Sergipe (UFS)
São Cristóvão – SE – Brazil
leonardopspl@gmail.com

Leonardo Matos
Departamento de Computação
Universidade Federal de Sergipe (UFS)
São Cristóvão – SE – Brazil
lnmatos@ufs.br

Maria Augusta Nunes
Departamento de Computação
Universidade Federal de Sergipe (UFS)
São Cristóvão – SE – Brazil
gutanunes@dcomp.ufs.br

Resumo

This article describes a real-time automatic facial expression recognition system using a technique for automatic facial expression detection based on Haartraning algorithm followed by automatic points placement that performs facial movements' reading using a technique for optical flow analysis on these points. The facial movements read are interpreted as action units, according to a classification presented by the researcher Paul Ekman. As a result, an automatic real-time facial expression detection system capable of selecting one among six expressions (anger, fear, happines, sadness, disgust and surprise) was generated.

1. Introdução

Há um interesse recente na melhoria da interação entre humanos e computadores. Para que uma efetiva interface humano-computador inteligente seja alcançada é necessário que o computador interaja naturalmente com o usuário, semelhante à maneira que os humanos interagem. Os humanos interagem principalmente através da fala, mas também utiliza gestos corporais para enfatizar uma parte do discurso ou para transmitir emoções. As emoções são demonstradas de forma visual, vocal, entre outras maneiras fisiológicas. Uma das maneiras mais importantes que os seres humanos demonstram emoções é através das expressões faciais. Com o objetivo de alcançarmos uma interface humano-computador inteligente mais eficiente fica evidente que o reconhecimento do estado emocional do humano pela face pode ser uma ferramenta valiosa.

Este trabalho descreve um sistema de detecção automática, em tempo real, que utiliza como entrada um vídeo ou as imagens capturadas através de uma *webcam*. Inicialmente ocorre a detecção da face do indivíduo na tomada de vídeo e a marcação de pontos, denominados pontos de leitura, em regiões de interesse (olhos e boca, por exemplo). E baseado nos movimentos faciais, ocorre a classificação da expressão facial apresentada pelo usuário.

2. Trabalhos Relacionados

Em um de seus trabalhos sobre as expressões faciais humanas, Ekman [4] demonstra evidências que suportam a universalidade das expressões faciais. Essas 'expressões faciais universais' são as que representam felicidade, tristeza, raiva, medo, surpresa e aversão [6]. Também foram estudadas as expressões em diversas culturas e tipos faciais (ocidentais, africanos, chineses, japoneses, por exemplo).

Ekman desenvolveu um sistema de codificação para as expressões faciais onde os movimentos faciais são descritos como um conjunto de *action units* (AU) [5]. Cada AU tem como base estudos relacionados à anatomia dos músculos faciais. Muitos pesquisadores se inspiraram nessas AU fazendo o uso de processamento de imagem e vídeo para automaticamente rastrear características faciais e então utilizá-las para categorizar as diferentes expressões. Pantic e Rothkrantz [13] fornecem um apanhado de abordagens diferentes utilizadas na tentativa de fazer o reconhecimento au-

tomático de expressões faciais. Essas abordagens têm algo similiar quanto ao fato de que todas elas utilizam o rastreamento de características faciais utilizando algum modelo de movimento da imagem (fluxo óptico, coeficientes DCT, entre outros). Baseado nessas características um classificador pode ser treinado. A principal diferença entre essas abordagens fica no conjunto de características extraídos das imagens de vídeo e do classificador utilizado.

O trabalho de Azcarate [1], assim como nosso trabalho, utiliza o detector de faces proposto por Viola e Jones [15] para a detecção inicial da localização da face. Nesse trabalho é construída na face uma máscara que é deformada com os movimentos faciais que são realizados quando o usuário muda a expressão facial. Essa máscara é baseada na máscara que utiliza deformação dos volumes de Bézier proposta por Tao e Huang [14]. Esse trabalho faz o uso de treinamento de classificadores baseado nos dados obtidos com a deformação dos volumes de Bézier. Os classificadores utilizados são o *Naive Bayes* e o *Tree-Augmented-Naive Bayes* (TAN).

Nosso trabalho não faz uso de treinamento de classificadores para a categorização das expressões faciais. Utilizamos as AU detectadas através da leitura do movimento de pontos colocados automaticamente na face do usuário utilizando inicialmente um detector de faces baseado no algoritmo Haartraining [15] e, em seguida, utilizando a técnica de fluxo óptico através do método Lucas-Kanade [12]. A categorização das expressões faciais ocorre com a aparição de conjuntos de AU, que caracterizam uma dentre as seis expressões faciais além da neutra, presentes simultaneamente num dado instante. A Tabela 1 mostra as AU utilizadas em nosso trabalho. Essas AU são listadas em [5], a primeira linha da Tabela 1, por exemplo, mostra o resultado de duas *action units* distintas que podem aparecer ao mesmo tempo (AU 1+2), a AU 1 é o levantamento da parte interna da sombrancelha enquanto que a AU 2 é o levantamento da parte externa da sobrancelha, resultando no levantamento arqueado das sobrancelhas.

AU	Descrição
1+2	levantamento arqueado das sobrancelhas
1+2+4+5	levantar e aproximar as sobrancelhas
4	abaixar as sobrancelhas
5	levantamento da parte superior dos olhos
6	apertar os olhos
12+25	boca no formato de sorriso
18	expremer os lábios
20	esticar a boca horizontalmente
26	queda do queixo
26b	queda menos acentuada do queixo

Tabela 1. *Action units* utilizadas no detector de expressões faciais implementado

3. Detecção da Face

O método de segmentação proposto por [15] e discutido em [10] é uma técnica de detecção baseada na aparência do objeto. Métodos desse tipo aprendem características a partir de conjuntos de imagens de treinamento que capturam a variedade da classe do objeto.

Como sugerido em [16] esse método pode ser usado para detectar faces humanas. Nesse método é feito uma etapa inicial de treinamento onde ocorre a seleção de dois conjuntos de imagens, um positivo e outro negativo. O conjunto de imagens positivas contém recortes onde está contido apenas o objeto a ser detectado. Já o conjunto de imagens negativas são imagens que não contêm o objeto, geralmente paisagens onde o objeto pode ser encontrado. Em seguida é feito um conjunto de exemplos onde as imagens positivas são combinadas com as negativas para formar imagens de provável aparição do objeto.

O processo de extração de características é baseado em máscaras convolucionais inspiradas em funções de Haar [8](Figura 1). Essas características são calculadas através da convolução na imagem do objeto baseadas numa decisão binária a partir de um *threshold*.

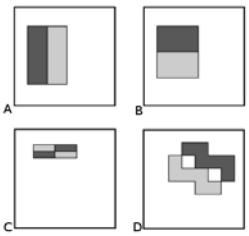


Figura 1. Exemplos de características retangulares [15]

Algumas características são mostradas na Figura 1, em (A) e (B), são características que capturam transições de nível de cinza verticais e horizontais respectivamente. Em (C) e (D) são mostradas máscaras que capturam características diagonais.

Na etapa de treinamento o método utiliza o AdaBoost [7] para construir o classificador. O objetivo desse algoritmo é construir um classificador eficiente a partir de uma série de classificadores fracos baseados em decisões binárias a partir da convolução dessas máscaras na imagem do objeto. O uso de características reduz a variedade de dados na classe e aumenta a variedade de dados fora da classe em comparação a dados de entrada crús, isto é, somente a informação con-

tida nos *pixels* da imagem. Além disso, as características geralmente codificam conhecimento sobre o domínio, o que é difícil aprender a partir de um conjunto crú e finito de dados. O conjunto de características usados pelo algoritmo são características como as descritas acima em todas as escalas e enquadramentos possíveis. Para se ter uma ideia, numa imagem de 64 por 64 *pixels* temos um conjunto de mais de 120 mil características possíveis.

Para o cômputo rápido dessas características é utilizado uma imagem chamada imagem integral, descrita em detalhes no trabalho de Lienhart e Maydt [10]. Essa imagem é construída utilizando o príncipio da programação dinâmica, onde cada pixel $(\imath, \jmath)$ da imagem integral guarda a soma dos valores de todos os *pixels* do canto superior esquerdo (a origem) até o pixel corrente, isto é, o pixel $(\imath, \jmath)$. Usando a imagem integral é possível realizar o cálculo de uma característica em tempo constante.

O AdaBoost seleciona as características que melhor classificam os objetos e chama um classificador fraco repetidamente numa série de turnos. Para cada chamada, uma distribuição de pesos é atualizada para indicar a importância de alguns exemplos no conjunto de dados para a classificação. A cada turno, os pesos de cada exemplo incorretamente classificado são incrementados, de maneira que o novo classificador tenha um maior foco nesses exemplos. Assim, após selecionar um classificador ótimo, baseado nessas características selecionadas e nessa distribuição de pesos, os exemplos que o classificador classifica incorretamente têm seus pesos aumentados e os exemplos classificados corretamente têm seus pesos diminuídos. Consequentemente, quando o algoritmo testa uma nova distribuição de pesos ele irá selecionar um classificador que melhor identifica esses exemplos que o classificador anterior errava. Cada classificador forte obtido a partir desses conjuntos de características é posto numa estrutura de cascata de rejeição (Figura 2) na ordem do menos complexo (menos características usadas) para o mais complexo (mais características usadas).

Na etapa de detecção utilizamos a arquitetura de cascata de rejeição. A Figura 2 ilustra a cascata de rejeição. A entrada é passada pelo primeiro classificador que decide entre verdadeiro ou falso (objeto encontrado ou não encontrado). Uma determinação de falso interrompe computação posterior e faz com que o detector retorne falso. Uma determinação verdadeira passa a entrada para o próximo classificador na cascata. Se todos os classificadores votarem em verdadeiro então a entrada é classificada como um exemplo verdadeiro. Dessa maneira, podemos economizar vários ciclos computacionais já que um dado de entrada, quando falso, requer a utilização de somente alguns nós da cascata.

A detecção ocorre numa janela deslizante desde o canto superior esquerdo da imagem até o canto inferior direito. A cada término de deslizamento na imagem a janela de detecção é aumentada 20% do seu tamanho inicial até que a janela enquadre ou exceda o tamanho da imagem. Caso mais de uma face seja encontrada praticamente na mesma região, já que poderá ser encontrada tanto na janela de detecção menor quanto na próxima aumentada, essas detecções são mescladas e consideradas uma só.

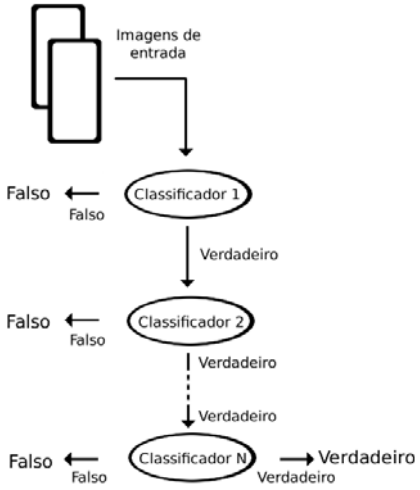


Figura 2. Modelo em cascata do algortimo proposto por Viola e Jones [15]

4. Colocação dos Pontos de Leitura

A colocação automática dos pontos de leitura, isto é, pontos de marcação de estruturas de controle como olhos, boca e nariz, ocorre logo após a etapa inicial de detecção automática da face e das regiões de interesse na face. De fato, ela só ocorre após a segmentação dessas regiões de interesse. A técnica utilizada para a segmentação dessas regiões é a mesma utilizada para a detecção da face utilizando a face detectada como região de interesse.

Para reduzir o custo computacional e a possibilidade de falhas utilizamos como região de interesse para a detecção da boca apenas a metade inferior da face detectada, já pra as regiões dos olhos esquerdo e direito utilizamos a metade de cada lado respectivamente. A região entre as sobrancelhas é encontrada imediatamente após o término da região do olho esquerdo e tem o mesmo comprimento e metade da altura. A região da testa é encontrada imediatamente acima da região entre as sobrancelhas e tem a mesma dimensão. A região abaixo da boca é encontrada logo após a colocação do ponto da parte inferior da boca e tem o mesmo compri-

mento da região do olho esquerdo e 25% da altura da mesma região. Os pontos são colocados nas seguintes regiões:

- Cantos esquerdo, direito, superior e inferior da boca totalizando quatro pontos sendo colocados na região de interesse correspondente à boca.
- Cantos superior e inferior do olho e um ponto acima da sobrancelha (tanto no olho esquerdo, como no direito) totalizando três pontos sendo colocados em cada uma dessas regiões de interesse.
- Região entre as sombrancelhas.
- Região da testa.
- Região abaixo da boca.

A Figura 3 mostra a tela inicial do programa após a etapa inicial de detecção e segmentação dessas regiões de interesse.

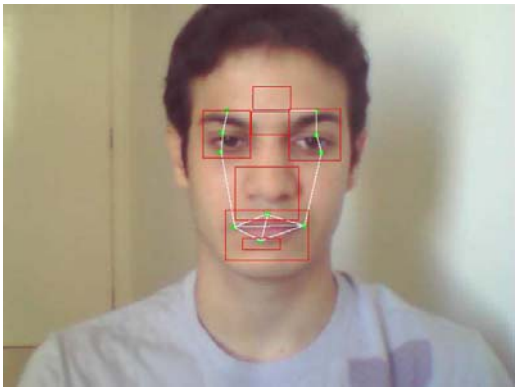


Figura 3. Regiões de interesse apontadas pelo sistema implementado na face inicialmente detectada

Na região da boca utilizamos o filtro de Sobel, que calcula a magnitude do gradiente da intensidade luminosa em cada *pixel* da imagem. Assim, obtemos uma noção de como varia espacialmente a luminância em cada ponto, se de modo mais suave ou abruptamente. Com isso, conseguimos estimar a presença de uma transição abrupta de nível de cinza. Filtros de Sobel direcionados permitem identificar a direção dessas variações (horizontal ou vertical) [8]. Como as variações intensas correspondem a fronteiras bem definidas entre objetos, conseguimos fazer a detecção de contornos .

Após a utilização do filtro configurado para detectar transições verticais (já que essa é a orientação perpendicular da boca) fazemos uma varredura da esquerda para a direita até que o primeiro canto seja encontrado para colocarmos o ponto esquerdo e uma varredura da direita para esquerda da mesma maneira para colocarmos o ponto direito após a binarização da imagem. A binarização acontece tornando branco *pixels* com tonalidade abaixo de um dado limiar e preto caso contrário. Assim, fica evidente as transições abruptas do nível de cinza. A mesma ideia é feita para encontrarmos o canto inferior e superior da boca utilizando o centro da boca como região inicial. A Figura 4 mostra a sequência da obtenção da posição inicial dos pontos na boca.

Figura 4. Na parte esquerda temos uma imagem em escala de tons de cinza que é utilizada para a convolução do filtro de Sobel, no centro temos a imagem após a convolução e binarização da imagem (para que fique evidenciado apenas as transições na horizontal) e na direita temos a colocação inicial dos pontos nos cantos encontrados

Para os olhos utilizamos o mesmo filtro e fazemos o mesmo tipo de varredura para colocarmos os pontos inferior e superior. Para o canto superior fazemos a varredura no sentido da esquerda para a direita e de cima para baixo partindo da metade da altura da região de interesse do olho, para o canto inferior fazemos o mesmo procedimento, porém partindo de baixo para cima. Para a sombrancelha fazemos a varredura de cima para baixo, partindo do meio até que a primera transição seja encontrada.

Nas regiões de interesse da área entre as sobrancelhas, da testa e da área abaixo da boca, colocamos o ponto no centro da região de interesse. Esses pontos são utilizados para que seja possível rastrear essas regiões constantemente a um baixo custo computacional. Na próxima seção é mostrada a técnica utilizada para a leitura contínua da posição atual desses pontos. A Figura 3 mostra os pontos colocados.

5. Fluxo Óptico nos Pontos

Para o cálculo contínuo da posição atual dos pontos, foi utilizado o algoritmo Lucas-Kanade (LK) [12] [2]. Esse método diferencial para o cálculo do fluxo óptico supõe que o fluxo é essencialmente constante numa vinzinhança local do *pixel* em consideração. Assim, as equações básicas de fluxo óptico para todos os *pixels* naquela vinzinhança são resolvidas utilizando o método dos mínimos quadrados [11].

O método LK supõe que o deslocamento do conteúdo da imagem entre dois instantes próximos é pequeno e aproximadamente constante com uma vizinhança do ponto p em consideração. Assim, podemos supor que a equação do fluxo óptico é mantida para todos os *pixels* com uma janela centrada em p. Isto é, o vetor (V_x, V_y) deve satisfazer

$$\begin{aligned} &I_x(q_1)V_x + I_y(q_1)V_y = -I_t(q_1) \\ &I_x(q_2)V_x + I_y(q_2)V_y = -I_t(q_2) \\ &\vdots \\ &I_x(q_n)V_x + I_y(q_n)V_y = -I_t(q_n) \end{aligned}$$

Onde $q_1, q_2, ..., q_n$ são *pixels* dentro da janela, e $I_x(q_i), I_y(q_i), I_t(q_i)$ são as derivadas parciais da imagem I em relação à posição x, y e o tempo t, calculados no ponto q_i e no tempo corrente.

Essas equações podem ser escritas na forma da matriz $Av = b$, onde

$$A = \begin{bmatrix} I_x(q_1) & I_y(q_1) \\ I_x(q_2) & I_y(q_2) \\ \vdots & \vdots \\ I_x(q_n) & I_y(q_n) \end{bmatrix}, v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}, \text{ e } b = \begin{bmatrix} -I_t(q_1) \\ -I_t(q_2) \\ \vdots \\ -I_t(q_n) \end{bmatrix}$$

Esse sistema tem mais equações que incógnitas, assim ele geralmente é indeterminado. O método LK usa uma solução ajustada utilizando o princípio dos mínimos quadrados. Isto é, ele resolve o sistema 2×2

$A^T Av = A^T b$ ou

$v = \frac{A^T b}{A^T A}$

Onde A^T é a matriz transposta, ou seja, ela computa

$$\begin{bmatrix} V_x \\ Vy \end{bmatrix} = \begin{bmatrix} \sum_i I_x(q_i)^2 & \sum_i I_x(q_i)I_y(q_i) \\ \sum_i I_x(q_i)I_y(q_i) & \sum_i I_x(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_x(q_i)I_t(q_i) \\ -\sum_i I_y(q_i)I_t(q_i) \end{bmatrix}$$

A matriz $A^T A$ é chamada de tensor de estrutura da imagem no ponto p.

5.1. Janela ponderada

A solução plana dos mínimos quadrados acima dá a mesma importância para todos os n *pixels* q_i na janela. Na prática é geralmente melhor dar mais peso para os *pixels* que são próximos do *pixel* central p. Para isso, utilizamos uma versão ponderada da equação dos mínimos quadrados,

$A^T WAv = A^T Wb$
ou
$v = \frac{A^T Wb}{A^T WA}$

onde W é uma matriz diagonal $n \times n$ contendo os pesos $W_{ii} = w_i$ a serem assinalados na equação do pixel q_i. Isto é, computamos

$$\begin{bmatrix} V_x \\ Vy \end{bmatrix} = \begin{bmatrix} \sum_i w_i I_x(q_i)^2 & \sum_i w_i I_x(q_i)I_y(q_i) \\ \sum_i w_i I_x(q_i)I_y(q_i) & \sum_i w_i I_x(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i w_i I_x(q_i)I_t(q_i) \\ -\sum_i w_i I_y(q_i)I_t(q_i) \end{bmatrix}$$

O peso w_i é geralmente ajustado como sendo uma função Gaussiana da distância entre q_i e p.

5.2. Método LK piramidal

Para o cálculo do fluxo óptico dos pontos na imagem foi utilizada a versão "piramidal" do método LK, como descrito em [3]. A técnica consiste em criar uma pirâmide Gaussiana da imagem onde no topo da pirâmide temos a menor versão da imagem (com menos detalhes) até chegarmos à base onde atingimos os *pixels* crús da imagem (maior riqueza de detalhes). O fluxo óptico é calculado para os pontos no topo da pirâmide e os resultados do cálculo desse nível servem como ponto de partida para os cálculos do nível seguinte. Utilizar o método LK dessa forma nos permite que movimentos maiores sejam capturados por janelas locais, o que torna o método mais eficiente e robusto, pois janelas maiores ferem a suposição inicial do método de que o *pixel* rastreado e seus vizinhos pertecem à mesma superfície fechada [3]. O método é ilustrado na Figura 5.

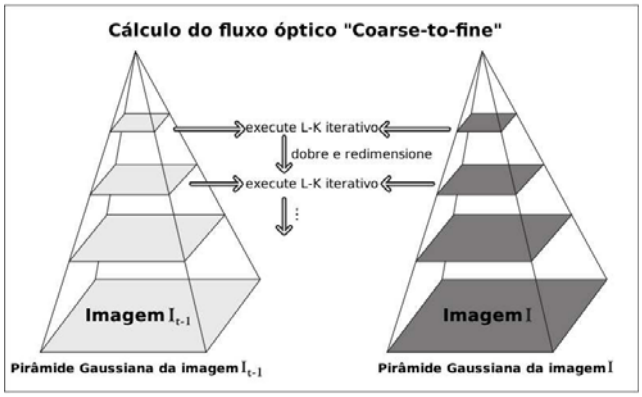


Figura 5. Fluxo óptico piramidal de Lucas-Kanade [3]

6. Implementação

Para a implementação do sistema proposto utilizamos a linguagem C++ com a biblioteca livre de visão computacional OpenCV [9]. Essa biblioteca possui diversos algoritmos e rotinas de processamento de imagem e visão computacional. Em nosso trabalho fizemos o uso de alguns desses algoritmos, assim como a captura de vídeo através da *webcam* que também é suportada por essa biblioteca.

6.1. Visualização

Para a utilização da implementação basta que o usuário execute a aplicação. A captura das imagens através da *webcam* ou do vídeo é iniciada e nessa etapa o algortimo detector automático de expressões faciais, descrito na Seção 3, é utilizado até que a face do usuário seja detectada. Após isso, o mesmo algoritmo é executado na região da face encontrada para segmentar as regiões dos olhos esquerdo e direito, boca e nariz, como descrito na Seção 4. No término desta etapa esse algoritmo deixa de ser executado. Na próxima etapa o algoritmo de fluxo óptico, descrito na Seção 5, entra em execução para realizar a leitura contínua da movimentação dos pontos colocados nas diferentes regiões da face.

As mudanças na movimentação dos pontos em relação à distância inicial entre esses pontos na face inicial, que é interpretada como neutra, caracteriza a presença de algumas AU, além da presença ou ausência de rugas, que são detectadas utilizando filtro de Sobel, nas regiões entre as sombrancelhas, testa e abaixo da boca. A combinação de algumas AU representam a aparição de uma emoção na face. A Figura 6 mostra o sistema sendo utilizado.

Figura 6. Exemplos da aplicação em execução

7. Experimentos

A etapa de experimentos consistiu na coleta de dados a partir do uso da aplicação por três usuários distintos. Nenhum deles é autor ou tinha conhecimento prévio do sistema. As expressões foram feitas de forma não espontânea, cinco vezes cada expressão facial para cada um dos três usuários e de forma consecutiva iniciando com a expressão facial neutra. A Tabela 2 mostra a matriz de confusão para as seis expressões faciais. Podemos ver na tabela a taxa de acerto (expressão facial detectada corretamente) em negrito em cada uma das linhas onde coincidem as expressões intensionadas pelos usuários com as colunas correspondentes às expressões detectadas pelo sistema.

	F	M	S	T	R	A
F	**73,3%**	6,7%	0%	0%	0%	20%
M	0%	**80%**	6,7%	0%	0%	13%
S	0%	40%	**60%**	0%	0%	0%
T	0%	0%	0%	**67%**	0%	33%
R	6,7%	0%	0%	0%	**53,3%**	40%
A	0%	20%	0%	0%	0%	**80%**

Tabela 2. Matriz de confusão para as seis expressões faciais: felicidade (F), medo (M), surpresa (S), tristeza (T), raiva (R) e aversão (A). As linhas representam a expressão intensionada pelo usuário, já as colunas representam a expressão detectada pelo sistema. A precisão média foi de 68,89%

A taxa de atualização dos quadros de vídeo após a detecção da face foi próxima de 30 quadros por segundo. O computador utilizado nos experimentos é constituído de um processador de 2 núcleos de 2.2 GHz e 2 GB de memória RAM. A *webcam* utilizada tem 2 MB *pixels* de resolução.

8. Conclusão

Apresentamos neste trabalho mais uma abordagem para a problemática da detecção automática de expressões faciais baseada em algoritmos conhecidos de processamento de imagem e visão computacional. Essa abordagem tem como principal vantagem a não intervenção do usuário na utilização da aplicação. Nos experimentos foram obtidos uma taxa de atualização dos quadros de vídeo compatível com uma aplicação de tempo real. Porém, existem ainda algumas limitações que devem ser levadas em consideração nessa abordagem. As expressões faciais são detectadas apenas com o indivíduo de frente para a câmera e sem objetos ou características muito distintas na face (óculos escuros, cabelo cobrindo os olhos, por exemplo), já que a técnica

detectora de faces tem sua cascata de rejeição treinada para detectar apenas a face em posição frontal (faces de perfil ou parcialmente oclusas não são detectadas) sem peculiaridades na aparência da face. A luminosidade do ambiente deve ser constante para não ferir a suposição de constância do brilho da técnica proposta por Lucas-Kanade [3]. Para avaliar melhor a eficiência da detecção das expressões faciais, pretendemos realizar teste com uma base de vídeos extensa e com uma quantidade maior de usuários.

Referências

[1] A. Azcarate, F. Hageloh, K. V. D. Sande, and R. Valenti. **Automatic facial emotion recognition**. *Universiteit van Amsterdam*, 2005.

[2] J.-Y. Bouguet. **Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm**. *Intel Corporation Microprocessor Research Labs*, 2000.

[3] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 1 edition, September 2008.

[4] P. Ekman. *Strong evidence for universals in facial expressions*. Psychol. Bull., 1994.

[5] P. Ekman and W. V. Friesen. *Facial Action Coding System (FACS): Manual*. Consulting Psychologists Press, 1978.

[6] P. Ekman and W. V. Friesen. *Emotion in the Human Face*. Cambrige Univ. Press, 1982.

[7] Y. Freund and R. Schapire. **A Short Introduction to Boosting**. *Journal of Japanese Society for Artificial Intelligence*, pages 1401–1406, 1999.

[8] R. C. Gonzalez and R. E. Woods. *Digital Image Processing 2nd Edition*. Prentice Hall, 2 edition, 2002.

[9] C. Intel. **Open Source Computer Vision Library Reference Manual**. *Intel Corporation Microprocessor Research Labs*, 2001.

[10] R. Lienhart and J. Maydt. **An Extended Set of Haar-like Features for Rapid Object Detection**. *ICIP*, pages 900–903, 2002.

[11] B. D. Lucas. *Generalized Image Matching by the Method of Differences*. PhD thesis, Robotics Institute, Carnegie Mellon University, July 1984.

[12] B. D. Lucas and T. Kanade. **An iterative image registration technique with an application to stereo vision**. *Journal of Japanese Society for Artificial Intelligence*, pages 121–130, 1981.

[13] M. Pantic and L. Rothkrantz. **Automatic Analysis of Facial Expressions: The State of the Art**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 2000.

[14] H. Tao and T. Huang. **Connected vibrations: a modal analysis approach to non-rigid motion tracking.** *Proceedings in IEEE Conference on CVPR*, 1:735–740, 1998.

[15] P. Viola and M. Jones. **Rapid Object Detection Using a Boosted Cascade of Simple Features**. *Conference on Computer Vision and Pattern Recognition*, 2001.

[16] P. Viola, M. Jones, and D. Snow. **Detecting Pedestrians Using Patterns of Motion and Appearance**. *Proceedings of the Ninth IEEE International Conference on Computer Vision*, 2003.

Detecção Automática da Área em Imagens de Impressão Digital

Marcos William da Silva Oliveira*
oliveira.mw@sjrp.unesp.br

Maurílio Boaventura*
maurilio@ibilce.unesp.br

Inês Aparecida Gasparotto Boaventura*
ines@ibilce.unesp.br

*Universidade Estadual Paulista
São José do Rio Preto - SP, Brasil

Abstract

In this paper we proposed a novel automatic method for area detection in fingerprint images, which is based on active contours. The segmentation is performed with the application of a model of active contours without edges, a morphological operation is used to filter the results. In latents images, we proposed a semi-automatic extension of this procedure in which a manual input of a simple initial contour is requested.

1. Introdução

O interesse da humanidade pelas impressões digitais data da pré-história, como indicam desenhos arqueológicos fazendo alusão a seus padrões e placas de cerâmica antigas com impressões gravadas no sentido de selar acordos ou assinar autorias. Porém, o primeiro sistema de classificação de impressões digitais é do fim do século XIX, realizado pelo francês Francis Galton, que motivou Juan Vucetich, a criar um sistema de identificação para a polícia de La Plata, na Argentina [2]. O Brasil foi o primeiro país a adotar oficialmente tal processo de identificação datiloscópica em 1903 [9], mesma época em que se noticiava em Nova York a solução de um crime na Índia a partir das digitais do suspeito [8].

Com o desenvolvimento dos computadores, as instituições policiais *Federal Bureau of Investigation* (FBI), nos Estados Unidos, *Home Office*, no Reino Unido, e *Paris Police Department*, na França, investiram no início da década de 1960 no desenvolvimento de um sistema automatizado de identificação a partir de impressões digitais, originando o *"Automatic Fingerprint Identification System"* (AFIS) [10]. Na atualidade, esses sistemas de identificação não se resumem às aplicações criminais e são alternativas aos meios comerciais tradicionais de identificação que se mostram falhos, como senhas ou cartões magnéticos.

Com essa motivação, este trabalho tem as imagens de impressões digitais obtidas em ambiente controlado como objeto de pesquisa. O objetivo traçado neste trabalho é propor uma técnica automática para detecção da área da impressão digital em imagens roladas ou planas, baseada no modelo de contornos ativos. Para tanto, realiza-se uma segmentação através do modelo sem detectores de bordas de Chan e Vese [5] e efetua-se uma filtragem deste resultado através da operação morfológica abertura. Sugere-se também uma extensão do processo para latentes, originárias de ambiente não controlado, na qual é exigida uma predefinição manual de um contorno inicial retangular simples.

Outros trabalhos publicados recentemente para detecção da área de impressões digitais utilizando *snakes* podem ser encontrados em [13] e [4], porém as abordagens utilizadas são diferentes das aqui propostas. Além disso, esses artigos não fazem menção a aplicações a latentes.

Os testes foram realizados sobre o banco de dados NIST SD27 [1], em cerca de 50 imagens, e alguns dos resultados são apresentados na seção apropriada.

2. Biometria de Impressões Digitais

A mais evidente característica estrutural de uma digital é um padrão intercalado de cristas e vales dos sulcos papilares [10]. Chamam-se *cristas* aos picos dos sulcos, que formam as linhas escuras na impressão, e *vales* às depressões, que ocasionam as linhas claras.

O registro da digital em um papel ou imagem digitalizada é chamado *impressão digital* e, por ser feito de várias formas, gera imagens de várias qualidades. As impressões digitais podem ser classificadas pela maneira como foram registradas ou situação que foram encontradas. Podem ser:

- *Rolada*: quando a digital é rolada sobre um papel, depois de embebida em tinta, ou sobre a superfície de captação do sensor óptico;
- *Plana*: quando a digital é apenas pressionada sobre um papel, depois de embebida em tinta, ou sobre a superfície de captação do sensor óptico;
- *Latente*: é o fragmento de impressão digital deixado por descuido do indivíduo em um local de crime. Esse fragmento é revelado através de processos químicos e registrado, por exemplo, por fotografia.

Como é intuitivo, uma impressão digital rolada contém maior área do dedo que uma plana, mas está sujeita a algumas distorções por causa do movimento realizado pelo indivíduo. Além disso, uma imagem de impressão digital plana ou rolada geralmente tem melhor qualidade que latentes, como pode-se observar na Figura 1.

Figura 1. Impressões digitais rolada, plana e latente do mesmo dedo (n°38 no NIST SD27 [1]).

2.1. Segmentação de Impressão Digital

Maltoni et al. [10] destacam que o termo *segmentação* denota, geralmente, a separação da área da impressão digital do fundo da imagem, ou seja, destacar a região exata da imagem onde ocorre a impressão digital. Essa separação é útil para evitar a extração de características em áreas ruidosas que aparecem principalmente no fundo da imagem.

A máscara da área da impressão digital pode ser usada em processos como extração do campo de orientação, determinação de minúcias, detecção de pontos singulares, dentre outros.

A vantagem em utilizar um procedimento automático para detecção da área da impressão digital é obter precisamente a região da imagem que deseja-se processar. No caso deste trabalho, a abordagem baseada em contornos ativos propicia a independência de pré-processamento para obter a segmentação de impressão digital e, portanto, maior aplicabilidade e desempenho na extração de características.

Nas duas seção seguintes são discutidos o conceito de morfologia matemática e o modelo de Chan-Vese proposto para segmentação de objetos em imagens digitais. Ambos os tópicos são essenciais para o trabalho desenvolvido e descrito nas demais seções.

3. Morfologia Matemática

O termo *morfologia* é associado ao estudo de formas e estruturas de animais e plantas, na biologia, de palavras, na gramática, e da vida social, na sociologia. *Morfologia matemática* reúne técnicas de extração de elementos da imagem que permitem a descrição das formas de uma região ou técnicas que alteram as formas presentes na imagem. Essas ferramentas extraem fronteiras, esqueletos e o fecho convexo e também podem realizar pré ou pós-processamentos de filtragem, afinamento e poda (*pruning*) [6]. Nesta seção, há breves comentários acerca de algumas operações binárias, utilizadas durante o trabalho.

A *dilatação* faz com que um *pixel* receba valor 1 se algum *pixel* em sua vizinhança também possuir 1; caso contrário, o valor do *pixel* é nulo. Em outra operação, a *erosão*, o valor do *pixel* resultante é 1 se todos os *pixels* em sua vizinhança possuem 1 e nulo caso contrário. A *abertura* é a dilatação da imagem erodida, considerando a mesma vizinhança. Essa vizinhança é definida pelo chamado *elemento estruturante*, que possui formas e tamanhos diversos.

Como os nomes sugerem, a dilatação amplia a quantidade de *pixels* com valor 1, enquanto a erosão os diminui. Quando aplica-se a abertura, os contornos são suavizados e as estruturas finas desconexas de regiões homogêneas eliminadas.

4. Modelo Chan-Vese

A segmentação de objetos em imagens digitais, em geral, baseia-se na detecção de regiões da imagem que são homogêneas em algum sentido. Essa homogeneidade pode ser quanto às diferenças de tonalidade de cinza ou de texturas em cada região, por exemplo. Pode-se afirmar que essas técnicas são capazes de encontrar automaticamente objetos na imagem.

Nesse sentido, o modelo de *contornos ativos*, primeiramente proposto por Kass et al. [7], é uma técnica de segmentação que objetiva detectar características como linhas e bordas de objetos em uma imagem u_0 a partir da evolução de uma curva inicial, definida no domínio de u_0, que move-se sujeita à imagem na direção normal de sua fronteira até parar sobre a borda do objeto. Uma das formulações desse modelo, denominada *level set* [3], é quando a curva é representada como nível zero de uma função de dimensão maior e a função velocidade varia, assumindo valores positivos e negativos. Devido ao movimento de curvas, modelos desse tipo também são conhecidos como *snakes* (cobras, em inglês).

Em geral, para determinar a convergência, esses métodos utilizam detectores de bordas baseados no gradiente da imagem. Porém, o objetivo deste trabalho é obter o contorno subjetivo de uma impressão digital, ou seja, um contorno não detectado pelo gradiente. Para tanto, propõe-se aplicar o modelo de *snakes* sem detectores de bordas apresentado por Chan e Vese [5].

Dessa forma, considere um conjunto aberto e limitado $\Omega \subset \mathbb{R}^2$, sendo $\partial\Omega$ sua fronteira e $\overline{\Omega}$ seu fecho. Sejam $u_0 : \overline{\Omega} \to \mathbb{R}$ uma dada imagem e $C(s) : [0,1] \to \overline{\Omega}$ uma curva parametrizada.

A *snake* proposta por Chan e Vese [5] é a minimização de uma energia baseada na segmentação. Considerando a curva $C \subset \Omega$ como a fronteira de um conjunto aberto $\omega \subset \Omega$, o método minimiza as energias no interior de ω e no exterior de ω, isto é, em $\Omega \setminus \overline{\omega}$. Ou seja, a ideia é obter C tal que

$$\inf_{c_1,c_2,C} F(c_1,c_2,C), \tag{1}$$

sendo o funcional F definido por

$$\begin{aligned} F(c_1,c_2,C) &= \mu.\text{Comprimento}(C) + \nu.\text{Área}(\omega) \\ &+ \lambda_1 \int_{\omega} |u_0(x,y) - c_1|^2 dxdy \\ &+ \lambda_2 \int_{\Omega\setminus\overline{\omega}} |u_0(x,y) - c_2|^2 dxdy, \end{aligned} \tag{2}$$

onde $\mu \geqslant 0$, $\nu \geqslant 0$, $\lambda_1, \lambda_2 > 0$ são parâmetros fixados e

$$\begin{cases} c_1 = & \text{média}(u_0(x,y)), \quad (x,y) \in \omega \\ c_2 = & \text{média}(u_0(x,y)), \quad (x,y) \in \Omega \setminus \overline{\omega}. \end{cases}$$

Escrito como na Equação (2), o funcional F pode ser associado ao *funcional Mumford-Shah* [11], considerando $\lambda_1 = \lambda_2 = \lambda$ e $\nu = 0$, e o modelo de contorno ativo torna-se um caso particular de *problema da partição mínima* [5].

Outra formulação do modelo, como método *level set*, é representar $C \subset \Omega$ pela função Lipschitziana $\phi : \Omega \to \mathbb{R}$, tal que

$$\begin{cases} C = \partial\omega = \{(x,y) \in \Omega | \phi(x,y) = 0\}, \\ \omega = \{(x,y) \in \Omega | \phi(x,y) > 0\} \text{ e} \\ \Omega \setminus \overline{\omega} = \{(x,y) \in \Omega | \phi(x,y) < 0\}. \end{cases}$$

Dessa maneira, usando a função Heaviside H e a medida de Dirac δ_0, definidos respectivamente por

$$H(z) = \begin{cases} 1, \text{ se } z \geqslant 0 \\ 0, \text{ se } z < 0, \end{cases} \quad \text{e} \quad \delta_0(z) = \frac{d}{dz} H(z),$$

para reescrever o funcional F em (2) com respeito a ϕ, obtém-se que, para $\epsilon \to 0$, o problema de minimização (1) é solucionado resolvendo a equação de evolução

$$\begin{cases} \dfrac{\partial\phi}{\partial t} = \delta_\epsilon(\phi)[\mu \text{div}\big(\dfrac{\nabla\phi}{|\nabla\phi|}\big) - \nu - \lambda_1(u_0 - c_1(\phi))^2 \\ \qquad +\lambda_2(u_0 - c_2(\phi))^2] = 0, (t,x,y) \in (0,\infty) \times \Omega, \\ \phi(0,x,y) = \phi_0(x,y), \quad (x,y) \in \Omega, \\ \dfrac{\delta_\epsilon(\phi)}{|\nabla\phi|} \dfrac{\partial\phi}{\partial\overrightarrow{n}} = 0, \quad (x,y) \in \partial\Omega, \end{cases} \tag{3}$$

onde δ_ϵ é o delta de Dirac, $\mu, \nu \geqslant 0$ e $\lambda_1, \lambda_2 > 0$ são parâmetros, $\overrightarrow{n}$ é a direção normal externa à fronteira $\partial\Omega$ e c_1 e c_2 são dados por

$$\begin{cases} c_1(\phi) = & \text{média}(u_0(x,y)), \quad \phi(x,y) \geqslant 0, \\ c_2(\phi) = & \text{média}(u_0(x,y)), \quad \phi(x,y) < 0. \end{cases}$$

5. Método Proposto

A detecção automática da área da impressão digital de forma precisa é importante no pré-processamento de diversos modelos de extração de características, principalmente por eliminar ruídos específicos do fundo desta imagem.

A ideia é considerar a impressão digital como um objeto na imagem e, dado um contorno inicial, aplicar o método de contornos ativos Chan-Vese (3) e realizar uma filtragem para obter o contorno subjetivo da impressão digital e, consequentemente, sua área precisa na imagem.

O contorno inicial é uma máscara binária que pode ou não envolver toda a área da impressão digital. A proposta é determiná-lo automaticamente quando o foco são imagens de boa qualidade, como as roladas e planas, tomando-se para isso um ponto qualquer no interior da impressão digital, geralmente o centro da imagem, e comparando-se a variância bloco a bloco da imagem. Como geralmente as impressões digitais são capturadas em ambiente controlado, supor que o ponto central da imagem esteja contido na impressão digital não parece uma imposição muito restritiva. No caso de latentes, o contorno inicial deve ser pré-definido manualmente, devido à pouca qualidade das imagens.

A filtragem é realizada através da operação morfológica binária abertura, aplicada sobre o resultado do modelo *snakes*. Isso é necessário para corrigir incorreções dos contornos que foram atraídos pelos ruídos específicos, como escritas ou desenhos manuais e caracteres impressos, presentes no fundo da imagem.

O parâmetro μ que aparece do modelo Chan-Vese (3) foi definido experimentalmente com valores altos, em torno de 2×10^4, pois os testes realizados mostraram resultados insatisfatórios no caso contrário. A quantidade de iterações foi limitada em cada caso, para evitar a segmentação das cristas e consequente detecção de suas bordas. Veja na Figura 2 que as *snakes* com um valor muito pequeno para μ ($\mu = 0.1$) e com quantidades crescentes de iterações, iguais a 100, 500

e 1000, segmentam gradativamente, de fora para dentro, as cristas da impressão digital.

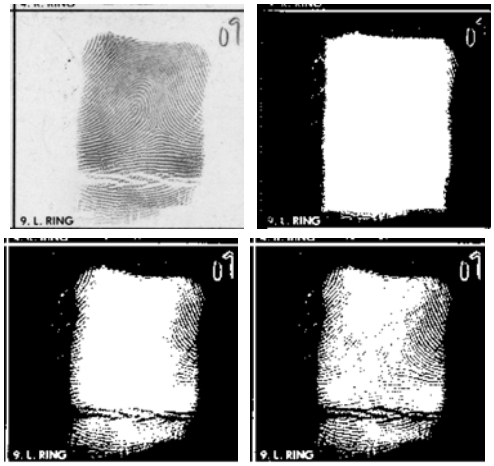


Figura 2. Imagem original n°24 no NIST SD27 [1] e resultados da aplicação do contorno ativo com $\mu = 0.1$ e 100, 500 e 150 iterações, respectivamente.

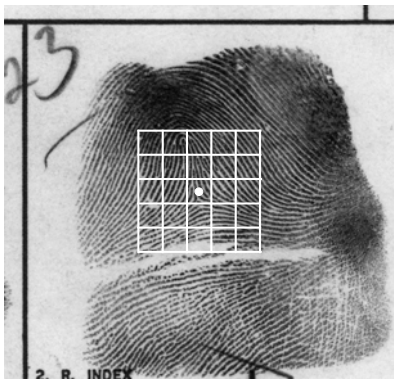


Figura 3. Ponto C e vizinhança considerados no ALGORITMO 1. O ponto C está em branco e as linhas representam os 25 blocos centrais.

6. Algoritmos

Primeiramente, o algoritmo para definição da máscara inicial automática é apresentado como ALGORITMO 1 e é utilizado apenas para imagens de boa qualidade. Nessa implementação, são fornecidos a imagem original I e o valor de $dimblc$ que determina a dimensão dos blocos utilizados durante o processo. A matriz m define a máscara e é inicializada como uma matriz nula com as dimensões de I. O ponto $C = (c_1, c_2)$ é o centro da imagem sendo o referencial da posição da impressão digital na imagem. Quando esta não estiver centralizada, deve-se fornecer um ponto pertencente à área da impressão. O procedimento seguinte define uma região retangular que contém a impressão digital. Para isso, faz-se uma busca, nas duas direções cartesianas e nos quatro sentidos a partir de C, dos blocos com variância menor ou igual a um limite (*lim*). Devido ao fato que a variância em blocos com *pixels* no interior da impressão digital é maior que em blocos com *pixels* do fundo, define-se *lim* como 20% da média da variância dos blocos vizinhos ao $bloco(c_1, c_2)$, centrado em C, e do $bloco(c_1, c_2)$, conforme Figura 3. O valor da porcentagem em *lim* é essencialmente experimental, não sendo o método sensível às suas alterações. A cada bloco encontrado, em cada sentido, sua posição é guardada e reinicia-se a busca em outro sentido. Tendo as quatro posições extremas em cada direção, são atribuídos valores 1 ao retângulo de vértices $\{(x_1, y_1), (x_1, y_2), (x_2, y_2), (x_2, y_1)\}$ em m.

ALGORITMO 1 (FUNÇÃO *mascara_inicial*)

Início

1. **Entrada:**
   I: impressão digital em escala de cinza.
2. **Parâmetros Fornecidos:**
   $dimblc$: tamanho do bloco.
3. **Máscara Inicial:**
   m = `zeros(` *dimensão*(I));
   (c_1, c_2) := posição (x, y) do centro de I;
   $mvar$:= média das variâncias dos 25 blocos centrais;
   $lim = mvar * 0.20$;
   $i = c_1$; $j = c_2$;
   enquanto (*variância*(*bloco*(i, j))$> lim$)
   $i = i - dimblc/2$;
   fim enquanto
   $x_1 = i$; $i = c_1$;
   enquanto (*variância*(*bloco*(i, j))$> lim$)
   $i = i + dimblc/2$;
   fim enquanto
   $x_2 = i$; $i = c_1$;
   enquanto (*variância*(*bloco*(i, j))$> lim$)
   $j = j - dimblc/2$;
   fim enquanto
   $y_1 = j$; $j = c_2$;
   enquanto (*variância*(*bloco*(i, j))$> lim$)
   $j = j + dimblc/2$;
   fim enquanto
   $y_2 = j$;
   $m(x_1 : x_2, y_1 : y_2) = 1$;
4. **Saída:**
   m: máscara binária inicial.

Fim

A proposta para detecção da área da impressão digital, esquematizada no ALGORITMO 2, é fornecer a ima-

gem I, o número de iterações nit e a constante μ para o método Chan-Vese. No caso de imagens roladas ou planas, o valor de $dimblc$, descrito acima, deve ser inserido para obtenção automática de m. Para latentes, a matriz de entrada m possui valores 1 formando uma região retangular que, sobreposta à imagem, envolve a latente e 0 nos demais elementos. No último caso, utiliza-se de uma imagem auxiliar, $I_{aux} = I(\texttt{find}^1(m))$, para ignorar as posições de I correspondentes aos elementos nulos em m, a fim de evitar incorreções no próximo passo. Na sequência, aplica-se o modelo Chan-Vese (encontrado em [12]) sobre o contorno inicial m para obter a variável $area$ e realiza-se sua correção efetuando a operação abertura com o elemento estruturante se.

ALGORITMO 2 (DETECÇÃO DA ÁREA)

Início

1. **Entradas:**
   I: impressão digital em escala de cinza.
   m: máscara inicial não automática.
2. **Parâmetros Fornecidos:**
   nit: número de iterações do método Chan-Vese.
   μ: constante do método Chan-Vese.
   $dimblc$: tamanho do bloco.
3. **Rolada/Plana:**
   se (I é imagem rolada/plana)
   m = *mascara_inicial*($I, dimblc$);
   fim se
4. **Latentes:**
   se (I é imagem latente)
   $I_{aux} = 255 \times \texttt{ones}($ *dimensão*(I));
   $I_{aux} = I($ $\texttt{find}($ m));
   $I = I_{aux}$;
   fim se
5. **Contorno Ativo:**
   $area$ = *chan-vese*(I, m, nit, μ);
6. **Abertura:**
   se := elemento estruturante disco de raio 16;
   $area$ = *erosão*($area, se$);
   $area$ = *dilatação*($area, se$);
7. **Saída:**
   $area$: máscara binária que define a área.

Fim

As matrizes resultantes da aplicação dos algoritmos 1 e 2, m e $area$, são matrizes binárias que possuem 1 nas posições que envolvem a impressão digital na imagem e 0 nas posições de fundo. Ou seja, a imagem produzida por m possui uma região retangular branca, que define a máscara inicial para aplicação do modelo *snakes*, em um fundo preto. No caso da imagem produzida por $area$, também temos uma região branca em um fundo preto, mas com a forma da área da impressão digital da imagem de entrada I, como pode-se ver nos resultados apresentados a seguir.

O *software* MATLAB® foi utilizado para realizar as implementações computacionais necessárias.

7. Resultados

O método proposto foi aplicado a impressões digitais roladas encontradas no banco de dados NIST SD27 [1]. A título ilustrativo, são mostrados apenas alguns resultados, entretanto muitos outros experimentos foram realizados. As Figuras 4 a 7 mostram os resultados dispostos de forma a apresentar a imagem original, a máscara inicial automática obtida pelo ALGORITMO 1 e a aplicação, sobre a imagem original, da área detectada pelo ALGORITMO 2. A quantidade de iterações é destacada em cada descrição enquanto o valor de μ é constante e igual a 20000.

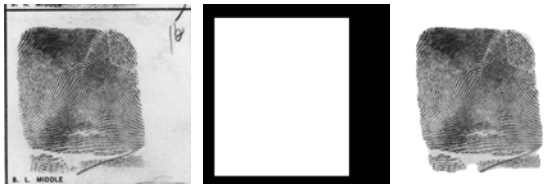


Figura 4. Imagem original, máscara inicial e resultado da proposta aplicada à rolada n° 4 em [1] ($nit = 150$).

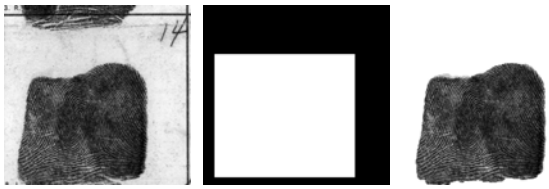


Figura 5. Imagem original, máscara inicial e resultado da proposta aplicada à rolada n° 9 em [1] ($nit = 90$).

Figura 6. Imagem original, máscara inicial e resultado da proposta aplicada à rolada n° 18 em [1] ($nit = 180$).

1 `find` é função padrão do MATLAB® que retorna as posições da matriz que são verdadeiras. No caso, onde m possui elemento igual a 1.

As latentes utilizadas também são do banco de dados NIST SD27 [1]. As Figuras 8 a 9 destacam a imagem original, a máscara inicial pré-definida manualmente e a aplicação, sobre a imagem original, da área detectada pelo ALGORITMO 2. A quantidade de iterações é destacada em cada descrição e $\mu = 20000$, novamente.

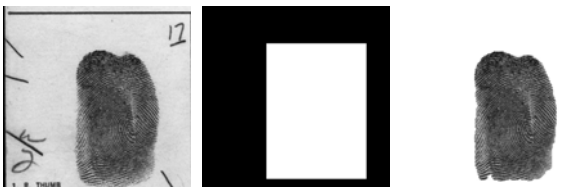


Figura 7. Imagem original, máscara inicial e resultado da proposta aplicada à rolada n° 25 em [1] ($nit = 200$).

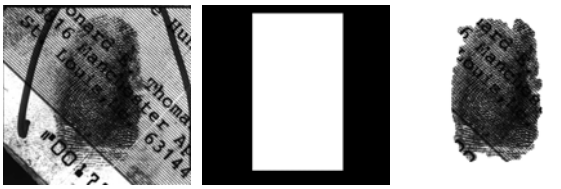


Figura 8. Imagem original, máscara inicial e resultado da proposta aplicada à latente n° 22 em [1] ($nit = 150$).

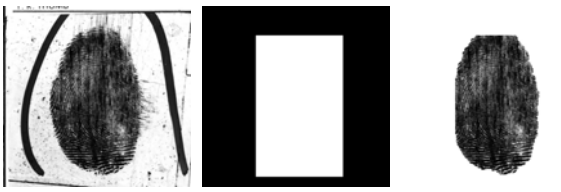


Figura 9. Imagem original, máscara inicial e resultado da proposta aplicada à latente n°111 em [1] ($nit = 200$).

8. Conclusões

O objetivo deste trabalho foi apresentar uma técnica para detecção automática da área da impressão digital em imagens de boa qualidade, baseada no modelo de contornos ativos, bem como uma extensão semi-automática para imagens latentes, através da pré-definição manual de uma máscara inicial retangular simples.

Apesar de ainda não terem sido aplicadas avaliações quantitativas, os testes da determinação da área com imagens roladas mostraram-se bastante satisfatórios. Verifica-se claramente que o fundo é desligado da impressão digital pelo processo aplicado, eliminando as informações indesejadas de riscos ou desenhos manuscritos e caracteres impressos fora da área da impressão digital. Apenas em alguns casos, em que os ruídos do fundo são muito densos e escuros, o procedimento não foi capaz de definir a área da impressão digital precisamente. Trabalhos futuros devem ser realizados para evitar esses problemas, provavelmente focando em processos para homogeneizar o fundo das imagens.

Além disso, o método semi-automático para detectar a área de latentes também obteve resultados relativamente bons, sendo a definição manual da máscara inicial retangular bastante simples. As dificuldades com imagens latentes são maiores principalmente por serem mais frequentes regiões com fundo bastante escuro. Trabalhos futuros deverão seguir no mesmo sentido que os de imagens roladas, focando inclusive em comparações com outras técnicas para validação da proposta.

Referências

[1] Fingerprint minutiae from latent and matching ten-print images. NIST Special Database 27, 2010. http://www.nist.gov/srd/nistsd27.htm.

[2] *Papiloscopia*. Wikimedia, 2011.

[3] K. L. S. Alexandre. Aplicação de contornos ativos em modelagem baseada em imagens. Master's thesis, Universidade de São Paulo - São Carlos, 2005.

[4] W. Bian, D. Xu, and Y. wei Zhao. Fingerprint segmentation based on improved active contour. *Networking and Digital Society, International Conference on*, 2:44–47, 2009.

[5] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(2), 2001.

[6] R. C. Gonzalez and R. E. Woods. *Digital Image Processing 2nd ed.* Printcie-Hall, New Jersey, 2002.

[7] M. Kass, A. Witkin, and T. Demetri. Snakes: Active contour models. *International Journal of Computer Vision*, pages 321–331, 1988.

[8] B. Laufer. History of finger-print system. *The Print*, 16(2):1–13, March/April 2000.

[9] C. W. Lima and M. F. Galvao. *Identificação Humana*. Academia de Polícia Civil do Distrito Federal, 2008.

[10] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer Publishing Company, Incorporated, 2nd edition, 2009.

[11] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[12] W. Yue. Active contour with Chan-Vese method for image segmentation. http://www.mathworks.com/matlabcentral/fileexchange/23445, 2009.

[13] X. Zheng, Y. Wang, and X. Zhao. Fingerprint image segmentation using active contour model. In *Proceedings of the Fourth International Conference on Image and Graphics*, ICIG '07, pages 437–441, 2007.

Fractal Image Encoding Using a Constant Size Domain Pool

Vitor de Lima[1] William Robson Schwartz[1] Hélio Pedrini[1]

[1]University of Campinas
Institute of Computing
Campinas-SP, Brazil, 13084-971

Abstract

Image compression techniques play an important role in data reduction and transmission. This work proposes a fractal image encoder based on a small domain pool with constant size constructed from the neighborhood of each range block and an efficient spatial subdivision data structure. This method is compared to a searchless algorithm using well-known grayscale images. The proposed approach is capable of performing fast compression and decompression, while maintaining high visual fidelity and operating at low bit-rates.

1. Introduction

Most fractal image encoders are characterized by extremely slow encoding times (taking a few hours to encode a single image) and fast decompression [5]. This is caused by the fact that constructing a fractal to approximate the content of an image is much more complex than displaying it [10]. Unfortunately, they are not competitive with other encoders based on transform coding both in encoding speed and rate-distortion performance.

A fractal encoder outputs a transform called *collage* that, given a downsampled version of the original image, it can create an approximation of the signal in its original scale. The usual method to create this is based on numerous searches for similar blocks between the image in these two scales which can be accelerated by specially designed heuristics [5].

A fast method for constructing a collage, called *searchless fractal compression*, was proposed by Furao and Hasegawa [3]. Each block in the original scale (called *range block*) is associated to a single block in the higher scale (called *domain block*). If the similarity between these two blocks is acceptable, the range block is encoded, otherwise, it is partitioned and the process is applied to the resulting sub-blocks. The encoders based on this concept are faster by orders of magnitude when compared to other fractal encoders employing more complex fractal-based algorithms, taking less than one second to encode an image instead of hours.

This work proposes a fast method for encoding images that is not as restrictive as the searchless encoders while keeping the encoding speed approximately at the same order of magnitude and improving the image quality at similar bit rates.

The next sections are organized as follows. Section 2 presents a brief review of fractal coding and searchless methods. The proposed method is described in Section 3. Experimental results comparing the proposed method against a searchless encoder are presented in Section 4. Finally, the conclusions of the work are given in Section 5.

2. Background

This section presents a brief review of fractal image coding and the searchless fractal encoders.

2.1. Fractal Coding

Fractal image encoders [5] generate a transform as output, called *collage*, that approximates the original image given a downsampled version of it. This transform is created in such way that when applied several times to any image, it will converge to an approximation of the original signal. Based on this property, starting with an arbitrary image, the decoder only needs to apply the collage and downsample its results repeatedly until it converges to the desired output.

Most fractal encoders construct the collage through a subdivision of the image into blocks (range blocks) that are matched against higher scale blocks (domain blocks) in the downsampled image by means of costly searching algorithms. This process makes the encoding process extremely expensive, specially when compared to the decoding time.

The collage can apply any affine transform to map a domain block into a range one. The collage also changes the

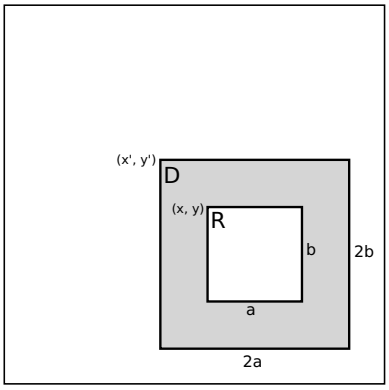


Figure 1. Relationship between range (R) and domain (D) blocks in a searchless fractal encoder.

intensities of the pixels contained in the domain blocks by using a transform such as the one proposed by Øien and Lepsøy [7], shown in Equation 1, where G is the resulting pixel intensity, D is the intensity in the domain block, $\bar{D}$ is the mean intensity of the pixels in the domain block, $\bar{r}$ is the mean intensity of the pixels in the range block and α is a parameter determined by a least squares estimation.

$$G(D) = \alpha(D - \bar{D}) + \bar{r} \tag{1}$$

2.2. Searchless Fractal Coding

Initially proposed by Furao and Hasegawa [3], the searchless fractal coding aims at quickly constructing a collage avoiding the complex search for the best match between range and domain blocks. Each range block at the coordinates (x, y) with dimensions $a \times b$ is matched against only one domain block with dimensions $2a \times 2b$ located at $(x - a/2, y - b/2)$, as shown in Figure 1. To encode a range block, only α and $\bar{r}$ must be transmitted, instead of the large number of parameters used in other state-of-the-art fractal encoders.

The image is partitioned into equally sized blocks using a quadtree, which are encoded by the process previously described. For each block. if the obtained error is larger than a certain threshold, the block is recursively divided into four sub-blocks until a certain minimum size is reached.

The searchless encoder proposed by Furao and Hasegawa [3] was later refined in [16] by employing a more flexible data structure to subdivide the image, in which a range block can be split into half either in the horizontal or vertical directions. There is no limit to the size of the range blocks, but smaller regions always have their α parameter set to zero and $\bar{r}$ is encoded using fewer bits.

3. Proposed Method

This paper proposes a new method for encoding the relationship between the domain and range blocks in images that is not as restrictive as the searchless methods, but is not as complex as the classical brute force or heuristic algorithms [1].

A more flexible domain pool with a larger selection of possible domain blocks is created to avoid the necessity of the searchless encoders to split certain regions of the image several times until they reach the desired reconstruction quality. This selection avoids such excessive use of subdivisions and also can reduce the number of bits used to encode the image.

For each range block with dimensions $a \times b$ located at (x, y), there are 9 possible domain blocks that can be used to represent it. The position of these blocks can be expressed as

$$\begin{aligned} x' &= x - a + p_x \times a/2 \\ y' &= y - b + p_y \times b/2 \end{aligned} \tag{2}$$

where p_x and p_y must be equal to 0, 1 or 2. All possible domain blocks given by this equation are illustrated in Figure 2. These candidate domain blocks have dimensions equal to $2a \times 2b$ and only one of them is chosen as the definitive mapping by testing which one of them has the lowest error when compared to the original range block, after having its gray values properly transformed by Equation 1.

If the range block must be split to achieve a lower reconstruction error, then it is divided into two equally sized sub-blocks. However, to do so, it must decide in which direction this subdivision will be performed. The heuristic used to choose how to properly partition a block tries to encode both sub-blocks separately without actually subdividing them, but by evaluating the two possible directions that can be used to split the original block. The direction resulting in the lowest estimated error is chosen.

The encoder initially subdivides the image into a uniform grid of blocks with 64×64 pixels. The rate-distortion heuristic proposed by Saupe et al. [12] is used for rate control, the blocks are inserted into a priority queue to sort the range blocks according to their sum of squared differences (SSD) instead of the mean squared difference (MSE), this replacement was proposed by Fisher and Menlove [2] to improve both the PSNR and the perceptual quality of the resulting image. At each iteration of the decoder, the range block with the largest SSD is subdivided into two blocks, which are reinserted into the queue. The total number of iterations, N_{it}, is defined by the user.

After the last iteration, the range blocks in the queue are transmitted by a context-adaptive arithmetic encoder [11], where α parameters are quantized into one of the values in the set $\{0.25, 0.5, 0.75, 1.0\}$, as suggested by Tong and

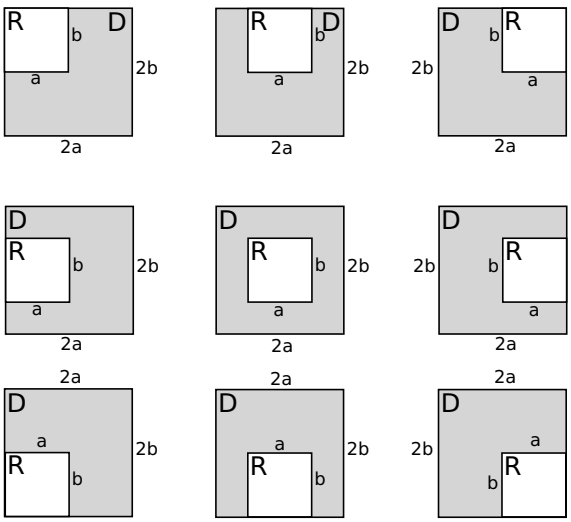


Figure 2. All possible relationships between the range (R) and the domain blocks (D) in the proposed encoder.

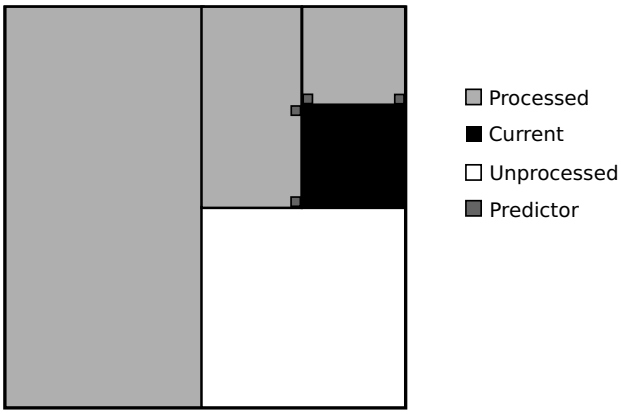


Figure 3. The prediction rule used to encode the mean values.

Pi [15], and encoded by an adaptive context chosen by the $\lfloor \log_2 A \rfloor$, where A is total area of the block, p_x and p_y values are encoded by their own adaptive contexts. Range blocks with only one or two pixels have their α parameter set to zero as suggested by Wu et al. [16].

The parameter $\bar{r}$ is encoded by a prediction method similar to the one proposed by Teuhola [14], in which the range blocks are scanned recursively from the left to the right and from top to bottom. Each visited range block predicts its mean value by averaging four mean values from neighboring blocks (as shown in Figure 3), quantizes the prediction and the real mean value using a uniform quantizer and an adaptive context chosen by the $\lfloor \log_2 A \rfloor$, as shown in Table 1 to encode the difference between them.

The spatial subdivision binary tree is transmitted by two different symbols: a binary flag for each node to mark it as either a leaf or an internal node, which is encoded by an adaptive context based on the $\lfloor \log_2 A \rfloor$ and another symbol to indicate the direction that node was split, which has two different contexts, one for blocks with b larger than a and another one for the remaining blocks.

The decoding process is accelerated by 3 different and complementary methods. The initial image that is used in the first iteration of the decoder is composed by filling each range block with its mean (for more details, see [6]). The used pixel intensity transform is the one proposed by Øien

Decision Table for the Quantization of $\bar{r}$		
area	quantization step	number of used bits
1	16	4
2	16	4
4	8	5
8	8	5
16	4	6
32	4	6
64	2	7
≥ 128	1	8

Table 1. Quantizers applied according to the area of the range block.

and Lepsøy [7] with additional proofs and details given by Pi et al. [9]. Each iteration is applied according to the Gauss-Seidel inspired method proposed by Hamzaoui [4], which uses only one image during the iterations overwriting each range block with its updated contents. The use of these methods assures that the decoding process converges in 4 iterations or less, instead of the usual 8 to 10 iterations used by other fractal decoders.

Finally, the image is post-processed by the same deblocking filter proposed by Fisher and Menlove [2] that adapts itself according to the size of the range block, ignoring the ones with smaller areas and using more aggressive parameters in the larger ones.

Images	Searchless		Constant-Sized Domain Pool	
	encoding time (ms)	decoding time (ms)	encoding time (ms)	decoding time (ms)
Baboon	21.9	19.2	100.6	19.3
Barbara	21.0	19.0	92.3	19.2
Boat	20.3	18.8	94.5	19.2
Goldhill	21.6	19.0	98.0	19.2
Lena	21.6	19.2	99.3	19.5
Peppers	21.4	19.0	99.3	19.3

Table 2. Average encoding and decoding time of the proposed and the searchless method for the benchmark images.

4. Experimental Results

All experiments were conducted on an Intel Core 2 Duo E6750 processor, 2.66 GHz with 3GB of RAM running the Linux operating system. The method was implemented in C++ programming language using only integer values and running on a single thread.

The proposed approach is compared to a searchless encoder using the standard grayscale benchmark images *Baboon*, *Barbara*, *Boat*, *Goldhill*, *Lena*, and *Peppers*, with 512×512 pixels. The metric used to compare the original and the decoded images is the peak signal-to-noise ratio (PSNR), which can be calculated as:

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right) \tag{3}$$

where MSE is the mean squared error between the compared images. The resulting PSNR for each image using both methods at different bit-rates is shown in Figure 4.

The searchless encoder used in the comparisons is based on exactly the same base code from the constant size domain pool encoder, but assigning a fixed value for p_x and p_y of each range block, it is extremely similar, but not exactly equal to the one proposed by Wu et al. [16]. This allows a fair comparison between the two proposed domain-range mappings, avoiding possible differences in the implementation.

Both encoders were compared by varying N_{it} from 100 to 10000 iterations. The rate-distortion curves show that the proposed method maintains a lower distortion in the compressed images at the same rates even though there are fewer range blocks in the transmitted data, since each block uses a larger amount of bits.

Table 2 presents the average encoding and decoding times for each image. Although the proposed domain pool is nine times larger if compared to the searchless method, the total encoding time is only increased by a factor of five since there are fewer range blocks. The number of range blocks, as shown in Table 3, is smaller because they are subdivided less frequently and consequently they cover larger areas.

Images	Number of Range Blocks	
	searchless	proposed method
Baboon	6005	4223
Barbara	5807	4025
Boat	5708	4025
Goldhill	6005	4223
Lena	5708	4025
Peppers	5708	4124

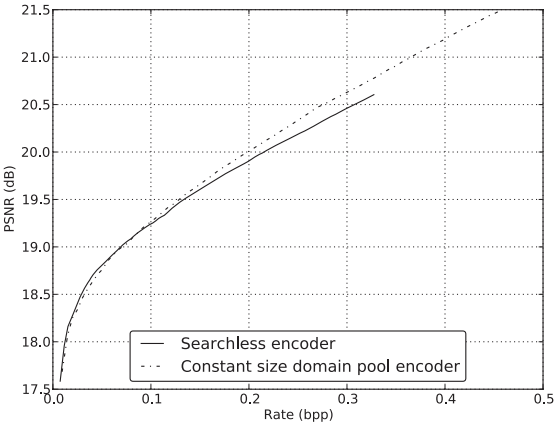
Table 3. Number of range blocks generated at 0.20 bpp for each tested image.

5. Conclusions and Future Work

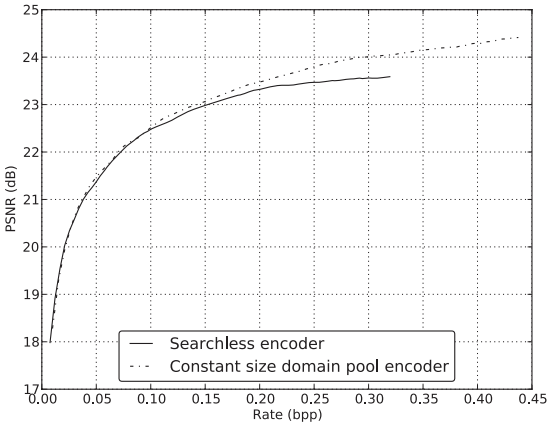
This paper proposed a novel domain pool for fractal coding which is an intermediary between the searchless encoders and the usual methods employing brute force or heuristics in a large domain pool.

Due to the smaller number of encoded range blocks, the proposed method outperforms the searchless encoder even though the latter uses less bits to encode each block. The additional complexity of the larger domain pool has a significant performance impact, however, the encoder still takes approximately 100 ms to encode a 512×512 image compared to several minutes of the original fractal encoders based on large domain pools.

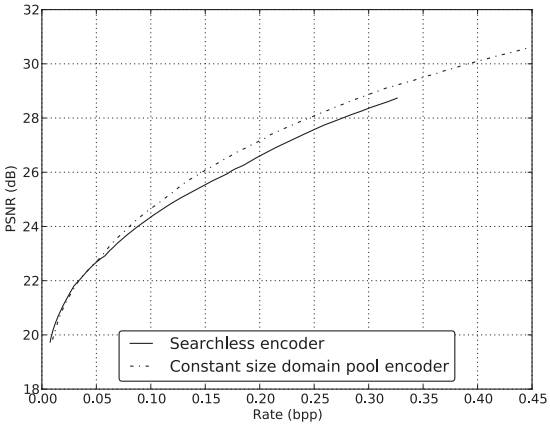
It is important to note that p_x and p_y parameters are transmitted without any form of optimized compression and



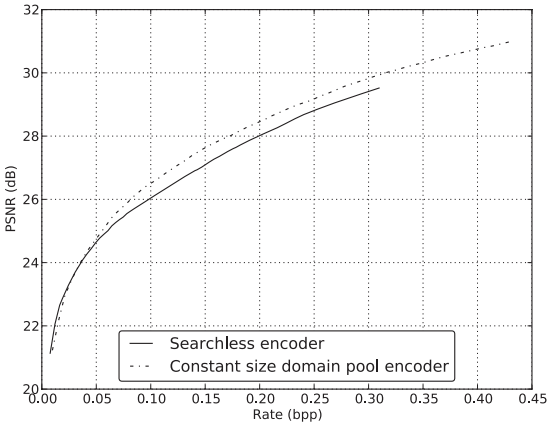
(a) Baboon



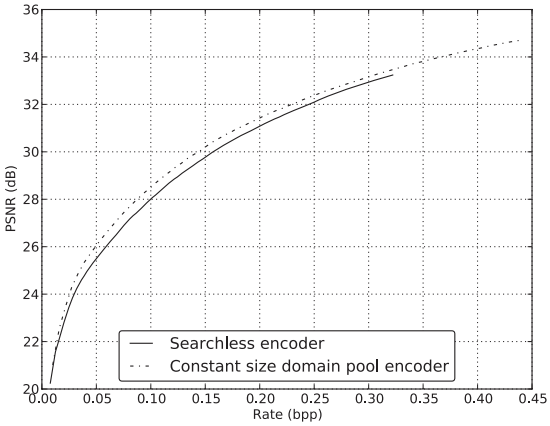
(b) Barbara



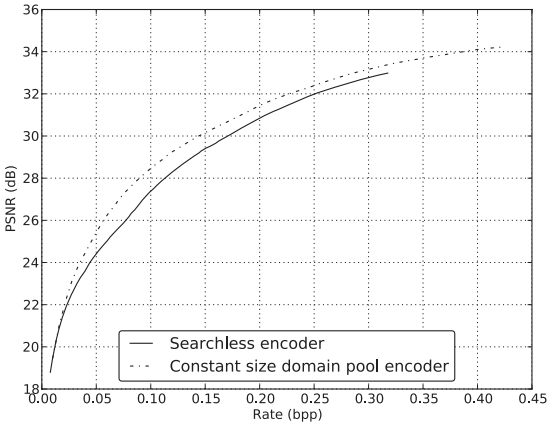
(c) Boat



(d) Goldhill



(e) Lena



(f) Peppers

Figure 4. Rate-distortion curve for a number of tested images.

the domain block is chosen by exhaustive search, which causes a major decrease in the rate-distortion performance and an increase in the encoding time of the proposed approach. Future experiments will analyze possible heuristics to choose the proper domain block and compress its relative position.

Other proposals for future work include the use of a pyramidal algorithm during the image decoding [13], a more efficient method for the quantization and coding of $\bar{r}$ parameter, improvements on the heuristics and data structures used in spatial subdivision, and the use of rate-distortion optimization methods to choose between different domain pools and quantizers [8].

Acknowledgements

The authors are thankful to FAPESP, CNPq and CAPES for the financial support.

References

[1] Y. Fisher. *Fractal Image Compression: Theory and Application*. Springer-Verlag, London, UK, 1995.

[2] Y. Fisher and S. Menlove. Fractal Encoding with HV Partitions. In Y. Fisher, editor, *Fractal Image Compression - Theory and Application*, pages 119–126. Springer-Verlag, London, UK, 1995.

[3] S. Furao and O. Hasegawa. A Fast No Search Fractal Image Coding Method. *Signal Processing: Image Communication*, 19(5):393–404, 2004.

[4] R. Hamzaoui. Fast Iterative Methods for Fractal Image Compression. *Journal of Mathematical Imaging and Vision*, 11:147–159, October 1999.

[5] A. Jacquin. Image Coding Based on a Fractal Theory of Iterated Contractive Image Transformations. *IEEE Transactions on Image Processing*, 1(1):18–30, Jan. 1992.

[6] Y. H. Moon, H. S. Kim, and J. H. Kim. A Fast Fractal Decoding Algorithm based on the Selection of an Initial Image. *IEEE Transactions on Image Processing*, 9(5):941–945, May 2000.

[7] G. Øien and S. Lepsøy. *A Class of Fractal Image Coders with Fast Decoder Convergence*, chapter Fractal Image Compression, pages 153–175. Springer-Verlag, London, UK, 1995.

[8] A. Ortega and R. K. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine*, 15(6):23–50, 1998.

[9] M. Pi, A. Basu, and M. Mandal. A New Decoding Algorithm based on Range Block Mean and Contrast Scaling. In *International Conference on Image Processing*, volume 2, pages II – 271–4 vol.3, Sept. 2003.

[10] M. Ruhl and H. Hartenstein. Optimal Fractal Coding is NP-Hard. In *Data Compression Conference*, pages 261–270, Mar. 1997.

[11] A. Said. *Lossless Compression Handbook*, chapter Arithmetic Coding. Communications, Networking, and Multimedia. Academic Press, 2003.

[12] D. Saupe, M. Ruhl, R. Hamzaoui, L. Grandi, and D. Marini. Optimal Hierarchical Partitions for Fractal Image Compression. In *IEEE International Conference on Image Processing*, 1998.

[13] I. Sutskover and D. Malah. Hierarchical Fast Decoding of Fractal Image Representation using Quadtree Partitioning. In *International Conference on Image Processing and Its Applications*, volume 2, pages 581–585, 1999.

[14] J. Teuhola. Fast Image Compression by Quadtree Prediction. *Real-Time Imaging*, 4:299–308, Aug. 1998.

[15] C. Tong and M. Pi. Fast Fractal Image Encoding based on Adaptive Search. *IEEE Transactions on Image Processing*, 10(9):1269–1277, Sept. 2001.

[16] X. Wu, D. Jackson, and H. Chen. Novel Fractal Image-Encoding Algorithm Based on a Full-Binary-Tree Searchless Iterated Function System. *Optical Engineering*, 44(10), 2005.

Feature Space Visualization to Image Analysis: Preliminary Results

Bruno Brandoli Machado, Danilo Medeiros Eler and João do E.S.B. Neto
Institute of Mathematical Sciences and Computing (ICMC)
University of São Paulo (USP)
Avenida Trabalhador São-carlense, 400
13560-970 São Carlos, SP - Brazil
{brandoli,eler,jbatista}@icmc.usp.br

Abstract

Image applications rely on the fact that the dataset under investigation is correctly represented through features. Despite of the advances in this sense, the definition of a set of features which properly represents a dataset is still a challenging and usually an laborious task. The available techniques considere a large number of features, and mostly, they are based on purely quantitative statistical measures or artificial intelligence-based approaches, and normally are black-boxes to the user. The approach proposed here seeks to inspect this black-box using visual representations, enabling users to interactively change the parameters of the feature extraction algorithms and to visually evaluate the outcome, getting insights about the representativeness of such features. The experiments using this approach shows that as the quality of the visual representation increases, changing the employed features or considering different parameters, the precision of distance based-classifiers also increases. Therefore, experiments show this approach can be used as a good guide to understand and to define a set of features which properly represents image datasets when classification tasks are considered.

1. Introduction

Computing feature vectors is a crucial task for many image-related applications, including *pattern recognition*, *content-based image retrieval* or *image mining*. It is able to determine the accuracy of the results achieved by aplications of this nature [14]. There are several algorithms to compute such vectors, each of which with their own set of parameters, reflecting different properties of the image data set under investigation. Therefore, the choice of a descriptor that provides the highest classification rates or the most efficient retrieval is a challenging task. Mostly, it requires *a priori* knowledge of the image data set [1]. Further, to create a coherent model several experiments may be required and this poses as a very expensive computational cost [9].

Usually, users start by selecting a labeled dataset and define the parameters for the feature extraction algorithm. Then, feature vectors are computed and classification is carried out. Results will only exhibit high classification rates if the set of features proves adequate for the image data set. However, this can only be assessed after the classification task has been perfomed. Thus, users cannot infer in advance how effective the computed features will be for the classification. Our main goal is to provide an effective scheme to help user to find the coherent feature space.

In this paper we propose a new approach to the visual analysis of feature spaces using information visualization techniques. Initially, different sets of features are extracted using distinct techniques or varying the parameters set. Then, the computed feature vectors are visualized in a 2D representation that reveals similarity relationships, forming clusters of similar images. To reduce subjectivity during analysis, a measure of the quality of the visual representation, called *silhouette coefficient* [13] is computed. The *silhouette* indicates how properly grouped similar images are, and also how far apart dissimilar ones are. The proposed approach is validated by comparing the classification rates and the *silhouette*. It is shown that whenever classification rates rises, the quality of the visual representation increases. Thus, the main contributions of this paper are: an approach for visual feature spaces analysis aimed at converging to representative features for image classification purposes and a method for objectively assessing visual representations for that same purpose.

This paper is organized as follows. Section 2 describes related work on visual analysis of feature spaces. Section 3 presents background information on feature extraction and information visualization techniques used in our approach. Section 4 presents the results of the experiments performed. Finally, conclusions and directions for future research are

presented in Section 5.

2. Related Work

Employing visualization techniques to explore and to draw knowledge from data sets is an efficient way to combine human intelligence with the powerful force of computation [1]. Several visualization techniques and tools have been developed enabling users to interact with abstract data [4], some of which are specially designed to explore multi-dimensional spaces resulting from the feature extraction process.

One of these approaches was proposed by Rodrigues et. al. [11] (further extended in [12]). The aim is to give support to the analysis of the features employed on similarity queries for a content-based image retrieval system. Once the feature space is compounded, a visual representation is created showing that the best visual representation conveys the best precision and recall measures when a query is executed on the data. However, the visual representation is not used to help an user to interactively define the best set of features or the best set of parameters to extract the features, but only to confirm that the query precision and recall matches with the visual representation quality. In our approach, the visual representation is meant as an interactive guide to explore, define and refine the set of features, giving insight on how the extraction algorithm parameters or the weighting of the features affect the similarity relationships between groups or individual images.

PEx-Image [6] is a similar tool, which employs a point placement visual representation to explore image collections. It was applied to create the best visual representation given a feature space. In contrast to the approach we present, in PEx-Image the quality of a visual representation is defined according to the user's point of view, while here we seek to create a visual representation which best reflects the feature space. Therefore, our approach gives more precise insight in the similarity relationships between images than [6]. In addition, we use a well-know measure, borrowed from the data-mining community, to assess such quality, reducing the subjectiveness of conclusions based purely on the visual analysis.

3. Background

3.1. Feature Extraction

Feature extraction is the process of capturing quantitative characteristics of an image and place them into a dimensional feature vector. In this paper, traditional texture analysis methods as co-occurrence matrix [7] and Gabor filters [2] features are used. On the former, 5 measures (*energy*, *entropy*, *inertia*, *inverse difference moment*, and *correlation*), 5 distances and 4 directions are considered, totaling 100 distinct features. On the latter, 16 features using energy of the responses of Glabor filters are computed (4 orientations and 4 scales).

3.2. Multidimensional Projection Techniques

Multidimensional projection techniques [10] aim at creating visual representations that enable users to employ their visual skills to recognize patterns and structures present in the data set. Each data instance, in this case an image, comprises to a visual element, represented as a circle, point or sphere, and mapped into a visual space that may be either {1,2,3}-D. The geometric positions of these elements reflect some type of relationship between data instances, such as similarities or neighborhood relationships [10]. Thereby, if the elements are closely placed on the visualization, it indicates that the data instances they represent are similar according to a certain distance. Conversely, if the elements are projected far apart, it indicates that the objects they represent are not related.

Currently, there is a number of different techniques considering different aspects of data distribution (we refer to [10] for more details). In this paper, we are interested in a technique that reflects as much as possible on the visual layout the distance relationships between the data instances in the original space. One that receive a great deal of interest for this task is *Classical Scaling* [5]. In *Classical Scaling*, a doubly-centered distance matrix between all pairs of data instances is defined, and a spectral decomposition is applied to recover the Cartesian coordinates of the elements in the visual space. It is possible to prove that if the distance function is Euclidean, the projected space presents the smallest mean quadratic deviation from the original space between all possible reduced spaces. That is, between all matrices $\mathbf{A}_{n\times p}$, the projection $Y_{n\times p}$ is the one that minimizes $\|\mathbf{X}-\mathbf{A}\|^2 = \sum_{i,j}(x_{i,j}-a_{i,j})^2$ [8].

3.2.1. Projection Quality with Cluster Validity The visual analysis approach proposed in this work aims at verifying whether similar images, according to the user point of view, are also similar according to the extracted set of features used to represent them. Although users may employ their visual abilities to determine the quality of a projection, it is rather difficult to tell small differences resulted from slightly modified set of features. In order to reduce this subjectivity problem, a measure called *silhouette coefficient* [13], which was originally proposed to evaluate results of clustering algorithms, is employed.

The silhouette coefficient measures both the cohesion and separation between instances of clusters (groups). Given a data instance d_i, its cohesion a_i is calculated as

the average of the distances between d_i and all other instances belonging to the same cluster of d_i. The separation b_i is the minimum distance between d_i and all other instances belonging to the other clusters. The silhouette of a projection is given as the average of the silhouette of all instances, as shown in Equation (1).

$$S = \frac{1}{n} \sum_{i=1}^{n} \frac{(b_i - a_i)}{max(a_i, b_i)} \quad (1)$$

The silhouette can vary between $-1 \leq S \leq 1$. Larger values indicate better cohesion and separation between clusters. In our approach, clusters are composed taking into account an image pre-classification process, and the silhouette indicates whether images belonging to the same class are more similar between themselves than images belonging to other classes. Therefore, the best set of features is the one which yields the projection with the largest silhouette coefficient.

The diagram of Figure 1 summarizes the approach suggested here. First, a set of image features is extracted and then projected onto a 2D space. Upon user's agreement on the goodness of the provided result (based on visual inspection or observation of the silhouette coefficient), the set of feature can then be used for other tasks, such as classification, retrieval or data mining. If significant misplacement is found, the parameters used to compute features ought to be changed, and the process re-executed. This process is performed until the expected outcome is reached, that is, until the projection separates what is similar from what is not.

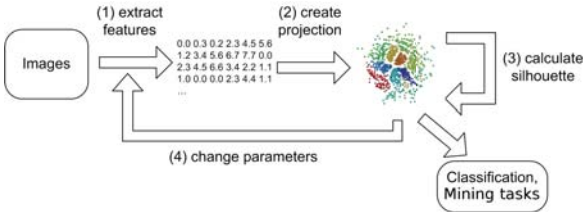


Figure 1. Our approach to visual exploration of the feature space.

4. Results

A set of 70 texture images from the Brodatz data set [3], grouped into 7 classes with 10 images each, is employed in this experiment. Features are extracted using Gabor filters, for 4 different orientations (0°, 45°, 90° and 135°) and 4 scales (numbered 1 to 4). The feature space visualization is depicted in Figure 2. Images are represented as circles, and are colored according to the class they belong to. It is important to notice that the class information available for each sample is not used to generate the projection, but only for display purposes. For clarity, one image sample of each class is placed near its corresponding cluster. In this projection 5 classes are well separated, while 2 are mixed up (top right-hand corner). This is a clear indication of a limited discriminant power of the chosen feature set. A closer look, on the other hand, reveals that in terms of texture pattern, both mixed up classes are very similar. Therefore, it is up to the user to decide whether such classes should in fact be considered as a single class. If that is the case, this can be taken as a good set of features. Otherwise, a new set should be evaluated.

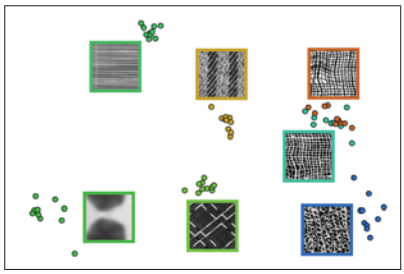
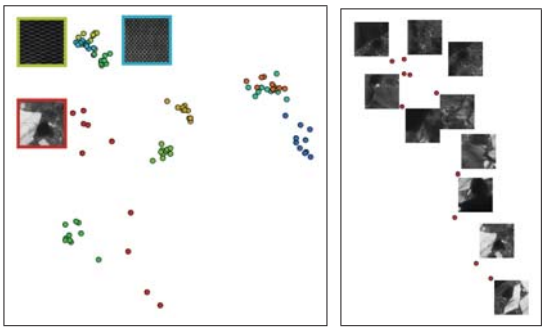


Figure 2. Analysis of Gabor features with a subset of Brodatz.

The experiment is extended with the addition of 3 new texture classes (marble and two wired-frame classes), and the resulting projection is shown in Figure 3(a). For clarity, image samples of such classes are exhibited with enlarged icons. The same set of parameters is used. Observe that the marble class, whose sample is highlighted in red, appears scattered throughout the projection. A user without much expertise can easily conclude that the chosen feature set is not suitable for representing the marble class, which probably matches with the outcome expected by an expert user. Images from the marble class normally exhibit non-uniform texture elements, which vary both in size and orientation. This characteristic is illustrated in Figure 3(b) that shows the zoomed-in area of the marble class depicted in Figure 3(a). Since a fixed set of parameters (for orientation and scales) is used, it is unlikely that all subtleties in marble textures with varied sizes and orientation can possibly be captured.

For the wired-frame classes, the projected samples are mixed up, as can be observed at the top left-hand corner of Figure 3(a). To reach better separability, only features with 90° orientation are now selected (user ignored the other three orientations by setting their weights to zero) and a new projection is computed. The resulting visualization is shown in Figure 4. Notice that these two classes are now



(a) (b)

Figure 3. Intraclass variation can be easily verified for Gabor filtering.

separated. This is expected as the texture patterns for orientation 90° are different in each class. However, when compared to the previous projection generated considering different features, the remaining classes are more spread. In this case, the user can conclude that the orientation 90° is effective in separating these 2 new classes of images, but the cohesion of the other classes is impaired. This outcome is not easy to reach without visualizing the feature space. In terms of silhouette coefficient, the features employed to generate Figure 4 are found to be better than the features employed to generate Figure 3(a). However, the difference is small, 0.429 for the former and 0.474 for the latter, which matches with the visual inspection.

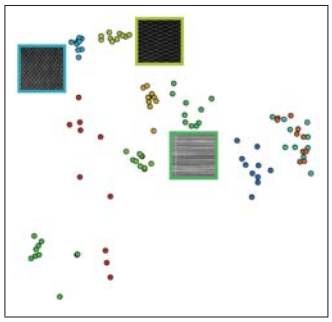


Figure 4. Gabor features for 90° orientation only.

Whichever orientations have been chosen in the previous experiment, it is possible to notice that Gabor features failed to separate samples of the two Brodatz's Linen classes. Their samples are colored in brown and cyan and are best seen at the top-right corner of Figure 2. They are also present in Figures 3(a) and 4.

In fact, these two classes exhibit very similar texture patterns and closer inspection reveals a slight variation in the pixel intensities only. Seeking better separability, new features extracted from co-occurrence matrix technique are investigated. To compute the gray-level co-occurrence matrices 4 orientations are considered 0°, 45°, 90° and 135°; as well as 5 distances, ranging from 1 to 5. Feature vectors are calculated over the matrices using 5 different statistical measures: entropy, correlation, energy, inverse difference moment, and inertia, yielding a feature space with 100 dimensions. The resulting projection is shown in Figure 5. For clarity, two enlarged images, each representing a distinct Linen class, are shown at the right. Notice that the brown and cyan samples are now well separated. Visual inspection indicates that co-occurrence matrix is the best choice, when compared with Gabor, providing better separability. The silhouette coefficient reinforces this perception. Projection with co-occurrence features yields 0.583 coefficient, against 0.474 for the Gabor experiment, as seen in Figure 4.

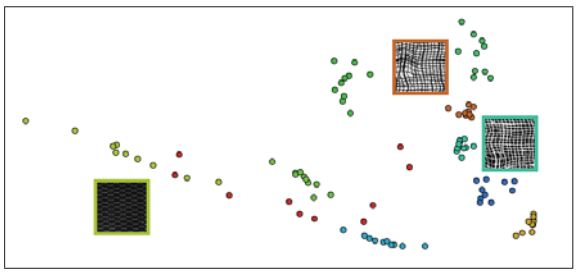


Figure 5. Projection of the Broadtz textures using co-occurrence features.

Although the co-occurrence feature set provides better separability for the two Linen classes, the wired fence texture images (Figure 5, left-hand side of the projection), exhibit inferior cohesion if compared to that provided by Gabor filters. In order to identify which statistical measures mostly influenced the scattering of this texture class, we generated projections considering each statistical measure separately. Silhouettes of these projections revealed that the energy measure is the one which yields the worst results. As a consequence, a new projection with all features, except energy, was created and is shown in Figure 6. The wired frame class is better grouped, while the cohesion of the other classes are preserved. One can infer that energy is a piece of information that impairs the representation of the data set under analysis. The silhouette coefficient for this projection, 0.607, is higher than those presented in the previous experiments and indicates that the feature set used here is not only quantitatively the best of the representa-

tions, but also matches with visual inspection.

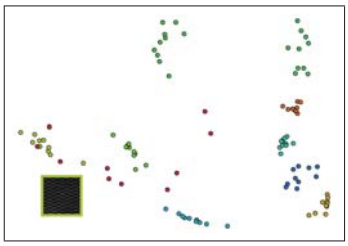


Figure 6. Projection of the Brodatz textures using co-occurrence features without energy.

The following examples seek to present how the proposed approach fits to classification tasks based on similarity. In this experiment, a texture data set[1] of 280 images and 7 different classes, each containing 40 distinct samples, is used. Table 1 shows the results of applying a K-NN with ten fold cross-validation classifier for Gabor features with different directions. $Test_1$, $Test_2$, $Test_3$ and $Test_4$ correspond to orientations 0°, 45°, 90° and 135°, respectively. $Test_5$ is the combination of 0° and 90° orientations, while $Test_{All}$ consists of all orientations combined. The silhouette coefficients and the classification rates matches, supporting the idea that if an user employs the best projection, the best results in terms of classification (based on similarity, such as K-NN) will follow.

θ	0°	45°	90°	135°	Silhouette	%
$Test_1$	1	–	–	–	0.1117	69.28
$Test_2$	–	1	–	–	-0.0164	60.01
$Test_3$	–	–	1	–	0.0164	68.57
$Test_4$	–	–	–	1	-0.0179	53.93
$Test_5$	1	–	1	–	0.2204	84.28
$Test_{All}$	1	1	1	1	0.2758	87.85

Table 1. Combinations of Gabor features.

In the second test we extract co-occurrence features from the same image data set and combine them with the Gabor features (0° and 90° orientations). Entropy (H), Correlation (COR), Inverse Difference Moment (IDM) and Inertia (I) are the chosen statistical measures for the co-occurrence matrix technique. Table 2 shows the results of silhouette coefficient and the classification rates. $Test_1$ combines COR and IDM, while $Test_{All}$ combines all features. Additionally, $Test_{Best}$ consider the best features based on the silhouette values of projections considering each individual feature. The employed features are 0° and 90° orientations for the Gabor filter, and COR and I for the co-occurrence matrix. Again, the idea of better projections (larger silhouettes) conveying better classification rates is confirmed.

Match	0°	90°	H	COR	IDM	I	Silhouette	%
$Test_1$	–	–	–	1	1	–	0.2590	81.96
$Test_{All}$	1	1	1	1	1	1	0.3023	83.74
$Test_{Best}$	1	1	–	1	–	1	0.3899	84.61

Table 2. Different combinations of features.

To assess the robustness of this approach in other scenarios, we create different feature subsets combining distinct Gabor and co-occurrence features. Then, for each space we create a projection and compute the silhouette. Also, different distance-based classifiers, other than K-NN, are applied on the resulting subsets. Results of this test are presented in Figure 7. The X-axis indicates the different sets of features, and the Y-axis shows the silhouette measures and classification rates obtained from K-means, Expectation Maximization (EM), and K-nearest neighbor (K-NN) classifiers. For the K-NN classifier a value of $K = 25$ is employed. For the remaining, the number of clusters is set to 7, which is the amount of available classes. Despite some mismatches, it can be observed that both the silhouette and the classification rates curves present the same trend, that is, when the silhouette increases the classification increases, and vice-versa.

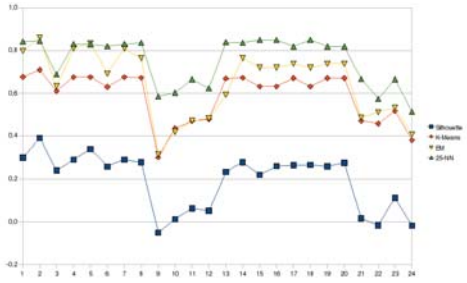


Figure 7. Relation between silhouette and classification rate for three different classifiers.

Finally, our last test checks how applying projections to select features on a reduced set of samples only (hereafter referred to as training set), compares with the task performed over the whole data set (test set). We employ a subset of the ImageCLEF data set[2], composed of 3,870 images and 35 classes (each class containing more than 50 and less

1 Available at `http://www-cvr.ai.uiuc.edu/ponce_grp/data/`

2 Available at `http://ir.shef.ac.uk/imageclef/2006/`

than 200 images). We use 15 images of each class as training set, yielding a total of 525 images. From this data set, 3 different sets of features are extracted: a) 16 Gabor features, b) 100 co-occurrence features and c) 1,024 wavelet features. Projections are created during the test and training stages for each feature set. Table 3 presents the results of the silhouette coefficient and classification rates for a K-NN classifier. The larger coefficients calculated for the training sets, also result on better classification rates for the whole data set. This correlation shows how the quality achieved for feature selection with the subset is preserved in the entire data set.

	Extractor	%(1-NN)	%(10-NN)	Silhouette
Training	Gabor	18.86	16.95	-0.29
	Co-occurrence	38.67	37.91	-0.22
	Wavelet	77.14	71.62	-0.19
Test	Gabor	14.11	15.46	-0.34
	Co-occurrence	35.16	36.08	-0.21
	Wavelet	77.10	75.67	-0.18

Table 3. Analysis of silhouettes and classification rates of 3 different sets of features.

5. Conclusions

This work has presented a novel approach, which combines interactive exploration of feature spaces using visual representations with quantitative analysis given by silhouette coefficients, in order to assess the quality of feature spaces representing image data sets. Several experiments were performed, showing that as the silhouette coefficient increases, by changing feature extraction algorithms or their parameters, the quality of the feature space, in terms of separability and cohesion of the image classes which belongs to it, increases too. Further, distance-based classifiers presented better classifications rates when feature spaces were the ones that produced displays with larger silhouettes.

Although there exist automatic methods for selecting features in high dimensional spaces, user interaction and perception play an important role in the process. Many times, different classes of images have, in fact, very similar samples. Other times, samples within a given class may be very non-uniform. The impact that such samples have on features and, consequently, in classification or feature selection algorithms, depends on the nature of the descriptor. Being capable of visually analyzing samples and image classes cohesion is a powerful approach to overcome the limitations imposed on traditional feature selection schemes and by the very nature of representing different types of images through features. As future work, it is necessary to seek theoretical basis to adapt the visual representation in order to be possible to use other non-purely-distance-based classifiers, such as SVM (Support Vector Machines).

Acknowledgments. The authors gratefully acknowledge the financial support of CNPq and FAPESP.

References

[1] C. C. Aggarwal. Towards effective and interpretable data mining by visual interaction. *ACM SIGKDD Explorations Newsletter*, 3(2):11–22, 2002.

[2] F. Bianconi and A. Fernández. Evaluation of the effects of gabor filter parameters on texture classification. *Pattern Recognition*, 40(12):3325–3335, 2007.

[3] P. Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover Publications, New York, 1966.

[4] S. K. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[5] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall/CRC, second edition, 2000.

[6] D. Eler, M. Nakazaki, F. Paulovich, D. Santos, G. Andery, M. Oliveira, J. Batista, and R. Minghim. Visual analysis of image collections. *The Visual Computer*, 25(10):923–937, 2009.

[7] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621, 1973.

[8] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis (Probability and Mathematical Statistics)*. Academic Press, 1995.

[9] E. Pampalk, W. Goebl, and G. Widmer. Visualizing changes in the structure of data for exploratory feature selection. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 157–166, New York, NY, USA, 2003. ACM.

[10] F. V. Paulovich, L. G. Nonato, R. Minghim, and H. Levkowitz. Least square projection: a fast high precision multidimensional projection technique and its application to document mapping. *IEEE Transactions on Visualization and Computer Graphics*, 14(3):564–575, May/Jun 2008.

[11] J. F. Rodrigues Jr., C. A. B. Castañón, A. J. M. Traina, and C. Traina Jr. Using efficient visual exploration techniques to evaluate features for content-based image retrieval. In *XVI Brazilian Symposium on Computer Graphics and Image Processing*, pages 183–190, Los Alamitos, CA, USA, 2003. IEEE Computer Society.

[12] J. F. Rodrigues Jr., A. J. M. Traina, and C. Traina Jr. Enhanced visual evaluation of feature extractors for image mining. In *The 3rd ACS/IEEE*, volume 1, pages 45–I, Washington, DC, USA, 2005. IEEE Computer Society.

[13] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.

[14] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, February 2006.

Um Algoritmo Rápido para o Reconhecimento de Caracteres

Francisco Assis da Silva[1,2], Almir Olivette Artero[1],
Maria Stela Veludo de Paiva[2], Ricardo Luís Barbosa[3]

[1]*Universidade do Oeste Paulista – Unoeste, Faculdade de Informática de Presidente Prudente – FIPP Presidente Prudente/SP*

[2]*Escola de Engenharia de São Carlos – Universidade de São Paulo – USP, Departamento de Engenharia Elétrica, São Carlos/SP*

[3]*Universidade Estadual Paulista – UNESP – Sorocaba/SP*

{chico, almir}@unoeste.br, mstela@sc.usp.br, rluisbarbosa@usp.br

Abstract

This paper presents a character recognition algorithm based on supervised classification. The strategy presented makes a description of characters classes, using it to modeling the pixel sequences behavior in the characters images. The transitions between levels of the adjacent pixels in the characters images that not occur are used to define the classes of each character. The experiments had indicated that the proposed strategy was very effective, even with a fairly difficult set of characters, with a very wide variation in the shape of characters in the same class.

1. Introdução

A identificação de caracteres manuscritos é uma área de intensa pesquisa [3][4][5][6][7][8], que ainda apresenta vários problemas, por causa da grande diversidade de formas que os caracteres podem assumir. A classificação supervisionada busca obter regras que permitam predizer, com precisão, a classe de cada amostra. Para isto, utiliza-se medições realizadas em caracteres, cujas classes são conhecidas. Quando a classificação é realizada de forma não supervisionada, é chamada de agrupamento e, neste caso, procura colocar os registros em classes, que apresentam grande similaridade entre seus integrantes e baixa similaridade entre os elementos de classes distintas. Embora tenha sido proposta uma grande variedade de algoritmos para realizar o agrupamento de dados [1], o número elevado de atributos, que deveria ser útil para uma melhor caracterização dos agrupamentos, geralmente compromete a execução da tarefa, o que é conhecido na literatura como a "maldição da alta dimensionalidade", por causa da incapacidade que a maioria dos algoritmos de agrupamento apresenta para tratar dados com muitos atributos.

Neste trabalho propõe-se uma estratégia simples e rápida para modelar o comportamento dos caracteres, usando apenas as transições que ocorrem entre os níveis de pixels adjacentes que formam os caracteres. As demais seções deste trabalho estão organizadas da seguinte maneira: na Seção 2 são apresentados alguns trabalhos relacionados com o reconhecimento de caracteres; na Seção 3, é apresentada a estratégia proposta neste trabalho para modelar o comportamento dos caracteres cujas classes são conhecidas (treinamento) e, então, obter as transições permitidas em cada classe e usar esta informação na classificação; a Seção 4 apresenta alguns experimentos, os resultados obtidos e a análise de desempenho com a aplicação desta proposta em um conjunto real de dados e na Seção 5, são apresentados os comentários finais e trabalhos futuros.

2. Trabalhos Relacionados

O trabalho de Jain e Ko [4] apresenta um algoritmo de classificação para reconhecer dígitos numéricos manuscritos (0-9), sendo utilizada uma implementação de uma PCA (Análise de Componentes Principais), combinada com o algoritmo 1-*Nearest Neighbor* (Primeiro Vizinho mais Próximo) para reconhecer os dígitos.

O trabalho de Montaña [5] realiza o reconhecimento padrão de dígitos, empregando redes neurais. Em seu trabalho são usadas duas diferentes redes neurais para alcançar os resultados, uma rede *Perceptron Multilayer* e uma rede baseada no mapa de *Kohonen*.

A contribuição do trabalho de Pereira et. al. [6] é melhorar a taxa de precisão de reconhecimento de caracteres manuscritos usando um novo método de extração de características, também baseado em Análise de Componentes Principais (PCA). Neste caso, aplica-se uma nova técnica que visa combinar os melhores aspectos de uma Análise de Componentes Principais Modular (MPCA) e uma Análise de Componentes Principais de Imagem (IMPCA).

No trabalho de Trentini et. al. [8] a partir da imagem segmentada de uma placa de automóvel, é realizada uma varredura em cada coluna da imagem e são contadas as quantidades de pixels pretos, representando a densidade correspondente a cada coluna. Para a segmentação dos caracteres, ou seja, separar os caracteres em relação ao fundo da placa é utilizada uma função de análise de máximos e mínimos locais. Para o reconhecimento dos caracteres é utilizado o algoritmo *Random Trees*, também chamado de *Random Forests*, o qual é um classificador baseado em árvores de decisão e pode reconhecer os padrões de várias classes ao mesmo tempo.

O trabalho de Shrivastava e Gharde [7] é utilizado para reconhecimento de números *Devanagari* manuscritos. *Devanagari* é um manuscrito alfabético usado por vários idiomas na Índia. Para a realização do trabalho os autores utilizaram a técnica de aprendizagem de máquina *Support Vector Machines* (SVM).

3. Modelagem do Comportamento das Sequências de Pixels

A estratégia proposta neste trabalho sugere descrever os caracteres, enquadrando-os em uma malha com dimensões definidas previamente e, em seguida, observar as transições entre os níveis de cinza (0 e 1 – imagens binárias) dos pixels adjacentes. Deste modo, uma imagem com dimensões *m* x *n* gera um vetor de transições (atributos) com *m* x *n* posições, sendo os pixels percorridos na sequência indicada na Figura 1. Assim, determina-se o comportamento das sequências de pixels em cada classe e, então, este conhecimento pode ser usado para classificar as poligonais de outros registros, para os quais não se conhece a classe.

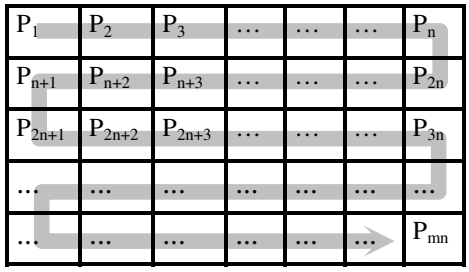
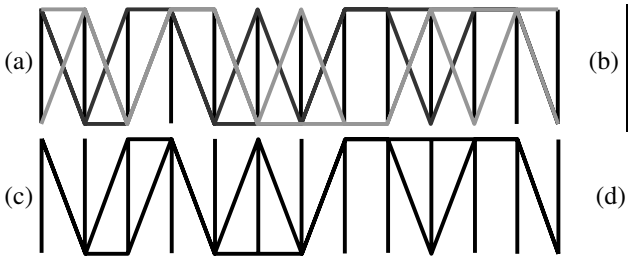


Figura 1. Sequência de pixels na imagem usada.

Nesta estratégia, é construída uma lista com as transições permitidas para cada caractere, anotando as transições que ocorrem entre os atributos adjacentes em cada uma das classes ou, então, as transições que não ocorrem em cada classe. Como as imagens usadas são binárias, as transições possíveis entre dois pixels que formam um caractere são: 00, 01, 10 e 11. A Figura 2 apresenta um exemplo em que as poligonais da classe 1 possuem apenas a transição 10, não ocorrendo as transições 01, 11 e 00 entre os atributos a_1 e a_2. Quanto à classe 2, nota-se que não ocorrem as transições 10 e 00. Assim, para cada classe são determinadas todas as transições que não ocorrem. Em seguida, a partir desta informação, os caracteres a serem reconhecidos são incluídos na classe que apresenta o menor número de inconsistências em relação às transições características anotadas em cada classe. A Figura 2 (a) exibe a visualização em coordenadas paralelas do conjunto de dados apresentado em (b), destacando as transições.



(b)

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}
classe 1	1	0	0	1	0	1	0	1	1	1	1	1	0
classe 1	1	0	1	1	0	0	0	1	1	0	1	1	0
classe 2	0	1	0	1	1	0	1	0	0	1	1	1	0
classe 2	1	1	0	1	1	0	0	0	0	1	0	1	1

Figura 2. a) Exibição dos registros da classe 1 em preto e da classe 2 em cinza; b) conjunto de dados contendo 4 registros com treze atributos em duas classes; c) Exibição dos registros da classe 1 e d) classe 2.

Em seguida, a classificação dos caracteres pode então ser conduzida usando as três regras de classificação propostas na Figura 3. A última regra (R_3) prevê um limiar definido pelo usuário para determinar quando o registro deve ser classificado como ruído. Por exemplo, classificando-o como ruído quando ele não atende ao menos 30% das restrições de regra alguma.

R_1 – Reconhecer o caractere na classe cujas transições não são violadas pelas transições do registro;

R_2 – Havendo mais de uma classe que satisfaz esta condição, a classe que possuir mais restrições (classe mais restritiva) deverá ser a escolhida;

R_3 – Quando o registro não atende todas as restrições de nenhuma classe, deverá ser inserido na classe menos violada, ou ser classificado como ruído em casos extremos.

Figura 3. Regras usadas para fazer a descrição dos caracteres.

As restrições das duas classes do conjunto de dados ilustrado na Figura 2 (b) são apresentadas na Tabela 1. No caso, as duas classes possuem quantidades diferentes de restrições, ou seja, não ocorrem as trinta transições na classe 1 e as trinta transições na classe 2, indicadas na Tabela 1.

Tabela 1. Transições que não ocorrem entre atributos adjacentes no conjunto de dados ilustrado na Figura 2 (b).

	Atributos adjacentes											
	a_1-a_2	a_2-a_3	a_3-a_4	a_4-a_5	a_5-a_6	a_6-a_7	a_7-a_8	a_8-a_9	a_9-a_{10}	a_{10}-a_{11}	a_{11}-a_{12}	a_{12}-a_{13}
classe 1	00 01 11	10 11	00 10	00 01 11	10 11	01 11	00 10 11	00 01 10	00 01	00 10	00 01 10	00 01 11
classe 2	00 10	00 01 11	00 10 11	00 01 10	00 01 11	10 11	01 11	01 10 11	00 10 11	00 01	00 10	00 01

Embora esta abordagem tenha sido proposta para operar com dados binários (com cardinalidade igual a dois), dados de maior cardinalidade também podem ser processados de duas formas diferentes. Na primeira os valores no conjunto de dados devem ser convertidos para binário, o que conduz a um aumento no número de atributos, melhorando o processo, pois aumenta o número de transições. A segunda possibilidade consiste em discretizar os valores dos atributos em número finito de níveis e então definir as transições para todos os níveis. Assim, dados dois atributos a_i e a_j, com cardinalidade c_i e c_j, respectivamente, o número de transições entre eles é dado pelo produto $c_i.c_j$. O número total de transições T para um conjunto de dados contendo n atributos é dado pela soma de todas as transições entre os atributos adjacentes c_i e c_{i+1} indicada pela Equação 1.

$$T = \sum_{i=1}^{n-1} c_i . c_{i+1} \qquad (1)$$

3.1 Melhorias no Processo

Duas alternativas que podem ser usadas para melhorar a qualidade das classificações usando esta estratégia são: 1) aumentar as dimensões da malha usada para representar os caracteres; 2) Anotar as transições entre três pixels no lugar das transições entre dois pixels, propostas incialmente. Assim, as transições a serem verificadas seriam: 000, 001, 010, 011, 100, 101, 110 e 111. Com estas duas estratégias, aumenta-se a quantidade de restrições usadas e, consequentemente, aumenta-se as chances de se classificar os caracteres corretamente.

4. Experimentos

Em seguida é apresentada uma análise do conjunto de dados *binaryalphadigs* (*UCI – Machine Learning*)[2], usando a estratégia apresentada. Este conjunto de dados é formado por 390 registros, obtidos a partir das imagens de 39 caracteres (números: 0, 1, ..., 9), escritos à mão, conforme ilustra a Figura 4.

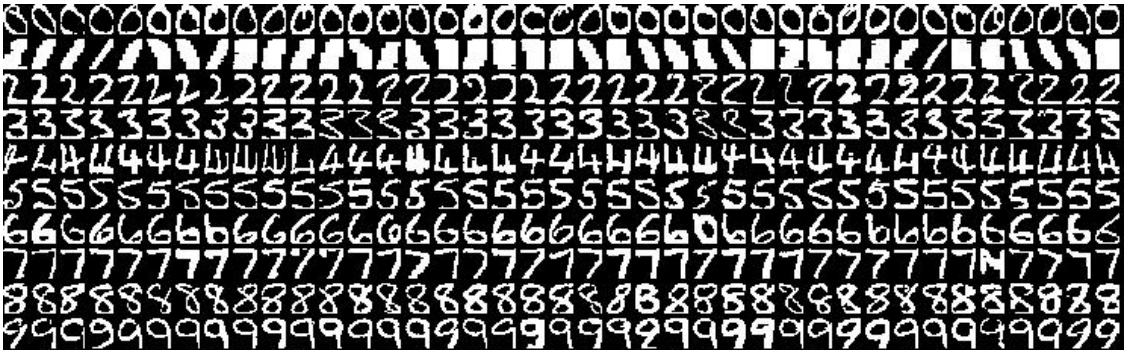


Figura 4. Trinta e nove caracteres (números: 0, 1, ..., 9) escritos à mão.

No experimento realizado, cada caractere foi digitalizado em uma grade de 16 x 20 pixels, usando apenas as cores preto e branco. Assim, cada caractere é representado neste conjunto através de 320 atributos que podem assumir os valores zero (preto) ou um (branco).

A Figura 5 mostra as transições entre os 320 atributos. A visualização dos registros em algumas de suas classes revela um comportamento particular para cada classe, evidenciando diferentes conjuntos de restrições para cada classe.

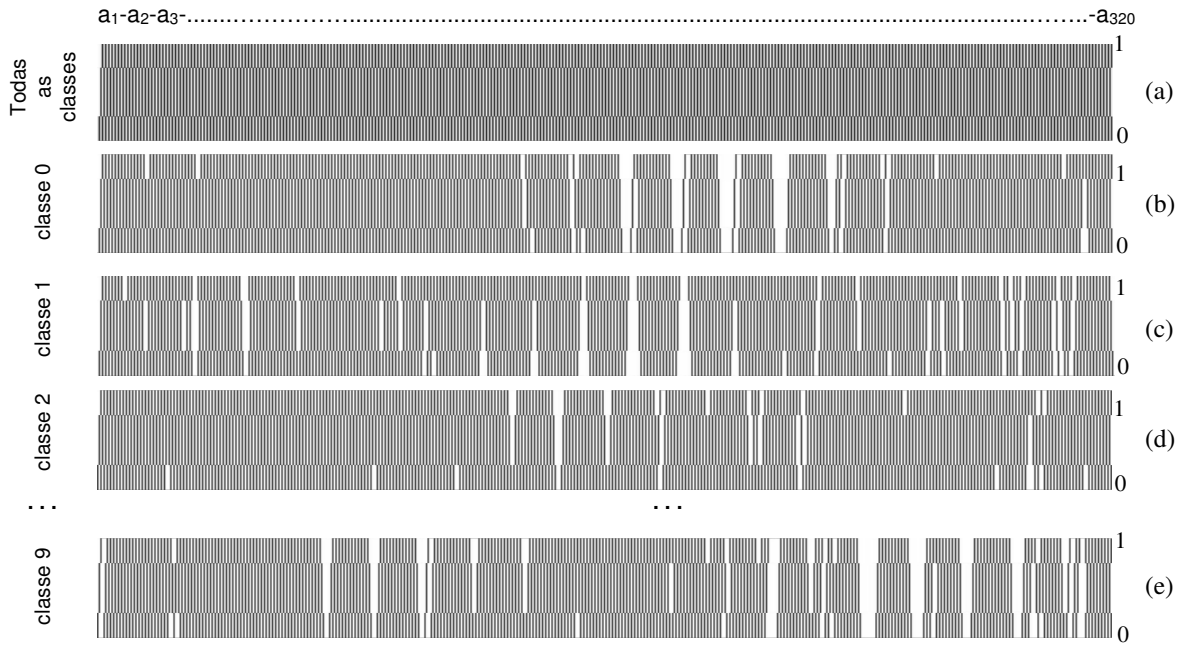


Figura 5. Visualização obtida usando: a) todas as transições de todos os caracteres nas dez classes; b) apenas os registros da classe do caractere "0"; c) apenas os registros da classe do caractere "1"; d) apenas os registros da classe do caractere "2"; e) apenas os registros da classe do caractere "9".

As quantidades de restrições nas classes zero até nove são, respectivamente, 228, 315, 155, 197, 117, 188, 196, 277, 88 e 272. Assim, a classe dos caracteres "1" é a mais exigente (possui um número maior de restrições), enquanto que a classe dos caracteres "8" possui o menor número de restrições. Como a cardinalidade de todos os atributos é igual a dois, e o conjunto tem 320 atributos, o número total de transições entre os atributos do conjunto, obtido usando a Equação 1, é dado por 1.276.

A matriz de confusão apresentada na Figura 6 ilustra a eficácia da estratégia para classificar os caracteres de acordo com as regras estabelecidas na Seção 3, ou seja, o registro na classe cujas transições não são violadas pelas transições do registro. Nos casos em que se encontra mais de uma classe satisfazendo esta condição, o registro é inserido na classe que possui mais restrições.

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	39	0	0	0	0	0	0	0	0	0
1	0	39	0	0	0	0	0	0	0	0
2	0	0	39	0	0	0	0	0	0	0
3	0	0	0	39	0	0	0	0	0	0
4	0	0	0	0	39	0	0	0	0	0
5	0	0	0	1	0	38	0	0	0	0
6	0	0	0	0	0	0	39	0	0	0
7	0	0	0	0	0	0	0	39	0	0
8	0	0	0	0	0	0	0	0	39	0
9	0	0	0	0	0	0	0	0	0	39

Figura 6. Matriz de confusão obtida com a classificação usando as transições entre os atributos adjacentes.

Apesar da dificuldade com este conjunto de caracteres (Figura 4), a matriz de confusão mostra que a maior parte dos registros foi colocada em suas devidas classes, exceto um registro da classe 5 (caractere "5"), que foi colocado na classe 3 (caractere "3"). Trata-se do 23º caractere "5" na Figura 4, que atende todas as restrições das classes 3 e 5 e, pela regra R_2 (Figura 3), foi reconhecido como da classe 3, porque ela possui um número maior de restrições que a classe 5.

Uma solução imediata para este problema é aumentar o tamanho da malha em que a imagem é processada, consequentemente, gerando uma quantidade maior de transições. Este efeito pode ser confirmado nas Figuras 7, 8 e 9 que mostram os resultados obtidos com diferentes tamanhos de malha. Na Figura 7 tem-se um resultado ruim usando a malha 10x8 (reduzida).

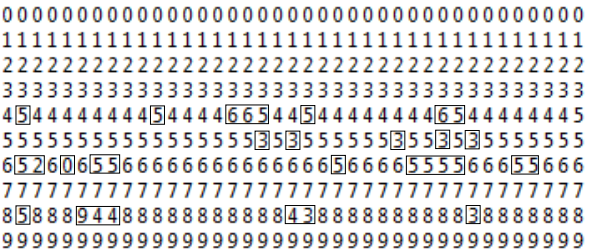


Figura 7. Resultado ruim da classificação usando uma malha 10x8.

Na Figura 8 tem-se o resultado já apresentado na Figura 6, que usa a malha 20x16, e classifica um caractere “5”, na classe dos caracteres “3”.

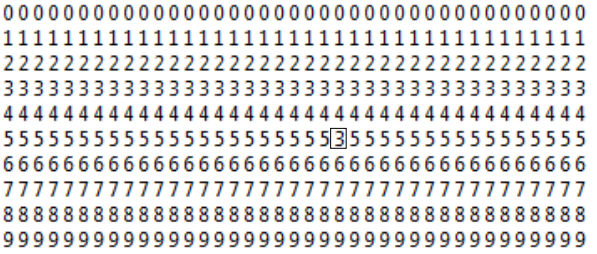


Figura 8. Resultado razoável da classificação usando uma malha 20x16.

Na Figura 9, usando uma malha 40x32, tem-se um resultado melhor, com todos os caracteres classificados em suas devidas classes. A explicação para a melhoria nos resultados é que quando se aumenta a malha, aumenta-se também a quantidade de transições, o que diminui a chance de se classificar os caracteres em classes erradas.

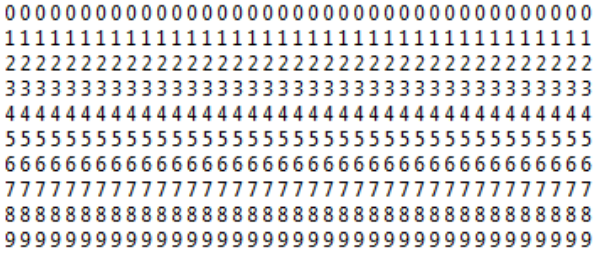


Figura 9. Resultado muito bom da classificação usando uma malha 40x32.

Em seguida, utilizando transições entre três pixels, tem-se o seguinte resultado para a malha 10x8, ilustrado na Figura 10, onde se verifica uma quantidade de erros bem menor, se comparada com o resultado da Figura 7, que também usa a malha 10x8, porém, usando transições entre dois pixels.

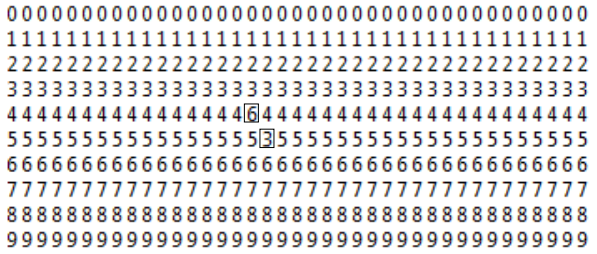
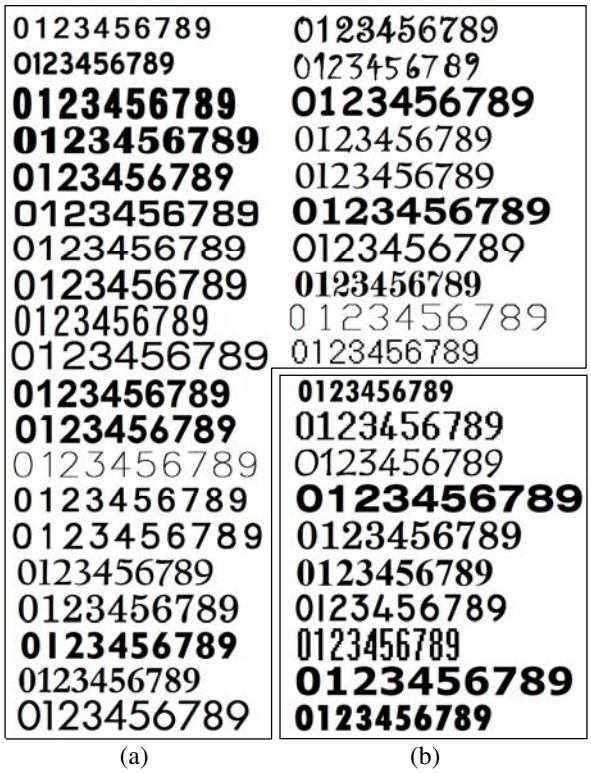


Figura 10. Resultado melhor para a classificação usando uma malha 10x8.

Para as malhas 20x16 e 40x32, o uso das transições entre três pixels elimina completamente os erros de classificação.

Em seguida, é apresentado um outro experimento, usando o conjunto de imagens de caracteres exibido na Figura 11. Neste experimento, o treinamento foi realizado usando apenas os caracteres em (a), enquanto que o teste foi feito usando apenas os caracteres em (b), de fontes diferentes.



(a) (b)

Figura 11. a) Conjunto de caracteres usados no treinamento do classificador (30 fontes diferentes); b) Conjunto usado no teste (outras 10 fontes diferentes).

Os resultados deste experimento são apresentados na Figura 12, sendo que em (a), são utilizadas as transições entre dois pixels, gerando um total de quatro erros de classificação, enquanto que em (b), usando as transições entre três pixels, foi obtida a classificação correta para todos os caracteres.

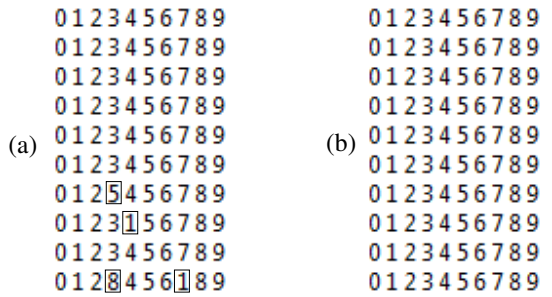


Figura 12. Resultado da classificação usando uma malha 40x32 com transições a) dois pixels; b) três pixels.

4.1 Análise de Desempenho

Os tempos médios de processamento, em segundos, do algoritmo proposto, usando o conjunto de dados *binaryalphadigs* (Figura 4) é apresentado na Tabela 2, usando transições entre dois e três pixels, com diferentes resoluções de malha. A máquina usada possui um processador Intel Core i3 M330 2.13 GHz e memória de 4 GB de RAM.

Tabela 2. Tempos médios de processamento (segundos) para um caractere usando o algoritmo proposto (conjunto de dados *binaryalphadigs* – Figura 4).

	Resoluções		
Transições	10 x 8	20 x 16	40 x 32
2 pixels	$3{,}4489.10^{-5}$	$1{,}5116.10^{-4}$	$6{,}2616.10^{-4}$
3 pixels	0,0000782	0,0002851	0,0010766

A Tabela 3 apresenta os tempos médios, em segundos, de processamento, usando os caracteres impressos da Figura 11. Neste caso, são usadas transições entre dois e três pixels, com a resolução de malha de 40x32.

Tabela 3. Tempos médios de processamento (em segundos) para um caractere usando o algoritmo proposto (caracteres da Figura 11).

	Resolução
Transições	40 x 32
2 pixels	0,000683
3 pixels	0,001083

Estes resultados mostraram que o algoritmo proposto consegue identificar caracteres em uma taxa média de 923,36 caracteres por segundo, usando a transição de 3 pixels, com a malha 40x32 (melhores resultados). Usando a transição de 2 pixels, com uma malha 10x8, o algoritmo consegue identificar caracteres em uma taxa média de 28.994,75 caracteres por segundo, o que é bastante razoável para tarefas que precisam realizar identificações de caracteres em tempo real.

5. Conclusões

A classificação de caracteres usando as transições entre os níveis dos pixels adjacentes se mostrou bastante eficiente, mesmo com um conjunto de caracteres tão difícil como os caracteres manuscritos, pois como se observa, o caractere "1" (ver Figura 4), por exemplo, apresenta uma grande variação, que desafia a maioria dos algoritmos conhecidos. A descrição dos caracteres a partir de transições de níveis entre os pixels se mostrou uma estratégia muito simples e rápida, que pode ser facilmente implementada em hardware e usada em aplicações em tempo real. O uso de malhas de maiores dimensões resulta em um aumento no número de transições entre pixels, contribuindo para a melhoria dos resultados. Do mesmo modo, o uso de transições entre três pixels também contribui para aumentar o número de combinações e transições, descrevendo melhor cada classe de caracteres. Este algoritmo também está sendo aplicado em reconhecimento automático de placas de sinalização de velocidade, para fins de georreferenciamento automático das mesmas, com resultados bem promissores.

6. Referências Bibliográficas

[1] Berkhin, P., Survey Of Clustering Data Mining Techniques, Accrue Software, San Jose, CA, *http://citeseer.ist.psu.edu/berkhin02survey.html*, 2002.

[2] Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository. http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science.

[3] Gonzalez, R.F., Woods, R.E., *Digital Image Processing*, Addison-Wesley, 1993.

[4] Jain, G.; Ko, J. "Handwritten Digits Recognition". Multimedia Systems, *Project Report*, University of Toronto, November 21, 2008, pp. 1-3.

[5] Montaña, E. G. Digits recognition via neural networks (2007) *http://www.pdf-finder.com/Digits-recognition-via-neural-networks.html#*.

[6] Pereira, J. F.; Alves V. M. O.; Cavalcanti, G. D. C. and Ren, T. I. "Modular Image Principal Component Analysis for Handwritten Digits Recognition". IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing, Center of Informatics Federal University of Pernambuco Recife, Brazil.

[7] Shrivastava, S. K.; Gharde, S. S. "Support Vector Machine for Handwritten Devanagari Numeral Recognition". International Journal of Computer Applications (0975 – 8887), Vol. 7– N.11, October 2010.

[8] Trentini, V. B.; Godoy, L. A. T.; Marana, A. N. "Reconhecimento Automático de Placas de Veículos". Anais do WVC 2010 - VI Workshop de Visão Computacional. Presidente Prudente, 2010, pp. 267-272.

Image decomposition with anisotropic diffusion applied to leaf-texture analysis

Bruno Brandoli Machado
ICMC – University of São Paulo (USP)
brandoli@icmc.usp.br

Wesley Nunes Gonçalves and Odemir Martinez Bruno
IFSC – University of São Paulo (USP)
wnunes@ursa.ifsc.usp.br
bruno@ifsc.usp.br

Abstract

Texture analysis is an important field of investigation that has received a great deal of interest from computer vision community. In this paper, we propose a novel approach for texture modeling based on partial differential equation (PDE). Each image f is decomposed into a family of derived sub-images. f is split into the u component, obtained with anisotropic diffusion, and the v component which is calculated by the difference between the original image and the u component. After enhancing the texture attribute v of the image, Gabor features are computed as descriptors. We validate the proposed approach on two texture datasets with high variability. We also evaluate our approach on an important real-world application: leaf-texture analysis. Experimental results indicate that our approach can be used to produce higher classification rates and can be successfully employed for different texture applications.

1. Introduction

Texture plays an important role in pattern recognition and computer vision. Applications with textures are found in several areas, including *remote sensing* [6] and *plant leaf identification* [2]. Though texture is easily perceived by humans, it has no precise definition due to its spatial distribution. In addition, physical surface properties produce distinct texture patterns. Thus the lack of a formal definition of texture is reflected into different methods for texture analysis.

Many methods for texture description have been proposed in the literature [23]. They are based on *statistical analysis* of the spatial distribution (e.g., co-occurrence matrices [13, 12] and local binary pattern [14]), *stochastic models* (e.g., Markov random fields [8]), *spectral analysis* (e.g., Fourier descriptors [1], Gabor filters [11] and wavelets transform [9]), *structural models* (e.g., mathematical morphology [21] and geometrical analysis [7]), *complexity analysis* (e.g., fractal dimension [19, 5]), *agent-based model* (e.g., deterministic tourist walk [3]). Despite there are effective texture methods, few papers are concerned in enhancing the richness of the texture attribute before computing features.

Inspired by biological vision studies, the community of computer vision has also shown a great deal in representing images using multiple scales. The basic idea is to decompose the original image into a family of derived images [17, 22]. The decomposition is obtained by convolving the original image with an image operator, for example, a simple way is to employ Gaussian kernels. Although the Gaussian filtering satisfies the heat equation, its derivatives cause spatial distortion in region boundaries. It implies that the diffusion process is equally in all directions, that is, the diffusion is linear or isotropic. On the other hand, Perona and Malik formulate a new concept that modified the linear scale-space paradigm to smooth within a region while preserving edges.

Due to the increasing interest in image analysis, we propose a novel framework to model textures. In the proposed approach, image decomposition using anisotropic diffusion of Perona and Malik is performed before feature extraction. The anisotropic diffusion process is mathematically modeled by partial differential equations (PDEs). The decomposition is applied to extract the texture component, obtained by the difference between the original image and cartoon approximations. Then, Gabor filters are used to extract features from the texture component, which presents more enhanced structures.

The remaining of this paper is organized as follows. Section 2 presents background information on nonlinear diffusion and Gabor filters. Section 3 details our approach in texture analysis. Section 4 presents the results of the experiments performed on two benchmark texture datasets. Finally, conclusions and directions for future research are given in Section 6.

2. Background

In general, texture analysis is studied into five groups: (1) *synthesis*, (2) *segmentation*, (3) *shape from texture*, (4) *compression* and (5) *classification*. All groups have been influenced by the use of decomposition and filter banks. Next we describe image decomposition using anisotropic diffusion and Gabor filters.

2.1. Anisotropic Diffusion

Scale-space theory has been investigated for representing image structures at multiple scales. The idea is to decompose the initial image into a family of derived images. According to [22] and [15], a family of derived images may be viewed as the solution of the heat equation and described using partial differential equations (PDEs). The successful use of PDEs in image analysis is assigned to the power to model many dynamic phenomenon, including diffusion. A new paradigm of nonlinear PDEs for image enhancement was introduced by Perona and Malik [20]. Their formulation, called anisotropic diffusion, uses a nonlinear scheme that smoothes images by creating cartoon approximations, while the region boundaries remain sharp. Formally, the discrete formulation of Perona-Malik is defined as:

$$I_{i,j}^{t+1} = I_{i,j}^{t} + \lambda \left[c_N . \nabla_N I + c_S . \nabla_S I + c_E . \nabla_E I + c_W . \nabla_W I \right]_{i,j}^{t} \quad (1)$$

where $0 \leq \lambda \leq 1/4$ is a scalar that controls the numerical stability, ∇I is the gradient magnitude, c is a constant value for the conduction coefficient, N, S, E and W are the mnemonic subscripts for North, South, East and West. The PDE equation above can be write as follows ($(i,j) \equiv s$):

$$I_s^{t+1} = I_s^t + \frac{\lambda}{\xi_s} \sum_{\rho \in \xi_s} g(\nabla I_{s,\rho}) \nabla I_{s,\rho} \quad (2)$$

where I_s^t is the cartoon approximation image, t denotes the number of iterations, s denotes the pixel position, ξ_s represents the number of neighbors of pixel s (usually 4-connectivity), and $g(\nabla I)$ is the conduction function. The value of the gradient is computed by linearly approximating its norm in a specific direction as:

$$\nabla I_{s,\rho} = I_\rho - I_s^t, \rho \in \xi_s \quad (3)$$

Perona and Malik proposed two functions of diffusion:

$$g\left(||\nabla I||\right) = e^{-(||\nabla I||/K)^2} \quad (4)$$

and

$$g\left(||\nabla I||\right) = \frac{1}{1 + \left(\frac{||\nabla I||}{K}\right)^2} \quad (5)$$

The parameter K controls the conduction. The first equation favours high contrast edges over low contrast ones, while the latter favours wide regions over smaller ones. Although Perona and Malik proposed two different functions, the smoothed images are quite similar.

A texture decomposition with anisotropic diffusion is shown in Figure 1. The first row shows the family of cartoon approximations from the original image I_0. We can observe that the information is gradually smoothed, while textures, in third row, are enhanced by the difference between the original image and cartoon approximations. The solution of the heat diffusion is depicted in rows 2 and 4. Note that the distribution of heat correponds to gray values in the image and the diffusion time is represented by the number of iterations t. For different scales t we obtain different levels of smoothing, as shown from t_1 (Figure 1(b)) to t_5 (Figure 1(f)).

2.2. Gabor Filters

A Gabor filter is a signal sinusoidal plane wave modulated by a Gaussian [11]. The filters used in image decomposition are created from a "mother" Gabor function of two dimensions, for a given space $g(x,y)$ and frequency $G(x,y)$ domains. Given the "mother" function, a bank of Gabor filters can be obtained in the $g(x,y)$ space domain from operations of dilatations and rotations.

Initially, the Gabor technique generates a filter bank $g_{mn}(x,y)$ for different scales $m = 1, \ldots, K$ and orientations $n = 1, \ldots, S$ parameters. Texture features are computed by convolving the original image I with the Gabor filter bank, as depicted in Equation (6). By tunning the values of m and n, some aspects of the image's underlying texture structure can be captured. In this work, a number of 40 Gabor features have been computed (8 orientations and 5 scales).

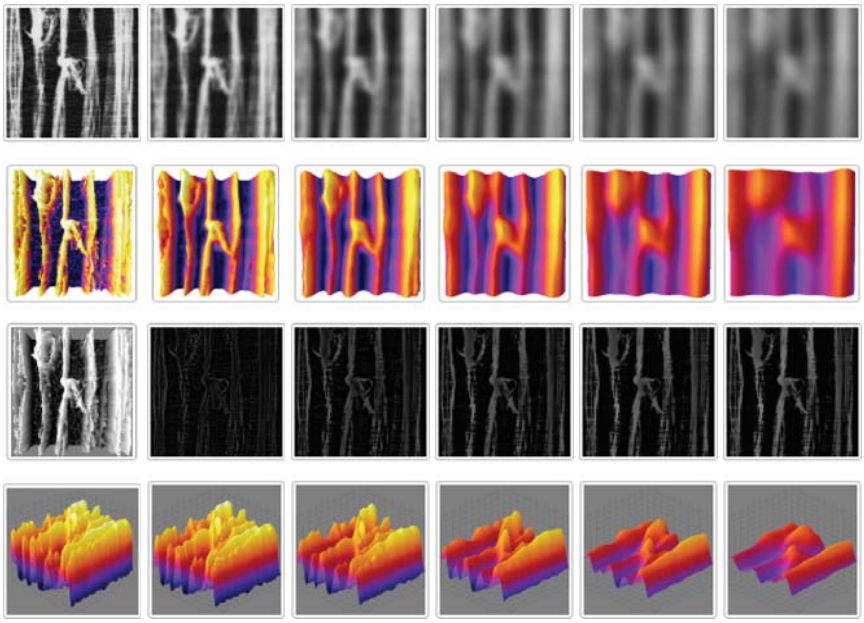
$$c_{mn}(x,y) = I(x,y) * g_{mn}(x,y) \quad (6)$$

The feature vector $\psi = [E_{11}, E_{12}, \ldots, E_{KS}]$ is finally obtained by computing the energy of the filtered images according to the Equation (7).

$$E_{mn} = \sum_{x,y} [c_{mn}(x,y)]^2 \quad (7)$$

3. An Approach to Texture Analysis

A widely strategy used to compute texture features with Gabor is to construct a bank of filters with different scales and orientations parameters. For each Gabor space is extracted statistical measures, such as energy and entropy. Instead of obtaining right the Gabor space, an original image (f) is decomposed in a set of derived images with



(a) I_0 (b) t_1 (c) t_2 (d) t_3 (e) t_4 (f) t_5

Figure 1. The essential idea with a scale-space representation of a image is to create a family of cartoon approximations. This figure shows an initial image I_0 (a) that has been successively smoothed with anisotropic diffusion [(b)–(f)]. The family of derived images may be viewed as the solution of the heat conduction, depicted in rows 2 and 4. The third row corresponds to the texture component.

anisotropic diffusion of Perona and Malik, described in Section 2.1. This procedure is executed with several levels of decomposition (t) in order to evidence high frequencies (v), while it preserves important structures such as edges. At each level (t), we obtain two components: cartoon approximation (u) and texture (v). The texture component is achieved by subtracting the original image and the cartoon approximation. An example of image decomposition using anisotropic diffusion is shonw in Figure 2.

The filtering process aim at evidencing high frequencies in the image in order to produce richer representations. Perona and Malik filtering overcomes the main restriction imposed by linear approaches, i.e., blur in region boundaries does not occur. The set of texture images v is then used to extract Gabor features and useful for a variety of tasks, for example, texture classification. The diagram of Figure 3 summarizes the approach proposed here.

4. Experimental Evaluation

In order to evaluate our approach, experiments are performed on two image datasets. First, the datasets used for evaluation are described. Then, implementation details of the descriptors and classifiers are discussed. Finally, the results are shown.

4.1. Datasets

The Brodatz album [4] is the most known benchmark for evaluating texture methods. Each class is composed by one image divided into nine new samples non-overlapped. These images have 200×200 pixels with 256 gray levels. A total of 100 texture classes with 10 images per class was used. Recently, the Brodatz dataset has been criticized for certain weaknesses, including lack of viewpoint and scale variation, and illumination changes. Thus, we also use the Vistex dataset.

The Vision Texture dataset [16] (or Vistex) contains a large set of natural colorful textures taken under several scale and illumination conditions. In addition, images are

(a) Original ($f = u + v$) (b) Cartoon (u) (c) Texture (v)

Figure 2. An example of image decomposition for the Barbara image (a). At each level of decomposition, it is generated a cartoon approximation u and a texture component v. v is obtained by subtracting the original image and the cartoon approximation.

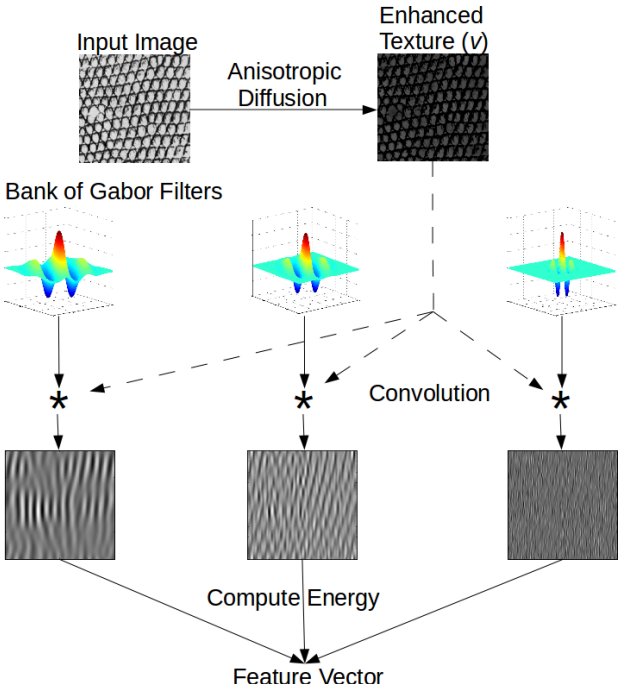


Figure 3. Our approach for texture analysis.

acquired with different cameras. For this dataset we use a total of 50 texture classes in gray scale. The size of the original images was 512×512, but we use the same number of samples as [18]. Each texture were split into 128×128 pixel images, with 16 sub-samples per class, totalizing 800 images.

4.2. Performance Evaluation

In the experiments, we compute the energy of Gabor filters with 8 orientations and 5 scales, resulting a feature vector with 40 dimensions. We adopt the K nearest-neighbor (K-NN) classifier, since it is a good reference classification method in the texture recognition. A initial value of $K = 5$ is used, with 10-fold cross validation and Euclidean similarity measure. Here, we change the levels of decomposition t of the anisotropic diffusion process (*scales*). The decomposition ranges from 10 to 200. The approach is evaluated using two texture datasets.

Experiment 1: First, we evaluate our approach on the Brodatz dataset and compare it to the original Gabor features. Features are computed with different levels of decomposition t. Figure 4 shows the classification rates in the y axis, while the levels of decomposition are indicated in the x axis. It can be observed that enhanced texture component (v), extracted using our approach, performs better than the original Gabor method. The highest classification rate ($t = 40$) is 94.29% for texture (v) and 91% for the original Gabor, respectively. Note that the performance of the cartoon approximations (u) get worst at each level of decomposition, which confirms our hypothesis that the component u can be discarded in order to improve the classification rate.

Experiment 2: In this experiment we evaluate our approach on the Vistex dataset. The setting for this experiment is the same as the previous one. In Figure 5, the classification rates are presented in the y axis, while the decompositions are

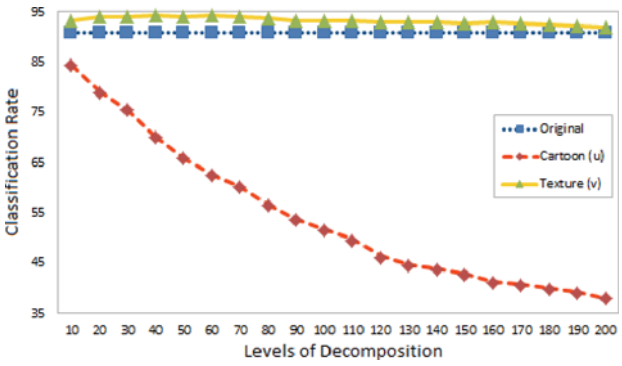


Figure 4. Comparison of different scales on the Brodatz dataset.

presented in the x axis. Our approach achieves the best performance with 88.96% ($t = 140$) against 83.66% for original Gabor. It is worth noting that the classification rates for the cartoon (u) component reduce at each iteration. This is associated to the gradual decomposition on the image.

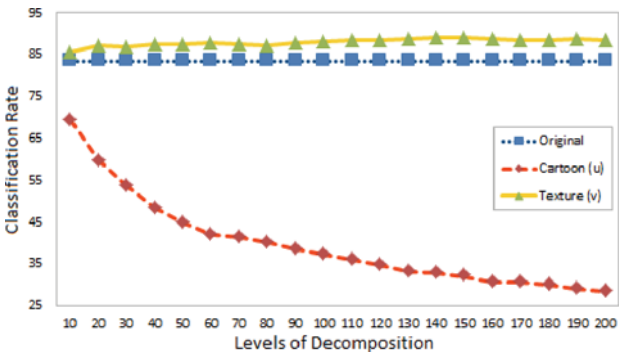


Figure 5. Comparison of different scales on the Vistex dataset.

Table 1 presents the average and standard deviation in terms of classification rates. It also shows results for $K = \{3, 5, 7\}$ on the original image (f), cartoon approximation (u) and texture (v). As we can see, our approach using the texture component (v) outperforms the others for all values of K on both datasets. Interesting results came out from the cartoon approximation experiments, which is discarded in the proposed approach. A classification rate of 67.80% and 31.16% are obtained on the Brodatz dataset and Vistex dataset, respectively. It clearly shows the poor classification power using the cartoon approximation.

Dataset	Component	%(3-NN)	%(5-NN)	%(7-NN)
Brodatz	(f) Original	92.53	91.00	89.04
	(u) Cartoon	71.61	70.06	68.96
	(v) Texture	94.88	94.29	92.87
	$(v - f)$	2.35	3.29	**3.83**
Vistex	(f) Original	84.71	83.66	83.10
	(u) Cartoon	31.85	32.95	32.35
	(v) Texture	89.21	88.96	86.65
	$(v - f)$	4.50	**5.30**	3.55

Table 1. Comparison of different values of nearest neighbors on both datasets.

To illustrate the potential of our approach, we compare it with three representative operators used for filtering edges: *Gaussian*, *Laplacian* and *Laplacian of Gaussian* (LoG) (we refer to [10] for more details). For all operators, the same procedure of the proposed approach was performed. In this setting, our approach achieved the highest classification rates for all values of K on both datasets. For the Brodatz dataset, an improvement of 3.14% compared to the Gaussian operator was obtained using $K = 5$. On the Vistex dataset with $K = 5$, our approach achieved a classification rate of 88.96%, which is significantly better than the classification rate of 83.63% achieved by the LoG operator. Experimental results demonstrate that our approach is an effective representation for texture modeling.

Dataset	Operator + Gabor	%(3-NN)	%(5-NN)	%(7-NN)
Brodatz	Gaussian	92.55	91.15	89.77
	Laplacian	91.17	89.49	87.93
	LoG	92.78	90.42	89.45
	Our approach	**94.88**	**94.29**	**92.87**
Vistex	Gaussian	85.14	82.75	81.72
	Laplacian	84.56	82.71	81.05
	LoG	85.24	83.63	82.11
	Our approach	**89.21**	**88.96**	**86.65**

Table 2. Comparison of different image operators on both datasets.

5. Leaf-Texture Enhancement: A Case Study

Although there exist some tools interested in identifying plant species, amost none of them are concerned in enhancing the texture attribute before computing features from images. Here, we show a case study using a subset of five classes, with 10 images per class. One example of each class is shown in Figure 6. Again our approach achieved highest classification rates, according to Table 3. The results show that our approach is consistent, being a useful method to enhance the texture attribute employed in real-world applications.

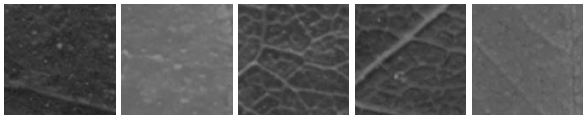


Figure 6. Leaf samples.

Operator + Gabor	%(3-NN)	%(5-NN)	%(7-NN)
Original	76.60	75.60	74.80
Gaussian	81.40	73.65	72.40
Laplacian	73.80	71.80	70.80
LoG	74.20	73.00	71.40
Our Approach	**86.00**	**80.60**	**76.20**

Table 3. Comparison of different values of nearest neighbors on the leaf dataset.

6. Conclusions

This paper proposed a new approach to enhance the richness of the texture attribute by applying anisotropic diffusion as an early step in the texture image modeling. We have also demonstrated how the Gabor process can be improved by using our approach. Promising results have been obtained on two databases of high complexity. In the Brodatz dataset, experimental results indicate that the proposed approach improves classification rate from 89.04% to 92.87% over the traditional approach. In addition, experimental results on Vistex dataset demonstrated that the proposed approach provides an improvement of 5.30% on classification rate. Our approach is able to successfully handle a wide range of texture methods, e.g. from Gabor filters to Markov random fields. In order to evaluate our approach, we performed it to enhance leaf-texture textures wide used in systems of plant leaf identification. As part of the future work, we plan to focus on investigating new nonlinear PDEs and texture image methods.

Acknowledgments. The authors gratefully acknowledge the financial support of CNPq and FAPESP.

References

[1] R. Azencott, J.-P. Wang, and L. Younes. Texture classification using windowed fourier filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19:148–153, February 1997.

[2] A. R. Backes, D. Casanova, and O. M. Bruno. Plant leaf identification based on volumetric fractal dimension. *IJPRAI*, 23(6):1145–1160, 2009.

[3] A. R. Backes, W. N. Gonçalves, A. S. Martinez, and O. M. Bruno. Texture analysis and classification using deterministic tourist walk. *Pattern Recogn.*, 43:685–694, March 2010.

[4] P. Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover Publications, New York, 1966.

[5] O. M. Bruno, R. de Oliveira Plotze, M. Falvo, and M. de Castro. Fractal dimension applied to plant identification. *Information Sciences*, 178:2722–2733, June 2008.

[6] C. H. Chen and P.-G. Peter Ho. Statistical pattern recognition in remote sensing. *Pattern Recognition*, 41:2731–2741, September 2008.

[7] Y. Chen and E. Dougherty. Gray-scale morphological granulometric texture classification. *Optical Engineering*, 33(8):2713–2722, 1994.

[8] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.

[9] I. Daubechies. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.

[10] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall series in artificial intelligence. Prentice Hall, New Jersey, 2003.

[11] D. Gabor. Theory of communication. *Journal of Institute of Electronic Engineering*, 93:429–457, November 1946.

[12] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

[13] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621, 1973.

[14] R. L. Kashyap and A. Khotanzad. A model-based method for rotation invariant texture classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8:472–481, June 1986.

[15] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, August 1984.

[16] M. M. Lab. Vision texture – vistex database, 1995.

[17] T. Lindeberg. Scale-space. In B. Wah, editor, *Encyclopedia of Computer Science and Engineering*, volume 4 of *EncycloCSE08*, pages 2495–2504, Hoboken, New Jersey, USA, September 2008. John Wiley and Sons.

[18] T. Mäenpää and M. Pietikäinen. Classification with color and texture: jointly or separately? *Pattern Recognition*, 37(8):1629–1640, 2004.

[19] B. B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Company, New York, August 1983.

[20] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, July 1990.

[21] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA, 1983.

[22] A. P. Witkin. Scale-space filtering. *International Joint Conference on Artificial Intelligence*, pages 1019–1022, 1983.

[23] J. Zhang and T. Tan. Brief review of invariant texture analysis methods. *Pattern Recognition*, 35(3):735–747, March 2002.

Segmentação de Íris utilizando Bag of Keypoints

Fabio Nascimento Brandão, Prof. Dr. Maurício Marengoni
Universidade Presbiteriana Mackenzie
Rua da Consolação, 930
São Paulo - SP - Brasil
fabio.nb@gmail.com, mmarengoni@mackenzie.br

Abstract

This paper presents an ongoing research that uses a Bag of Keypoints technique for the iris location in images. The idea is to use SURF points and learn how to identify points inside the iris region based on SURF features. Once the points inside the iris are selected they are used for region growing and grouping, such that the whole iris is segmented at the end. This is the first step performed in human recognition based on iris. The paper presents the basic concepts involved and preliminary results. The technique seems promising and can be easily adapted to different region types.

1. Introdução

A biometria é uma área de estudo que utiliza as características físicas e comportamentais das pessoas para o seu reconhecimento. Os sistemas de identificação biométrica atuais estão ficando cada vez mais populares devido às suas características como confiança, velocidade de execução, conforto de utilização e dificuldade de falsificação. [20]

Existem diversas características que podem ser utilizadas no reconhecimento de uma pessoa, entre elas estão: impressão digital, voz, modo de andar, face e íris. O reconhecimento através da íris se destaca pela dificuldade de se alterar acidentalmente ou intencionalmente as suas características, já que ela é um órgão protegido dentro do olho humano. Além disso, o padrão da íris se mantém inalterado durante toda a vida de uma pessoa e o seu padrão é único. [7]

Qualquer sistema de reconhecimento de íris segue basicamente os mesmos passos: aquisição da imagem, segmentação da íris na imagem, extração das características (isto é, obter os padrões da íris) e comparação das características extraídas com uma base de dados. O passo mais crítico entre os descritos é a segmentação, sendo que a segmentação incorreta da íris pode ocasionar o reconhecimento incorreto de uma pessoa. Os problemas que surgem durante esta etapa são: grande variabilidade do posicionamento da íris na imagem, bem como grande variabilidade do tamanho da íris; presença de oclusões (causadas pelos cabelos, cílios, pálpebras e óculos); borrões na imagem; problemas no foco; reflexos especulares e ângulos de visão oblíquos. [11]

Na Figura 1 **(a)**, vemos a imagem original e na Figura 1 **(b)** vemos a imagem segmentada. Esta imagem possui alguns tipos de ruídos: oclusões causadas pelos cílios e pálpebras e reflexos especulares. Além disso, a segmentação foi feita de maneira incorreta, pois não detectou corretamente a localização da íris e o seu tamanho.

Figura 1. Exemplo de segmentação incorreta, com diversos tipos de ruído

Os algoritmos geralmente são projetados para aquisição da imagem de forma cooperativa, isto é, o usuário que será reconhecido facilita o processo de captura da imagem de sua íris (ex.: usuário se mantém parado na frente da câmera). Os maiores problemas aparecem quando a aquisição da imagem é feita de forma não cooperativa, onde o usuário tem pouca ou quase nenhuma participação no processo de captura da imagem. Com isso, as imagens capturadas apresentam muitos ruídos, dificultando o processamento da imagem e, conseqüentemente, o reconhecimento da pessoa. [23]

O algoritmo utilizado na segmentação da íris deve ser rápido e robusto. A robustez está relacionada com a capacidade do algoritmo em lidar com diversos tipos de ruídos, já que o desempenho global do sistema de reconhecimento de íris está diretamente ligado com o desempenho do algoritmo de segmentação (pois é nesta etapa que o algoritmo extrai de maneira eficaz a região da imagem associada exclusivamente com a íris)[15].

A segmentaçao de íris é um problema com um alto grau de complexidade, sendo que, no período de Julho de 2007 à Dezembro de 2008 uma disputa de segmentação de íris chamada NICE.I[1] foi realizada para avaliar o desempenho de diversos algoritmos de segmentaçao de íris enviados pelos participantes. Boa parte das imagens das íris foram tiradas com close-up e em um ambiente com condições normais (onde a pessoa que está sendo fotografada coopera no processode aquisição da imagem, a iluminação utilizada no ambiente é adequada para o tipo de tarefa sendo executada, etc) e, além disso, as íris das imagens poderiam estar obstruídas por qualquer tipo de ruído (ex.: obstrução causada pelos cabelos, cílios, óculos, etc).

A base de imagens utilizada no NICE.I foi a UBIRIS.v2, que foi uma nova versão da base de imagens UBIRIS.v1 [17]. Esta base foi feita devido às diversas críticas sobre o realismo dos ruídos das imagens que foram descritos na literatura, sendo que esta base possui mais de 11000 imagens. As imagens foram obtidas em um ambiente sem restrição (a uma distância entre 3 e 10 metros, com o indivíduo em movimento e com ondas visíveis [luzes que o olho do ser humano é capaz de ver]), que corresponde a um ambiente mais real. Na Figura 2 podem ser vistos diversos tipos de oclusões e ruídos que podem ocorrer no processo de captura da imagem da íris.

A pesquisa apresentada neste trabalho procura responder às seguintes questões:

- É possível utilizar pontos característicos do tipo SUF para aprender a localizar objetos de interesse do tipo íris?
- É possível utilizar estes pontos aprendidos e fazer um agrupamento ou crescimento de região para segmentar uma região de interesse?

Este trabalho tem como objetivo principal a implementação das técnicas apresentadas para responder as questões levantadas. O trabalho irá utilizar o algoritmo SURF para detectar e descrever os pontos característicos das imagens, que servirá de entrada para a geraçao de um vocabulário e, após estes passos, será treinado um classificador SVM para a localizaçao das íris nas imagens. Com a íris detectada, a técnica de *Region Growing* será utilizada para que a íris seja segmentada. Serão apresentados ainda resultados preliminares obtidos nos experimentos já realizados.

Figura 2. Exemplos de oclusões e ruídos, onde podem ser vistas oclusões causadas por cílios, pálpebras, óculos e cabelos, reflexos especulares e imagem capturada com ângulo de visão oblíquo

2. Revisão Bibliográfica

2.1. Detecção e Representação de Pontos Característicos

Existem diversos algoritmos capazes de detectar e representar características de imagens, entre eles podemos citar o SIFT e o SURF [26]. Basicamente o algoritmo SIFT procura pontos de interesse na imagem e os representa através de um conjunto de valores numéricos que não levam em conta as mudanças na escala, iluminação e outras transformações geométricas.

Estes pontos são chamados de pontos característicos, ou pontos invariantes.

O SIFT[14] tem sido utilizado para o reconhecimento de objetos em ambientes onde podem existir oclusões parciais [18], localização de robôs e mapeamento de lugares [21], fotos panorâmicas, modelagem 3D, rastreamento, reconhecimento de gestos humanos, entre outros.

A idéia do algoritmo SURF é semelhante à do SIFT, porém o SURF é um algoritmo de menor complexidade. De acordo com seus autores o algoritmo SURF é mais rápido e mais robusto em relação a transformações geométricas do que o SIFT [4].

2.2. Dicionário de Pontos Característicos

No artigo de G. Csurka et al. [10] foi proposta uma nova metodologia de categorização visual chamada Dicionário de Pontos Característicos, ou "Bag of Keypoints". Esta nova metodologia é baseada em uma técnica de categorização de texto chamada "Bag of Words" [25]. A idéia de se adaptar técnicas de categorização de texto na área de visão computacional não é nova, Zhu et al [28] investigaram a quantização vetorial de pequenas imagens quadradas, que foram chamadas de blocos chaves. Foi mostrado que a quantização vetorial produz mais resultados semânticos que os métodos baseados na análise de cores e texturas.

O método descrito no artigo de G. Csurka [10] possui quatro passos fundamentais:

- Detecção e descrição de trechos da imagem
- Atribuição de descritores de trechos aos grupos pré-determinados (vocabulário) através de um algoritmo de quantização vetorial
- Construir um conjunto de *Keypoints*, que sumariza o número de trechos associados aos grupos
- Aplicar um classificador de múltiplas classes, considerando o *Bag of Keypoints* como um vetor de características, determinando em quais categorias a imagem deve ser classificada

Os descritores obtidos no primeiro passo devem ser invariantes às condições que são irrelevantes para a tarefa de categorização (transformações nas imagens, como por exemplo a rotação e a translação, e também variações de iluminação e oclusões parciais) mas rico o suficiente em informações que discriminem a categoria na qual o trecho analisado pertence. O segundo passo deve conter um vocabulário do tamanho ideal para distinguir mudanças relevantes nas partes das imagens e não distinguir mudanças irrelevantes que podem ocorrer nas imagens como por exemplo ruídos.

O objetivo principal deste algoritmo é utilizar um vocabulário (ou conjunto de características) que tenha um bom desempenho de categorização em um determinado conjunto de dados de treinamento. Os passos envolvidos no treinamento são:

- Detectar e descrever os trechos da imagem de um conjunto de imagens de treinamento
- Construir um conjunto de vocabulários: cada um sendo o centro de um grupo, em relação aos descritores
- Extração dos *bags of words* desses vocabulários
- Treinamento dos classificadores de múltiplas classes (por ex.: SVM) usando os *bags of words* como vetores de entrada
- Seleção do vocabulário e classificador com o melhor desempenho

Os resultados do trabalho de G. Csurka [10] mostraram que essa nova técnica se mostrou eficaz já que ela possui robustez diante de imagens que possuam um fundo complexo, e obteve um alto desempenho na categorização sem explorar as informações geométricas contidas nas imagens.

No trabalho realizado por D. Filliat [13] a técnica *Bag of Keypoints* foi utilizada no desenvolvimento de um sistema de localização e aprendizagem de mapas feito por robô. De acordo com o artigo esta técnica se mostrou confiável para reconhecer em que ambiente o robô estava após um curto período de tempo e, além disso, a técnica se mostrou estável após um longo tempo de uso.

2.3. Classificador SVM

O SVM (support vector machine ou support vector network) [9] é uma máquina de aprendizagem para problemas de classificação entre dois grupos. O algoritmo funciona mapeando um vetor de entrada definido em um espaço de muitas dimensões, onde é criada uma superfície para separar os vetores em dois grupos. Através de técnicas de classificação binária o SVM pode ser utilizado como um classificador de multi-classes [12].

2.4. Segmentação

A segmentação é o processo de divisão de uma imagem em grupos de pixels que simplificam a representação da imagem, facilitando assim a sua análise. A segmentação geralmente é utilizada como um pré processamento na localização de objetos e contornos (linhas, curvas ou outras formas geométricas) em imagens. Na figura 3 são exibidas diversas segmentações de íris, com baixa taxa de erro na área segmentada, obtidas no NICE.I.

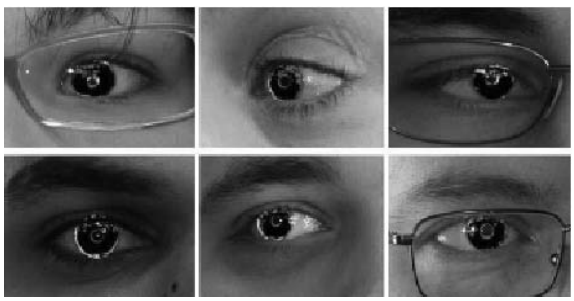


Figura 3. Exemplo de segmentações com baixas taxas de erro no NICE.I [1]

Existem diversas aplicações onde a segmentação de imagem é utilizada como um pré processamento,tais como:

- Imagens médicas
  - Localização de tumores ou outras patologias [8]
  - Medição de volume de tecidos
  - Cirurgia feita por computadores
  - Diagnósticos
  - Estudo da estrutura anatômica
- Localização de objetos em imagens de satélites (estradas, florestas, etc.) [22]
- Reconhecimento de face [2], impressão digital [6], íris [3]

Diversas técnicas têm sido utilizadas no processo de segmentação de imagens já que não há uma solução única ou geral para o problema de segmentação e, geralmente várias técnicas de filtragem e binarização são combinadas para resolver, de maneira eficaz, um problema específico.

Entre as técnicas mais utilizadas estão: clusterização [5], técnicas baseadas em histograma [24], detecção de arestas [19], crescimento de região [16], particionamento de grafos [27] e watershed [19].

A técnica que será utilizada neste trabalho é a de crescimento de região, pois ela é extremamente simples de implementar. O princípio básico desta técnica é colocar uma ou mais sementes em uma imagem e fazer com que elas "cresçam" de forma a agrupar pixels com alguma característica homogênea, criando assim uma ou mais regiões na imagem. Estas sementes são escolhidas com algum critério, como por exemplo, o valor de intensidade do pixel em uma imagem em nível de cinza ou até mesmo a posição de um pixel em uma imagem. [16]

Após a escolha das sementes, o conjunto de pixels ao redor de cada semente é analisado de acordo com um critério de vizinhança (geralmente 4 ou 8) para ver ser estes pixels pertencem ou não à região. Com isso, ocorre o crescimento da região até que todas as regiões da imagem não cresçam mais.

As vantagens desta técnica são: separação correta de regiões que possuam as mesmas propriedades e simplicidade, pois ela se baseia na comparação de pixels com as regiões formadas. Ruídos na imagem e alta variação na intensidade podem resultar em buracos no resultado da imagem, o que é uma das principais desvantagens desta técnica. Um exemplo de segmentação de imagem pode ser vista na Figura 4.

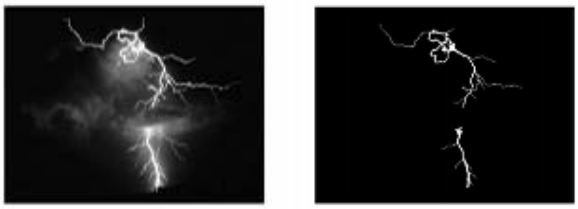


Figura 4. Crescimento de região utilizando 10 sementes colocadas nos pixels com níveis mais próximos ao branco

3. Metodologia Proposta

As imagens que foram utilizadas para os testes e treinamento dos algoritmos enviados ao NICE.I serão utilizadas no processo de aprendizagem e validação do algoritmo proposto. Os algoritmos enviados ao NICE.I devem processar a imagem de um olho e retornar uma imagem de mesma dimensão, onde os pixels pretos representam as regiões da íris como pode ser visto na Figura 5.

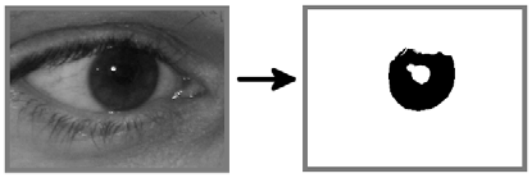


Figura 5. Exemplo da segmentação desejada

A taxa de erro da classificação do algoritmo para a imagem I_i é dada pela proporção dos pixels classificados incorretamente sobre a quantidade de pixels de toda a imagem como pode ser visto na equação 1.

$$E_i = \frac{1}{c \times r} \sum_{c'} \sum_{r'} O(c', r') \otimes C(c', r') \qquad (1)$$

Onde $O(c', r')$ e $C(c', r')$ são respectivamente os pixels da saída do algoritmo e os pixels da imagem esperada.

A taxa de erro da classificação E do algoritmo é a média dos erros E_i das imagens de entrada como pode ser visto na equação 2.

$$E = \frac{1}{n} \sum_{i} E_i \qquad (2)$$

O valor E está limitado no intervalo entre 0 e 1 e foi utilizado no NICE.I para classificar o desempenho dos algoritmos enviado à disputa. Neste contexto, 1 e 0 significam respectivamente o pior e o melhor resultado.

A segunda medida de erro utilizada se destina a compensar a desproporção entre os pixels que representam a íris e os pixels que não representam íris nas imagens.

Essa taxa de erro para a imagem E_i é dada através da média das taxas entre os falsos positivos **(FP)** e os falsos negativos **(FN)** e é representada pela equação 3.

$$E_i = \frac{FP + FN}{2} \quad (3)$$

Similarmente à taxa de erro E^1, a taxa de erro E^2 é calculada através da média dos erros E_i das imagens de entrada.

As taxas de erro E obtidas pelos vencedores do NICE.I ficaram entre 0,0131 e 0,0305.

O primeiro passo do treinamento do algoritmo consiste em aplicar o algoritmo de detecção e representação de features SURF para obter os pontos de interesse das imagens. Esses pontos serão utilizados para criar o vocabulário do algoritmo *Bag of Keypoints*. Esse vocabulário é formado através da clusterização dos pontos através do algoritmo k-means. Neste projeto foram gerados 50 grupos de pontos que pertencem à íris e 50 grupos de pontos que não pertencem à íris.

Com o vocabulário formado basta construir o *Bag of Keypoints* das regiões das imagens (regiões obtidas através do particionamento da imagem em grids de 20x15). O *Bag of Keypoints* de cada grid consiste em um vetor do tamanho do vocabulário, onde cada posição desse vetor é a quantidade de pontos da grid que mais se parecem com a keyword relacionada. As *Bag of Keypoints* obtidas são então utilizadas como entrada do treinamento do classificador SVM.

O processo de treinamento do algoritmo pode ser representado pela Figura 6.

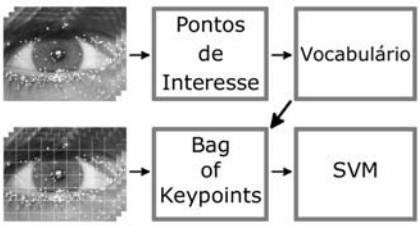


Figura 6. Processo de treinamento do algoritmo

O processo para a segmentação da íris consiste em separar a imagem em grids, obter o *Bag of Keypoints* de cada grid e verificar com o classificador SVM treinado se a grid possui ou não uma íris. Se a grid possuir íris, o ponto dessa grid que mais se parecer com um dos 50 pontos do vocabulário de íris é utilizado como entrada para o algoritmo de crescimento de região. Este processo pode ser representado pela Figura 7.

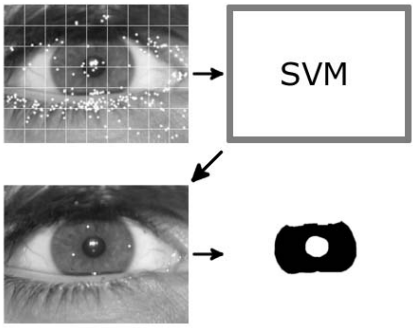


Figura 7. Processo de segmentaçao do algoritmo

O treinamento da técnica proposta no artigo utiliza 100 imagens de testes do NICE.I, enquanto a validação utiliza aproximadamente 400 imagens. Essas imagens foram fornecidas pelo NICE.I e, para cada imagem de teste existe uma imagem com a máscara onde a íris está posicionada. Essa máscara foi gerada pelos organizadores do NICE.I manualmente e são consideradas como a segmentação correta da íris.

4. Resultados

Até o momento de escrita deste artigo todo o processo de extração de features das imagens, geração do vocabulário, criação da *Bag of Keypoints*, treinamento do classificador e busca de áreas na imagem que possuim íris já foi implementada. O algoritmo de crescimento de região ainda não foi implementado, portanto não temos o resultado da eficácia da técnica como um todo.

O processo de busca de grids na imagem que possuem íris está com uma taxa de erro de aproximadamente 37%, sendo que o processo não está conseguindo encontrar íris em pelo menos 0.5% das imagens de validação.

5. Conclusões

A técnica de *Bag of Keypoints* se mostra muito boa para classificar diferentes tipos de imagens. Neste artigo a utilização dela para encontrar a íris em regiões das imagens se mostra promissora, mas por algum motivo os resultados esperados não estão aparecendo.

O próximo passo é verificar o motivo pelo qual a classificação das regiões entre região com íris ou sem íris está sendo feita de forma incorreta. Após isto devemos implementar algum tipo de escolha de ponto dentro da região classificada como íris para que este ponto seja

utilizado como uma semente no algoritmo de crescimento de região.

Referências

[1] Nice.i - noisy iris challenge evaluation, part i - http://nice1.di.ubi.pt/.

[2] B. B. Amor, M. Ardabilian, and L. Chen. Enhancing 3d face recognition by mimics segmentation. *Intelligent Systems Design and Applications, International Conference on*, 3:150–155, 2006.

[3] G. Annapoorani, R. Krishnamoorthi, and P. G. Jeya. Accurate and fast iris segmentation, 2010.

[4] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 – 359, 2008. Similarity Matching in Computer Vision and Multimedia.

[5] A. M. Bensaid, L. O. Hall, J. C. Bezdek, and L. P. Clarke. Partially supervised clustering for image segmentation. 29(5):859–871, May 1996.

[6] S. Bernard, N. Boujemaa, D. Vitale, and C. Bricot. Fingerprint segmentation using the phase of multiscale gabor wavelets, 2002.

[7] Y. Chen, M. Adjouadi, C. Han, J. Wang, A. Barreto, N. Rishe, and J. Andrian. A highly accurate and computationally efficient approach for unconstrained iris segmentation. *Image and Vision Computing*, 28(2):261–269, 2010.

[8] M. C. Clark, L. O. Hall, D. B. Goldgof, R. Velthuizen, F. Reed, and M. S. Silbiger. Automatic tumor segmentation using knowledge-based techniques. *IEEE Transactions on Medical Imaging*, 17:187–201, 1998.

[9] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995. 10.1023/A:1022627411411.

[10] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[11] R. Donida Labati and F. Scotti. Noisy iris segmentation with boundary regularization and reflections removal. *Image and Vision Computing*, 28(2):270–277, 2010.

[12] K.-B. Duan and S. S. Keerthi. Which is the best multiclass svm method? an empirical study. In N. C. Oza, R. Polikar, J. Kittler, and F. Roli, editors, *Multiple Classifier Systems*, volume 3541 of *Lecture Notes in Computer Science*, pages 732–760. Springer Berlin / Heidelberg, 2005.

[13] D. Filliat. A visual bag of words method for interactive qualitative localization and mapping. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2007.

[14] D. G. Lowe. Object Recognition from Local Scale-Invariant Features. *Computer Vision, IEEE International Conference on*, 2:1150–1157 vol.2, August 1999.

[15] M. A. Luengo-Oroz, E. Faure, and J. Angulo. Robust iris segmentation on uncalibrated noisy images using mathematical morphology. *Image and Vision Computing*, 28(2):278–284, 2010.

[16] M. Mancas, B. Gosselin, and B. Macq. Segmentation using a region-growing thresholding. In *Proceedings of the SPIE 5672 (2005) 388–398*, pages 12–13.

[17] H. Proenca and L. Alexandre. UBIRIS: A noisy iris image database. In *13th International Conference on Image Analysis and Processing - ICIAP 2005*, volume LNCS 3617, pages 970–977, Cagliari, Italy, September 2005. Springer.

[18] A. Ramisa, S. Vasudevan, D. Aldavert, R. Toledo, and R. Lopez de Mantaras. Evaluation of the sift object recognition method in mobile robots. In *Proceeding of the 2009 conference on Artificial Intelligence Research and Development: Proceedings of the 12th International Conference of the Catalan Association for Artificial Intelligence*, pages 9–18, Amsterdam, The Netherlands, The Netherlands, 2009. IOS Press.

[19] N. Salman. Image segmentation based on watershed and edge detection techniques. *Int. Arab J. Inf. Technol.*, 3(2):104–110, 2006.

[20] W. Sankowski, K. Grabowski, M. Napieralska, M. Zubert, and A. Napieralski. Reliable algorithm for iris segmentation in eye image. *Image and Vision Computing*, 28(2):231–237, 2010.

[21] S. Se, D. Lowe, and J. Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research*, 21:735–758, 2002.

[22] J. Senthilnath, M. Rajeshwari, and S. Omkar. Automatic road extraction using high resolution satellite image based on texture progressive analysis and normalized cut method. *Journal of the Indian Society of Remote Sensing*, 37:351–361, 2009. 10.1007/s12524-009-0043-5.

[23] T. Tan, Z. He, and Z. Sun. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image and Vision Computing*, 28(2):223–230, 2010.

[24] O. J. Tobias and R. Seara. Image segmentation by histogram thresholding using fuzzy sets. *IEEE Transactions on Image Processing*, 11(12):1457–1465, 2002.

[25] S. Tong and D. Koller. Support Vector Machine Active Learning with Applications to Text Classification. In P. Langley, editor, *Proceedings of ICML-00, 17th International Conference on Machine Learning*, pages 999–1006. Morgan Kaufmann Publishers, San Francisco, US, 2000.

[26] C. Valgren and A. J. Lilienthal. SIFT, SURF and seasons: Long-term outdoor localization using local features. In *Proceedings of the European Conference on Mobile Robots (ECMR)*, pages 253–258, September 19–21 2007.

[27] S. X. Yu, R. Gross, and J. Shi. Concurrent object recognition and segmentation by graph partitioning. In *in NIPS*, pages 1383–1390. MIT Press, 2002.

[28] L. Zhu, A. Rao, and A. Zhang. Theory of keyblock-based image retrieval. *ACM Trans. Inf. Syst*, 20:224–257, 2002.

Influência da padronização do sistema de cor RGB nos métodos de visão computacional

Dalcimar Casanova
Universidade de São Paulo
Instituto de Física de São Carlos
São Carlos - SP
dalcimar@gmail.com

Maurício Falvo
Universidade de São Paulo
Instituto de Física de São Carlos
São Carlos - SP
Faculdade Adventista de Hortolândia
Hortolândia - SP
mauricio.falvo@gmail.com

Odemir M. Bruno
Universidade de São Paulo
Instituto de Física de São Carlos
São Carlos - SP
bruno@ifsc.usp.br

Resumo

A aquisição de imagens é uma das principais etapas em um sistema de visão computacional. Embora os dispositivos para tal tarefa (e.g. scanners) estejam muito desenvolvidos ainda há um sério problema de padronização. Esta falta de padronização pode acarretar em sérios problemas, especialmente quando se trabalha com imagens biológicas que contém muito detalhes. O objetivo desta pesquisa é, portanto, investigar a variação da captura dos dispositivos de aquisição e medir o quanto o sistema de gerenciamento de cores pode minimizar essa influência sobre os métodos de visão computacional.

Resumo

The Image acquisition is a main task in computer vision systems. Although the devices for this task (e.g. scanners) are highly developed there is still a serious problem of standardization, especially regarding the capture of colors. This lack of standardization can lead to serious problems, especially when working with biological images that containing more detail. The objective of this research is therefore, investigate the variation of the capture of acquisition devices and how much the color management system can minimize this influence on methods of computer vision.

1. Introdução

Das diversas etapas de um processamento voltado à visão computacional, a aquisição de imagens é um dos pontos mais críticos de todo o processo. O uso de equipamentos de baixa qualidade ou de metodologias erradas, podem comprometer o sucesso das demais etapas envolvidas. Quando o objeto de estudo trata de imagens biológicas, cujos detalhes muitas vezes são de extrema importância, esta etapa de aquisição adquire ainda maior importância. Este é o caso da área de pesquisa de métodos de classificação de plantas baseados em imagens foliar, que atualmente vem ganhando uma atenção crescente [1, 2, 6].

Nessa área de pesquisa há três principais abordagens, análise de forma (ou contorno) [1, 2, 6], análise da estrutura de venação [3] e análise de textura [9]. Em trabalho recente (e.g. [9]) os autores demostram que a análise de textura é um poderoso meio discriminante de espécies, entretanto este meio possui uma grande dependência do processo de aquisição.

Em projetos de grandes proporções as coletas e digitalizações normalmente são realizadas em pontos geográficos distintos. Mesmo que realizadas por botânicos que utilizam metodologias de coleta padronizadas pouco preocupa-se com a qualidade na digitalização, considerando a influência das marcas e modelos dos equipamentos. Essa influência é presente em função da diversidade de mar-

cas e modelos e mesmo entre equipamentos de mesmo modelo, fato que compromete seriamente todo o processo de aquisição [10].

Assim, este artigo utiliza uma metodologia que permite a calibração de cores das imagens entre diferentes scanners e estuda sua influência, com relação aos erros de amostragem de cor, sobre os métodos de análise de textura. Também investiga qual o comprometimento na identificação de folhas em função dos diferentes dispositivos de digitalização. A qualidade desta metodologia é um procedimento chave para criação de um herbario digital padronizado, com imagens de qualidade confiável.

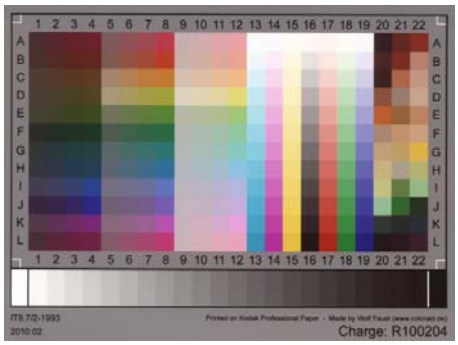


Figura 1. Imagem do cartão padrão IT 8.7/2 reflexivo

2. Materiais e métodos

2.1. IT8 target

Se duas cores possuem o mesmo CIE colorimétrico, então elas devem ser mesma cor se vistas sob as mesmas condições. Entretanto, em função das cores de uma imagem serem o resultado de um dispositivo de captura, estas sofrem a ação da variação de iluminação e dos próprios sensores. Faz-se necessário então uma correção para que a imagem capturada corresponda ao máximo ao objeto real que representa [10].

Atualmente a melhor maneira de garantir que se tenha a preservação das cores originais em uma imagem digital, é utilizar o sistema de gerenciamento de cores proposto pelo ICC(International Color Consortium) através de seus *colour profiles*. Esses profiles possuem a informação necessária para que o sistema de gerenciamento de cores compatibilize os dados entre dispositivos (e.g. scanners). Esses arquivos de profile são obtidos usando-se um cartão padrão de cores, o IT 8.7(ISO 12641). O cartão padrão IT 8.7 vem acompanhado com seu próprio arquivo de referência, o qual contém o exato valor esperado das cores que serão capturadas pelo scanner. Com a ajuda de um software específico é possível ler a imagem padrão capturada por um scanner e seu respectivo arquivo de referência e gerar um arquivo profile. Um bom software open-source que pode criar arquivos profiles é o SIPC (Scanner ICC Profile Creator) [7]. Esses cartões podem ser adquiridos de vários fornecedores, para este artigo foi utilizado o IT 8.7/2 (para dispositivos reflexivos) adquirido de http://www.targets.coloraid.de (Figura 1).

2.2. Dispositivios de digitalização

Para este trabalho foram utilizados dois scanners, de mesma marca e modelos distintos. O primeiro, HP G2710 (Scanner A) possui a capacidade de resolução ótica máxima de 2400dpi e profundidade máxima de cor de 48bits. O segundo, HP PSC1315 (Scanner B) possui resolução ótica máxima de 600dpi e 32bits de profundidade de cor. Para garantir garantir a qualidade da aquisição e a igualdade de parâmetros para efeitos de comparação, os dois scanners foram configurados para uma aquisição a 600 dpi e 24bits. Com relação ao formato das imagens geradas, foi utilizado o padrão PNG afim de que não houvesse perdas por compactação.

2.3. Material foliar

Como amostras para a realização do experimento, fora utilizadas cinco diferentes folhas de uma mesma espécie coletadas no campos de USP-São Carlos.

2.4. Padronização do sistema de cor

O experimento basicamente pode ser dividido em seis principais etapas. As duas primeiras (i. geração do profile do scanner; ii. geração da Estrutura de Transformação de Cores - ETC) visam obter o arquivo ETC utilizado para calibação das imagens que serão adquiridas. Deve-se notar que essa estrutura de correção é específica para o scanner no qual foi gerado, e no qual será realizada a digitalização das amostras.

As duas etapas seguintes (iii. digitalização das amostras foliares e iv. calibração das imagens) objetivam a aquisição e padronização das imagens das amostras. A sequência dessas ações resultam em dois conjuntos de imagens: um cujas cores foram calibradas e outro não.

Finalmente, as duas últimas etapas (v. extração do vetor de características das imagens não calibradas; vi. extração dos vetores de características das imagens calibradas) extraem características para a construção dos

respectivos vetores de cada amostra. Na finalização dessas etapas são obtidos dois conjuntos de vetores sendo que a única diferença entre esses é a origem dos dados (i.e calibradas e de não calibradas). Deve-se notar que essas etapas foram executadas duas vezes; uma para o scanner HP PSC1315 e outra para o scanner HP G2710.

A influência da padronização aplicada nas imagens é demonstrada pela comparação da variância entre os elementos dos vetores dos dois tipos de conjuntos gerados por um mesmo scanner e por scanners diferentes. A seguir cada par de etapas são descritas com mais detalhes, para o melhor entendimento do método aplicado.

2.4.1. Geração do profile e do ETC: Utilizado como padrão de calibração, o cartão IT 8.7/2 reflexivo foi digitalizado. A imagem obtida foi carregada juntamente com o arquivo de referência do cartão (IT8 File), pelo software SIPC (Scanner ICC Profile Creator), gerando seu respectivo arquivo de profile (Scanner profile). No MatLab, o arquivo de profile foi carregado e aplicado a função *makecform()* tendo como parâmetros: o próprio profile, e o padrão sRGB. O resultado obtido foi um arquivo contendo uma estrutura de transformação de cores, para um scanner específico, que permitirá calibrar as imagens adquiridas. Na Figura 2 é apresentado, de forma simplificada, o processo aqui descrito. Note que, para cada modelo de scanner foi obtido o seu respectivo arquivo ETC.

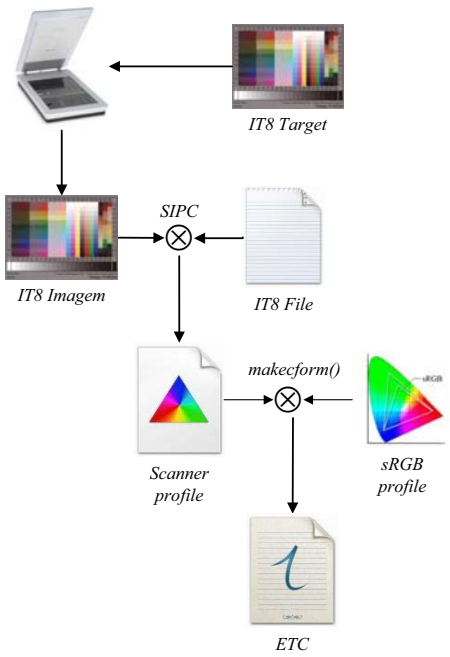


Figura 2. Processo de geração do arquivo de Estrutura de Transformação de Cores (ETC).

2.4.2. Aquisição e calibração das imagens: Cada amostra foliar foi digitalizada a uma resolução de 600dpi e 24bits de profundidade de cor. De cada imagem foi extraída cinco áreas distintas e não sobrepostas, com dimensões de 128x128 pixels. Esse procedimento foi realizado nos dois modelos de scanner citados neste trabalho, tomando-se o cuidado de que as cinco áreas extraídas fossem geograficamente as mesmas. Os dois conjuntos de imagens foram então carregados no MatLab e a função *applycform()* foi aplicada em cada imagem, utilizando como parâmetro o respectivo arquivo ETC do scanner de origem (Figura 3). Dessa forma ao final quatro conjuntos de imagens são obtidos: imagens corrigidas e não corrigidas obtidas pelo scanner HP PSC1315, imagens corrigidas e não corrigidas obtidas pelo scanner G2710. Das quais serão estraídos os vetores de características.

2.4.3. Extração dos Vetores de Caracterísiticas: Dois tipos de descritores foram testados para avaliar a influência da padronização do sistema de cor: Fourier e Gabor.

A extração dos descritores de Fourier [4] consistiu, para cada imagem, em dividir o espectro bidimensional em 64 setores utilizando 8 raios e 8 ângulos, totalizando 64 elementos que correspondem à energia obtida de cada setor.

Para a extração dos descritores de Gabor [5] cada imagem foi convoluida com uma família de filtros de Gabor que apresentavam diversas escalas e orientações. Neste experimento foram utilizadas 8 escalas e 8 rotações, totalizando 64 características, com frequência mínima e frequência máxima de 0.01 e 0.4, respectivamente. Os parâmetros individuais de cada filtro foram definidos de acordo com [8].

Para cada folha o vetor de características consiste na média dos descritores das 5 sub-texturas locais extraídas.

3. Resultados e discussão

O primeiro item observável em um processo de correção de cores é o aspecto visual das imagens transformadas. Como podemos observar pela Figura 4 as cores originais obtidas pelos referidos dispositivos tem um variação substancial (e.g. imagens A e C) e, como esperado, essa variação é minimizada após correção (e.g. imagens B e D). Embora não seja um resultado perfeito (i.e. um resultado perfeito seria se ambas as imagens B e D tivessem o mesmo aspecto de cor), essa

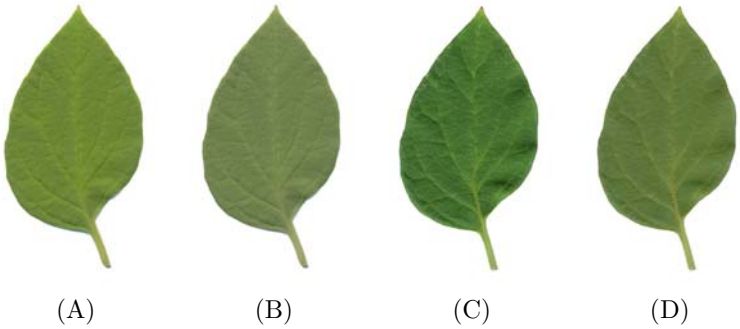


Figura 4. Exemplo de correção de cores sobre uma única folha nos 2 dispositivos testados. Observamos uma aproximação das tonalidades após correção. (A) Imagem original obtida pelo Scanner A. (B) Imagem A corrigida. (C) Imagem original obtida pelo Scanner B. (D) Imagem C corrigida.

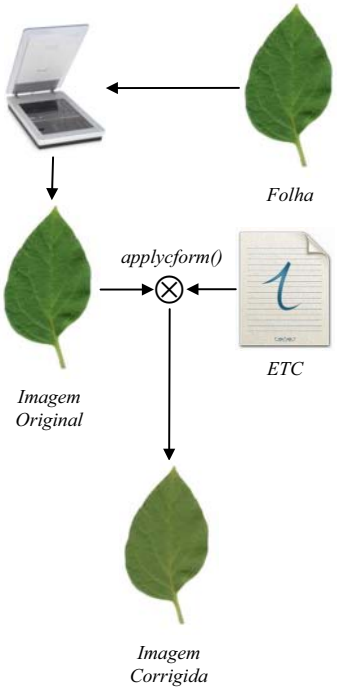


Figura 3. Processo de correção de imagens usando a ETC.

transformaçao pode ser considerada satisfatória.

Embora o aspecto visual seja uma importante fonte na avaliação da eficiência da metodologia da correção de cores, o que realmente queremos avaliar é se essa transformação minimiza a influência do dispositivo de aquisição nos métodos de visão computacional. Para tal, uma mesma imagem (folha) adquirida nos 2 dispositivos foi utilizada para extração de características com filtros de Gabor. Calculou-se a variância de cada descritor para os diferentes dispositivos (Figura 5, na linha azul). Paralelamente realizou-se a correção de cor dessa imagem e novas características com filtros de Gabor foram extraídas sobre a imagem transformada. Também calculou-se a variância de cada descritor entre os dispositivos (Figura 5, linha vermelha).

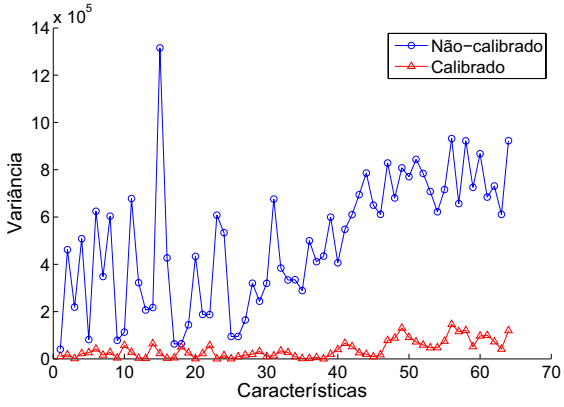


Figura 5. Variância observada de cada característica de Gabor antes (linha azul) e após correção de cores (linha vermelha). Quanto menor as variações, menor é a influência do dispositivo na aquisição.

O cenário ideal seria que, dada uma imagem, ela obtivesse os mesmos valores numéricos (características) independente do dispositivo de captura. Embora esse cenário seja intangível devido a diversos fatores, quanto mais próximas as características estiverem melhor. Di-

ante do exposto observou-se que a variância de quase a totalidade dos descritores caiu acentuadamente nas imagens corrigidas, demonstrando que numericamente as características obtiveram valores mais próximos.

A correção portanto trouxe bons resultados uma vez que aproximou os descritores. É importante observar que qualquer outra imagem foliar submetida à mesma análise irá obter resultados semelhantes. O mesmo comportamento podemos observar na Figura 6 com os descritores de Fourier.

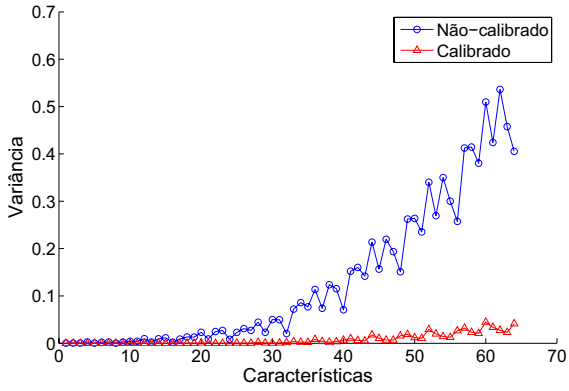


Figura 6. Variância observada de cada característica de Fourier antes (linha azul) e após correção de cores (linha vermelha). Correção de cores aproxima as características de uma mesma imagem adquirida em diferentes dispositivos.

Adicionalmente a correção de cores também minimiza a variância das características de diferentes folhas da mesma espécie. Calculando-se a variabilidade de cada característica obtida pelos métodos de Gabor e Fourier temos o gráfico das Figuras 7 e 8 respectivamente. Observa-se, especialmente no scanner B, uma queda bastante acentuada entre os valores numéricos de cada característica entre diferentes folhas da mesma espécie. Visto que para um sistema de classificação, supervisionado ou não, quanto mais próximas estiverem as características de uma mesma classe, mais fácil será sua classificação. Esse resultado portanto pode ser considerado bastante importante.

Também observa-se uma variabilidade menor do scanner A, tanto para as imagens corrigidas, quanto para as imagens originais. Esse fato talvez seja explicado pela baixa qualidade do dispositivo B, que pode acarretar em maiores erros de amostragem. A queda mais acentuada na variância das características

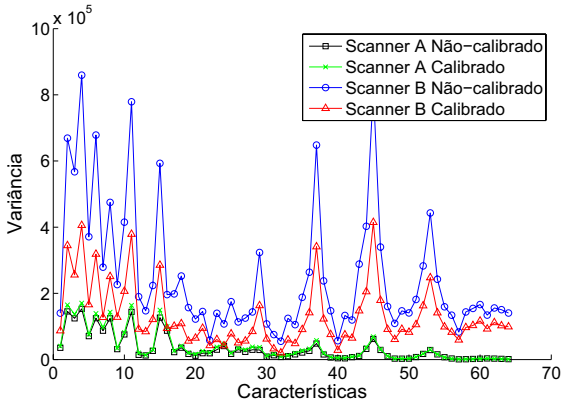


Figura 7. Correção de cores minimiza a variância das características de Gabor.

do scanner B pode também ser ocasionada por esse motivo, visto que esse erro de amostragem é minimizado após a correção.

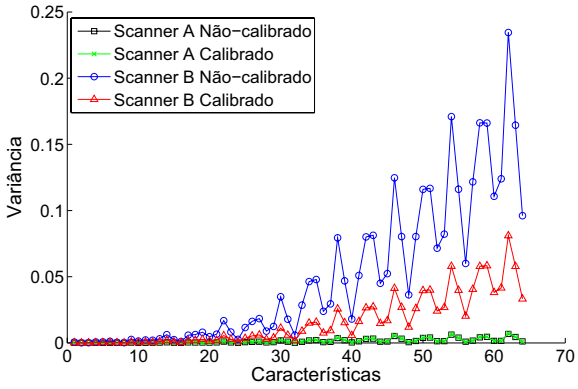


Figura 8. Correção de cores minimiza a variância das características de Fourier.

4. Conclusão

Este trabalho apresentou um estudo da influência da correção de cor entre dispositivos nos métodos de visão computacional. Observou-se que após correção de cores utilizando um sistema de gerenciamento de cores, ocorreu uma aproximação numérica das características entre dispositivos. Esta aproximação facilita o trabalho dos métodos de classificação empregados nos

sistemas de visão, visto que a variabilidade dos dados é diminuída. Embora não seja perfeita, no momento esse é o melhor mecanismo disponível para tentar garantir a fidelidade das cores nas imagens digitais.

Referências

[1] A. R. Backes, D. Casanova, and O. M. Bruno. Plant leaf identification based on volumetric fractal dimension. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(6):1145–1160, 2009.

[2] P. N. Belhumeur, D. Chen, S. Feiner, D. W. Jacobs, W. J. Kress, H. Ling, I. Lopez, R. Ramamoorthi, S. Sheorey, S. White, and L. Zhang. Searching the world's herbaria: A system for visual identification of plant species. In *ECCV 2008*, pages 116–129, 2008.

[3] R. de Oliveira Plotze, M. Falvo, J. G. Pádua, L. C. Bernacci, M. L. C. Vieira, G. C. X. Oliveira, and O. M. Bruno. Leaf shape analysis using the multiscale minkowski fractal dimension,a new morphometric method: a study with passiflora (passifloraceae). *Canadian Journal of Botany*, 83:287–301, 2005.

[4] R. C. Gonzales and R. Woods. *Digital Image Processing*. Addison Wesley, 1992.

[5] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[6] J. Liu, S. Zhang, and J. Liu. A method of plant leaf recognition based on locally linear embedding and moving center hypersphere classifier. In *ICIC 2009*, pages 645–651, 2009.

[7] A. S. Long. Scanner icc profile creator. http://www.nla.gov.au/preserve/dohm/sipc.html, 2009.

[8] B. S. Manjunath and W.-Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, 1996.

[9] D. R. Rossatto, D. Casanova, R. M. Kolb, and O. M. Bruno. Fractal analysis of leaf-texture properties as a tool for taxonomic and identification purposes: a case study with species from neotropical melastomataceae (miconieae tribe). *Plant Systematics and Evolution*, 291(1):103–116, 2010.

[10] K. Sokolowski. Color scanner calibration - a comparison of different methods. Master's thesis, Royal Institute of Technology, 2003.

Uma metodologia híbrida para segmentação de lesões de pele

Alex F. de Araujo, João Manuel R. S. Tavares

Faculdade de Engenharia, Universidade do Porto (FEUP) / Instituto de Engenharia Mecânica e Gestão Industrial (INEGI), R. Dr Roberto Frias s/n, 4200-465 - PORTO, Portugal
fa.alex@gmail.com, tavares@fe.up.pt

Roberta Barbosa Oliveira, Norian Marranghello, Aledir Silveira Pereira

Departamento de Ciências da Computação e Estatísticas (DCCE) / Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, R. Cristóvão Colombo, 2265, 15054-000 – São José do Rio Preto, Brasil
robyoliveira1@gmail.com, {norian, aledir}@ibilce.unesp.br

Ricardo Baccaro Rossetti

Clinica DERM, Av. José Munia 5950. 15054-000, São José do Rio Preto – SP – Brasil
rbrossetti2010@hotmail.com

Abstract

Computer aided methods are widespread demanded in medical applications. As such, methodologies to automatic extract contours are desired to aid the automatic diagnosis of skin lesions. In this work, we propose a hybrid method to detect and extract skin lesion contours from dermatoscopic images. In the proposed method, to obtain the contour that includes all lesion regions, the region growing technique, based on a Quadtree implementation, is used to extract an initial contour. Afterwards, this contour is refined by using a traditional active contour model. Experimental results indicate that the proposed method is promising to detect skin lesion areas and to extract their contours from dermatoscopic images. Actually, the extracted contours maintain the original lesion features that are usually used in their diagnosis. Additionally, the results allow concluding that the method is able to detect lesion regions even in images with strong noise, like in images of the scalp.

Palavras-chave: Segmentação de Imagens Médicas, Lesões de Pele, Crescimento de Regiões, Contornos Ativos.

1. Introdução

Metodologias computacionais para processamento e análise de imagens médicas têm sido extensivamente pesquisadas, e várias soluções desenvolvidas para auxílio aos profissionais da área médica. Estas soluções visam ajudar no diagnóstico e no acompanhamento da evolução de doenças e dos planos de tratamento a partir de imagens, de forma rápida e precisa.

As lesões de pele são cada vez mais frequentes e podem indicar doenças sérias, tal como o câncer de pele. Frequentemente, para obter o seu diagnóstico, os dermatologistas analisam visualmente imagens das regiões suspeitas. A possibilidade destas lesões serem um câncer, torna este diagnóstico inicial muito importante na definição de um plano de tratamento adequado. No entanto, alguns fatores, como cansaço visual do dermatologista, pequena dimensão das lesões, especialmente quando ainda em estágio inicial de desenvolvimento, e as variações nas imagens causadas pela presença de ruídos e reflexos, tornam tal diagnóstico difícil, e por vezes impreciso.

Para a análise de imagens de lesões de pele, os dermatologistas utilizam frequentemente as características da borda e da região interna das lesões para fazerem um diagnóstico inicial e visual das lesões. Além disso, pode ser considerada também a evolução da lesão com o passar do tempo. A análise da evolução pode ser justificada devido a um aumento considerável do número de casos de câncer de pele causados pela evolução de manchas ou lesões, que são afetadas por fatores externos, como a exposição excessiva ao sol [1]. Assim, as técnicas de processamento e análise computacional de imagens podem ser usadas para ajudar o dermatologista na realização de diagnósticos mais eficientes, extraindo os contornos e as características das lesões a partir das imagens em estudo.

Para que um método computacional de auxílio a diagnóstico de lesões de pele seja eficiente, o contorno extraído deve preservar as características de irregularidade da fronteira da lesão. Com o objetivo de detectar e extrair os contornos de lesões de pele a partir de imagens, mantendo sua irregularidade, propõe-se neste trabalho a aplicação do método de crescimento de regiões, seguido da aplicação do método tradicional de contorno ativo. Assim, o método de crescimento de regiões é aplicado para segmentar a imagem original e extrair um

contorno inicial da lesão. Posteriormente, este contorno é usado como a curva inicial a ser refinada pelo método de contorno ativo.

O crescimento de regiões foi implementado baseando-se no algoritmo *Quadtree*, adotando o desvio padrão dos níveis das componentes do espaço de cor RGB (*Red-Green-Blue*) de cada quadrante como parâmetro de controle de crescimento. Depois da sua execução, um algoritmo de união é aplicado baseado na intensidade de cor das áreas da lesão em estudo, tornando possível a extração do seu contorno inicial. A topologia deste contorno inicial é similar à topologia da borda desejada. No entanto, o contorno inicial deve ser refinado para representar convenientemente a borda da lesão. Para tal refinamento, o método tradicional de contorno ativo é usado, considerando o contorno resultante do algoritmo de união como a curva inicial a ser refinada.

Este artigo está organizado da seguinte forma: na próxima seção, trabalhos correlatos à segmentação de imagens são apresentados. Na seção 3, descreve-se o método proposto. Uma discussão sobre os testes realizados e os resultados experimentais obtidos é incluída na seção 4. Finalmente, são apresentadas as conclusões e sugestões para trabalhos futuros.

2. Trabalhos Correlatos

Geralmente, as técnicas de segmentação de imagens médicas tentam detectar estruturas, tais como órgãos, lesões, tumores e tecidos, representados em imagens, bem como extrair seus contornos de uma forma eficiente, robusta e automatizada. As pesquisas nesta área têm focado em métodos capazes de segmentar eficientemente as imagens afetadas por ruídos e outras interferências, evitando a perda das características principais da borda original, como a rugosidade, irregularidade e forma. Outra característica muito pesquisada é a automatização dos métodos na tentativa de evitar intervenções externas e subjetivas durante a etapa de processamento computacional das imagens. Há métodos de segmentação baseados em diferentes conceitos e técnicas, como a limiarização de imagens, crescimento de regiões, algoritmos genéticos, redes neurais artificiais e modelos de contornos ativos [2].

As metodologias de segmentação baseadas em limiares tentam separar as regiões de interesse do fundo da imagem, usando valores como classificadores de uma característica particular. Por exemplo, na limiarização simples de uma imagem em níveis de cinza, um dos níveis de cinza é escolhido como limiar, e então todos os pixels com intensidade maior do que este limiar são classificados como pertencentes à estrutura e os outros são classificados como fundo da imagem [3].

Na tentativa de unir o maior número possível de pixels em regiões homogêneas, têm sido propostos métodos baseadas em crescimento de regiões. Uma abordagem frequenteme usada é dividir a imagem de entrada em conjuntos de regiões disjuntas, tal como realizado pelo método de divisão e união. Esta técnica também é chamada de método de crescimento de regiões *Quadtree*, por usar uma *Quadtree* como estrutura de dados para a imagem. Esta abordagem consiste em dividir recursivamente a imagem original em quadrantes, até que um dado parâmetro de crescimento P seja verdadeiro. Usualmente, este parâmetro é baseado nos níveis de intensidade de cada quadrante, como a média por exemplo. Então uma árvore é construída, onde cada nó não-folha possui 4 (quatro) nós filhos, os quais podem ser unidos de acordo com sua similaridade [3]. Diversas outras abordagens para crescimento de regiões têm sido propostas [4, 5].

Algoritmos genéticos (AG) têm sido usados na segmentação de imagens com características variadas [6, 7]. Os AG usam funções, conhecidas como operadores genéticos, para gerar novas populações a partir de uma população inicial, com o objetivo de produzir indivíduos mais aptos. Os operadores mais comuns são o de cruzamento e mutação. O primeiro recombina as características dos pais durante o processo de reprodução, resultando na herança de características pelas gerações seguintes. Este operador pode ser implementado de diferentes formas: cruzamento de ponto único, onde um ponto é selecionado para dividir os cromossomos dos indivíduos pais em duas partes, e então as informações genéticas (genes) dos pais são trocadas, passando uma das partes de um pai para o outro; e o cruzamento multi-ponto, onde os genes são trocados considerando mais do que um ponto de corte. O operador mutação atua na manutenção da diversidade genética da população, evitando a estagnação na evolução dos indivíduos, que pode gerar resultados falsos-positivos.

As redes neurais artificiais (RNA) tentam simular o funcionamento do cérebro humano para interpretar e resolver problemas computacionais. Técnicas baseadas em RNA têm sido muito usadas no processamento e análise de imagens médicas. Iscan e colaboradores, por exemplo, apresentaram em [8] uma abordagem através do uso de uma rede neural incremental para a segmentação de imagens. O método realiza a atualização dos pesos dos nós (neurônios) da rede durante o treinamento da mesma, melhorando a performance da classificação das regiões na imagem.

Com o objetivo de desenvolver métodos de segmentação mais precisos e capazes de realizar a detecção aceitável de objetos e estruturas irregulares, várias técnicas de segmentação de imagens baseadas no modelo de contorno ativo proposto por Kass, Within e Terzopoulos [9] têm sido propostas. Usualmente, estes

métodos iniciam com uma curva inicial, definida dentro do domínio da imagem, deformando esta curva em direção à borda desejada, pela ação de forças internas e externas aplicadas sobre a curva. Esta deformação é obtida pela minimização da energia da curva, sendo a energia mínima encontrada quando a curva está sobre a característica a ser segmentada. O uso dos métodos de segmentação baseados em contornos ativos tem sido muito explorado no processamento de imagens médicas [10, 11].

Existe também cada vez mais uma tendência para unir diferentes técnicas [12, 13]. Estas metodologias híbridas têm ganhado atenção especial devido à sua habilidade para produzir resultados mais precisos, além de processar imagens mais complexas. Através da combinação das características de duas ou mais técnicas, estas metodologias tentam superar algumas dificuldades da segmentação, tais como a heterogeneidade das regiões, as interferências de ruídos, as variações de posicionamento ou oclusões. Jianli e Baoqi [14], por exemplo, utilizaram redes neurais genéticas para a segmentação de câncer de pele a partir de imagens. Os autores utilizaram o algoritmo genético para otimizar os pesos dos nós da rede neural, tentando acelerar o treinamento da mesma. O método para segmentação de imagens usando redes neurais genéticas se mostrou viável, mas o valor do limiar usado para segmentação da imagem não foi adequado para todos os tipos de imagens de câncer de pele, principalmente nos casos quando a estrutura da lesão é complexa e o contraste entre as regiões saudável e doente é baixo.

3. Método Proposto

As lesões de pele são analisadas a partir de imagens obtidas de duas maneiras: por fotografia tradicional e por meio de um exame de dermatoscopia. As imagens de dermatoscopia possuem mais detalhes das lesões do que as obtidas por fotografia convencional, porque são obtidas por um equipamente especial chamado dermatoscópio. Além disso, é comum também adicionar um gel sobre a lesão para realçar o contraste entre as áreas doente e saudável. No entanto, os exames por dermatoscopia não são acessíveis a toda população, sendo ainda muito usado o exame por fotografias convencionais. Na Figura 1, pode-se observar duas imagens, com lesões diferentes, obtidas através de fotografia convencional (a) e por um dermatoscópio (b).

Nota-se que em ambas as imagens da Figura 1, existe a interferências causada por ruídos e outros artefatos, como pêlos, bolhas e interferências da iluminação. Além disso, as lesões de pele possuem formas, dimensões e cores variadas. Assim, realizar a caracterização visual destas lesões é uma tarefa complexa. A extração eficiente das bordas de lesões é crucial para facilitar o seu diagnóstico pelo dermatologista. Neste trabalho, é proposta uma metodologia híbrida para extrair contornos de lesões de pele a partir de imagens dermatoscópicas e convencionais, usando os métodos de crescimento de regiões e contornos ativos.

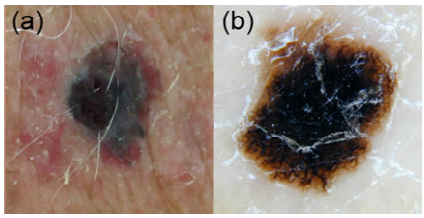


Figura 1: Exemplo de imagens de lesões de pele obtidas através de fotografia convencional (a) e por dermatoscópio (b).

Na Figura 2, apresenta-se o diagrama de fluxo do método proposto. Devido ao ruído e artefatos que normalmente afetam as imagens a analisar, estas são pré-processadas usando o filtro por difusão não-linear proposto por Barcelos e colaboradores [15]. Este método de remoção de ruídos suaviza a imagem original de forma não linear, retardando o efeito degenerativo da suavização nas regiões de borda.

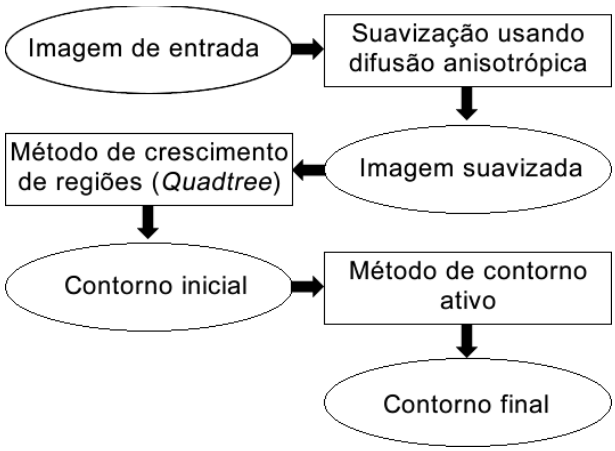


Figura 2: Diagrama de fluxo do método de segmentação proposto.

Em seguida, as imagens são segmentadas pelo uso do método de crescimento de regiões. Em particular, um método de crescimento de regiões *Quadtree* é usado para obter os contornos aproximados das lesões. Nesta segmentação inicial, cada componente do sistema de cores RGB é processada separadamente e unidas posteriormente. Na tentativa de aceitar apenas os quadrantes da imagem que possuam uma variação de intensidade reduzida, adotou-se o desvio padrão das intensidades dos pixels de cada quadrante como parâmetro de crescimento. Sendo a imagem suavizada

dividida até que o desvio padrão do quadrante seja inferior a 20% do desvio padrão da imagem antes da divisão. Este valor foi definido experimentalmente a partir de testes realizados sobre o conjunto de imagens analisadas neste trabalho. Assim, a divisão acontece até que este parâmetro seja satisfeito ou, no pior caso, até que o quadrante tenha dimensões iguais a 2x2 pixels e assim não seja mais divisível.

O passo anterior permite que as imagens suavizadas sejam representadas por um conjunto de regiões homogêneas, facilitando a detecção das regiões de pele que são potenciais áreas doentes. Para unir as várias regiões segmentadas e isolar o fundo da imagem, um método de união é aplicado considerando a distância entre as intensidades das regiões e agrupando as regiões com intensidades similares. A representação da imagem usando uma *Quadtree* foi adotada porque esta estrutura parte da maior porção da imagem (a imagem inteira) em direção à menor, assim os quadrantes que representam uma região homogênea não são totalmente divididos, e assim o custo computacional associado é reduzido. Em todas as imagens testadas, os contornos aproximados obtidos pelo método de crescimento de regiões foram quase sempre coincidentes com a topologia das bordas das lesões.

O resultado dos passos anteriores é uma imagem binarizada, a qual torna possível a extração dos contornos aproximados para as lesões pela avaliação das intensidades dos seus pixels. Estes contornos são então melhor ajustados para a fronteira da lesão pelo uso de um método de contorno ativo. Dois métodos de contorno ativo foram estudados: o método de contornos ativos tradicional (*snakes*) [9] e o método GVF (*Gradient Vecto Flow*) [16]. Os parâmetros usados relacionados com a elasticidade e a rigidez do contorno e o peso das forças aplicadas foram adotados como constantes e iguais para os dois métodos. O peso das forças externas adotado experimentalmente foi 0.1, e os parâmetros de elasticidade de rigidez foram considerados iguais e com valor 0.05. É importante ressaltar que os parâmetros adotados foram baixos para evitar que a topologia inicial da borda da lesão fosse muito afetada. Mesmo usando parâmetros de elasticidade e rigidez baixos, o método GVF gerou maiores deformações sobre os contornos iniciais do que o método tradicional de contornos ativos. Então, uma vez que o contorno inicial é já próximo do desejado, o modelo tradicional de contornos ativos foi escolhido para ajustar os contornos aproximados obtidos pelo método de crescimento de regiões às lesões em análise.

4. Resultados Experimentais

Durante os testes experimentais realizados com o método proposto, foram usadas 40 imagens coloridas, com 256x256 pixels, obtidas das bases de imagens Dermatlas [17] e Dermlite [18]. Destas imagens, 10 foram adquiridas por um dermatoscópio e 30 imagens por fotografia convencional. O conjunto de imagens adotado foi formado por imagens variadas, possuindo imagens com bom contraste, baixo contraste, afetadas por diferentes quantidades de ruído e artefatos, como pêlos, bolhas e reflexos de luz, e com lesões malignas, não-malignas e atípicas.

Na Figura 3, apresenta-se um exemplo de uma imagem retornada pela etapa de suavização. Nota-se que os ruídos presentes na imagem original (a) foram parcialmente removidos (b). Mesmo não sendo completamente eliminados os ruidos presentes na imagem original, a sua suavização foi suficiente para que a etapa de segmentação inicial pudesse detectar e extrair com sucesso o contorno aproximado das lesões representadas.

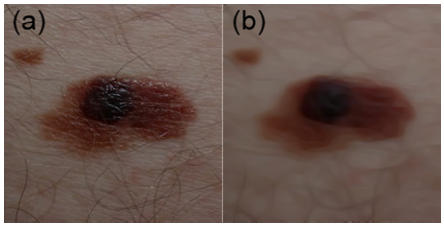


Figura 3: Exemplo da uma imagem original (a) e após a suavização por difusão anisotrópica (b).

Após a suavização da imagem original, o método de crescimento de regiões é aplicado, aumentando as áreas homogêneas, como apresentado na imagem (b) da Figura 4. Na figura referida, pode-se observar que as áreas internas da lesão e do fundo da imagem ficaram mais homogêneas (b) do que as correspondentes áreas na imagem original (a).

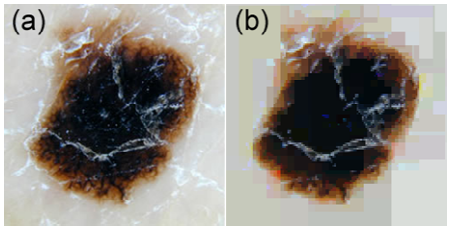


Figura 4: Exemplo do resultado da aplicação da etapa de divisão do método *Quadtree*: Imagem original (a) e imagem dividida em regiões (b).

Na Figura 5, são apresentadas algumas imagens originais (1a, 2a, 3a e 4a) bem como os respectivos contornos obtidos pelo método proposto (1b, 2b, 3b e 4b). Nas imagens apresentadas, nota-se que as lesões foram detectadas e envolvidas pelo contorno final, mesmo quando as lesões são de dimensão reduzida, como no caso da imagem 2b. Todos os contornos finais obtidos foram visualmente avaliados por um especialista que confirmou

que as lesões existentes nas imagens em estudo foram todas detectadas com sucesso. Vale ressaltar que as regiões com reflexo na parte interna das lesões, como na imagem 3b, não interferiram na segmentação devido à aplicação da suavização por difusão anisotrópica. No entanto, é importante destacar também, que se o reflexo sobrepor uma parte da borda da área doente, a segmentação será afetada, uma vez que serão perdidos alguns detalhes da borda naquela região.

Apesar de todas as regiões suspeitas de lesão terem sido detectadas eficientemente pelo método proposto, nas imagens com a presença de transições muito suaves entre as regiões saudável e doente e com muita interferência de pêlos, como no couro cabeludo, os contornos segmentados foram de qualidade inferior, como pode ser visto nos exemplos da Figura 6. Nesta figura, pode-se verificar que a região superior da lesão na imagem (a) (indicada pelo círculo amarelo) não foi corretamente envolvida pelo contorno final segmentado. Na imagem (b), tem-se o resultado do contorno obtido a partir da segmentação inicial, e nota-se que a região doente foi corretamente encontrada, mas várias bordas falsas ainda foram detectadas devido ao excesso de pêlos. No entanto, é importante salientar que, mesmo nos casos onde os contornos segmentados não foram bem ajustados às bordas das lesões, o método proposto foi capaz de detectar com sucesso todas as lesões de pele existentes.

5. Conclusões e Trabalhos Futuros

Um método híbrido para segmentação de imagens coloridas foi apresentado neste trabalho. O objetivo do método é segmentar imagens de lesões de pele, para detectar e extrair as bordas destas lesões e auxiliar no seu diagnóstico médico.

O método desenvolvido une as características das técnicas de crescimento de regiões e modelos de contornos ativos para detectar, extrair e refinar as bordas das lesões, preservando suas características principais, como a rugosidade e a irregularidade. Antes de aplicar os métodos para segmentação, a imagem original é pré-processada, usando um filtro de difusão anisotrópica, para amenizar as interferências causadas por ruídos e outros artefatos. Posteriormente, a imagem é pré-segmentada pelo método de crescimento de regiões *Quadtree*, definindo um contorno aproximado para as áreas suspeitas de lesão, o qual é seguidamente deformado pelo método de contornos ativos para obtenção do contorno final.

A partir da análise visual dos resultados obtidos, pode-se concluir que o método proposto é promissor, sendo capaz de detectar as prováveis regiões doentes em imagens de lesões de pele, tanto obtidas por uso de um dermatoscópio como por fotografia convencional.

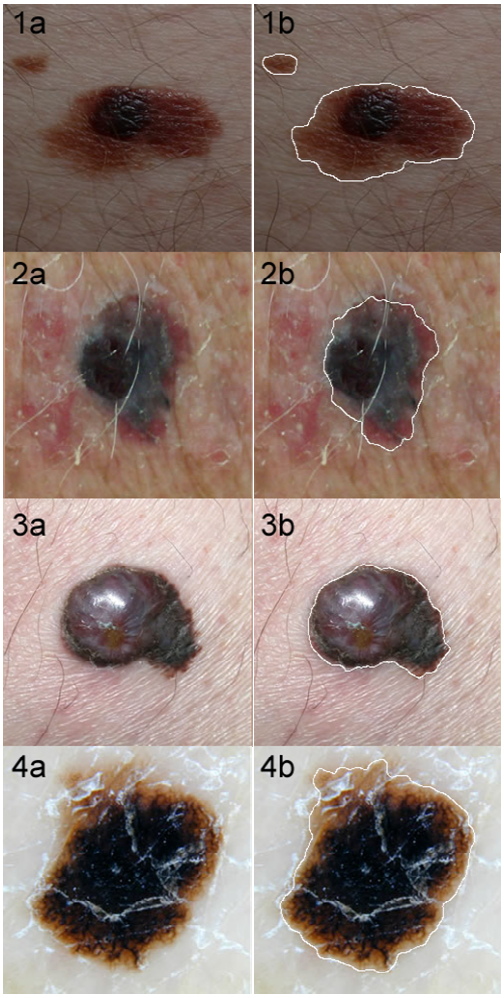


Figura 5: Alguns resultados obtidos pelo método proposto: imagens originais à esquerda (a); imagens originais com os respectivos contornos finais sobrepostos à direita (b).

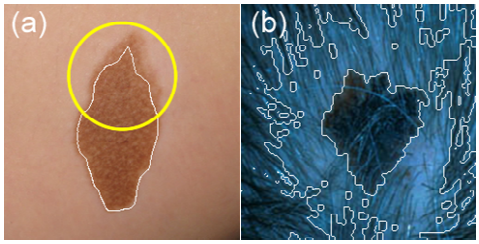


Figura 6: Exemplos de contornos que não foram corretamente ajustados à lesão existente.

No entanto, o método proposto ainda possui algumas limitações quando a imagem original possui transições muito suaves entre as regiões doente e saudável, ou quando as lesões encontram-se no couro cabeludo e existe uma presença excessiva de pêlos na região doente. Este último caso é comum em imagens obtidas por fotografia convencional, uma vez que no exame de dermatoscopia,

as áreas mais afastadas da lesão não são adquiridas, e a interferência de pêlos é menor. Para superar o problema das transições muito suaves, o método de contornos ativos usado será substituído por outro método capaz de realizar uma deformação mais adequada do contorno aproximado dado pelo método de crescimento de regiões. Além disso, na tentativa de solucionar as falhas para extrair o contorno das lesões no couro cabeludo, serão testados outros parâmetros para a metodologia de suavização adotada. Outra etapa a ser desenvolvida na sequência deste trabalho é uma melhor comparação do método proposto com os métodos existentes, bem como uma análise mais detalhada dos resultados, analisando as diferenças obtidas entre as imagens de fotografias convencionais e as imagens obtidas pelo exame de dermatoscopia.

Agradecimentos

O primeiro autor gostaria de agradecer à Fundação para a Ciência e a Tecnologia (FCT), em Portugal, pela sua bolsa de Doutoramento com referência SFRH/BD/61983/2009.

Os autores são gratos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), no Brasil, pelo suporte financeiro.

Este trabalho foi parcialmente desenvolvido no escopo dos projetos "Methodologies to Analyze Organs from Complex Medical Images – Applications to Female Pelvic Cavity", "Aberrant Crypt Foci and Human Colorectal Polyps: mathematical modelling and endoscopic image processing" e "Cardiovascular Imaging Modeling and Simulation - SIMCARD", com as referências PTDC/EEA-CRO/103320/2008, UTAustin/MAT/0009/2008 e UTAustin/CA/0047/2008, respectivamente, suportados pela FCT.

Referências

[1] Guide. Skin cancer self-examination. *Disponível em: http://www.skincancerguide.ca/prevention/self_examination.html*. (acessado em 2011).

[2] Ma, Z., Tavares, J. M. R., Jorge, R. N., Mascarenas, T. "A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Computer Methods in Biomechanics and Biomedical Engineering*. 2010. Vol. 13:2, p. 235-246.

[3] Gonzalez, R.C., R.E. Woods, and S.L. Eddins. "Digital Image Processing Using MATLAB". Prentice-Hall, Inc. 2003.

[4] Lee, W.-L., Chen, Y.-C., Chen, Y.-C., Hsieh, K.-S., "Unsupervised segmentation of ultrasonic liver images by multiresolution fractal feature vector". *Information Sciences: an International Journal*, 2005. Vol. 175:3, p. 177-199.

[5] M. E. Celebi, H. A. Kingravi, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz. S. W. Menzies. "Border detection in dermoscopy images using statistical region merging". *Skin Research and Technology*. 2008. Vol. 14:3, p. 347–353.

[6] Hashemi, S., Kiani, S., Noroozi, N., Moghaddam, M. E., "An image contrast enhancement method based on genetic algorithm". *Pattern Recognition Letters*, 2010. Vol. 31:13, p. 1816-1824.

[7] Mukhopadhyay, A., U. Maulik, "A multiobjective approach to MR brain image segmentation". *Applied Sof Computing*, 2011. Vol. 11:1, p. 872-880.

[8] Iscan, Z., Yüksel, A., Dokur, Z., Korürek, M.. Ölmez, T., "Medical image segmentation with transform and moment based features and incremental supervised neural network". *Digitital Signal Processing*, 2009. Vol. 19:5, p. 890-901.

[9] Kass, M., A. Witkin, D. Terzopoulos, "Snakes: Active contour models". *International Journal of Computer Vision*, 1988. Vol. 1:4, p. 321-331.

[10] Yoon, S.W., Lee, C., Kim, J. K., Lee, M., "Wavelet-based Multi-resolution Deformation for Medical Endoscopic Image Segmentation". *Journal of Medical Systems*, 2008. Vol. 32:3, p. 207-214

[11] Lu, R., Y. Shen, "Automatic Ultrasound Image Segmentation by Active Contour Model Based on Texture", *Proceedings of the First International Conference on Innovative Computing, Information and Control.* 2006, Vol. 2:1, p. 689-692.

[12] Dokur, Z., T. Ölmez, "Segmentation of ultrasound images by using a hybrid neural network". *Pattern Recognition Letters*, 2002. Vol. 23:14, p. 1825-1836.

[13] Kolár, R., J. Kozumplík, "Fuzzy Approach in Ultrasound Image Segmentation", *Proceedings of the International Conference, 7th Fuzzy Days on Computational Intelligence, Theory and Applications*. 2001, Vol. 1:1, p. 924-929.

[14] L. Jianli, Z. Baoqi. "The Segmentation of Skin Cancer Image Based on Genetic Neural Network". *Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering*. 2009. Vol. 5:1, p. 594-599.

[15] Barcelos, C.A.Z., M. Boaventura, E.C. Silva, "A well-balanced flow equation for noise removal and edge detection". *IEEE Transactions on Image Processing*, 2003. Vol. 12:1, p. 751-763.

[16] Xu, C., J.L. Prince, "Snakes, shapes, and gradient vector flow". *IEEE Transactions on Image Processing*, 1998. Vol. 7:3, p. 359-369.

[17] Dermatlas. "Dermatology image atlas". *Disponível em : http://dermatlas.med.jhmi.edu/derm* (acessado em 2011).

[18] DermLite. "Detect Skin Cancer – Early". *Disponível em: http://www.dermlite.com/* (acessado em 2011).

Geração de um Classificador para Detecção do HLB em Citros Utilizando Técnica *Box-Counting Multifractal*

Diego Carlos Pereira da Silva[1]; Lúcio André de Castro Jorge[2]; Adolfo Posadas[3]; Maria Stela Veludo de Paiva[4]

1- *Mestrando, Escola de Engenharia de São Carlos - Universidade de São Paulo - USP, Depto. de Engenharia Elétrica, São Carlos-SP, dicapesilva@usp.br.*
2- *Pesquisador, Embrapa Instrumentação Agropecuária, São Carlos – SP, lucio@cnpdia.embrapa.br.*
3- *Pesquisador, International Potato Center (CIP)- CGIAR, São Carlos-SP, a.posadas@cgiar.org*
4- *Professora, Escola de Engenharia de São Carlos - Universidade de São Paulo - USP, Depto. de Engenharia Elétrica, São Carlos-SP, mstela@sc.usp.br.*

Abstract

The Huanglongbing (HLB) is a serious disease that attacks citrus. It is caused by insect Diaphorina citri. Contamination occurs when the insect feeds on infected plants, from that moment he becomes the transmission vector of the same. HLB infected leaves with symptoms that can be mistaken with symptoms of some diseases and nutritional deficiencies. HLB to date is an incurable disease, and difficult to diagnose. In this way, this work developed a tool based on multifractal theory in order to characterize the self-similarities present in the leaves infected with HLB. Were analyzed 80 samples of leaves, 20 leaves infected with CVC, 20 with deficiency of manganese, 20 with deficiency of zinc and 20 infected with HLB. The results show that the method is quite effective for detection of HLB.

1. Introdução

O *Huanglongbing* (HLB, ex-*greening*) é uma doença destrutiva que ataca o citros. Essa doença tem como vetor transmissor o inseto *Diaphorina citri* na Ásia e na América, e o inseto *Trioza erytreae* na África. A contaminação se dá quando o inseto se alimenta de plantas doentes. A partir desse momento o inseto passa a ser o transmissor da doença [1]. As folhas das plantas contaminadas com o HLB apresentam manchas com formatos irregulares, de cor verde clara ou amarela, mescladas com o verde normal, sem apresentar uma nítida divisão entre elas. Os sintomas do HLB presentes nas folhas das plantas podem ser confundidos facilmente com sintomas de deficiências nutricionais como deficiência de Manganês e de Zinco e doenças como a CVC (Clorose Variegada dos Citros). A deficiência de Manganês faz com que as folhas apresentem clorose mais pálida e menos acentuada entre as nervuras. Na deficiência de Zinco as folhas apresentam clorose acentuada do limbo entre as nervuras. Na CVC surgem pequenas manchas amareladas e irregulares, espalhadas na frente da folha [2].

Para diagnosticar o HLB existem atualmente dois métodos, o método de inspeção visual e a análise do PCR (*Polymerase Chain Reaction*) [3]. O primeiro método apresenta a desvantagem de permitir diagnosticar a doença, somente vários meses após a contaminação da planta. Além disso, devido às condições de campo a que são acometidos os inspetores, os sintomas podem ser confundidos com outras doenças e deficiências nutricionais, proporcionando taxas de acerto inferiores a 50%. O método de análise do PCR é o mais eficaz, porém, o alto custo da análise impossibilita a sua realização em uma escala que possa permitir o controle da propagação da doença.

A prevenção do HLB se dá por meio do controle do vetor transmissor com a utilização de tratamentos químicos e pela erradicação das plantas contaminadas após o diagnóstico positivo para a doença.

Por se tratar de uma doença incurável até o presente momento, e de difícil diagnóstico, esforços vem sendo realizados na tentativa de se desenvolver ferramentas eficazes que auxiliem no diagnóstico dessa doença.

O presente artigo apresenta uma ferramenta baseada na teoria *multifractal* que caracteriza as auto-similaridades das folhas das plantas contaminadas pelo HLB, demais doenças e deficiências nutricionais de forma a permitir o reconhecimento de padrões e a identificação do HLB.

A teoria *multifractal* permite a caracterização de fenômenos complexos de maneira quantitativa, mesmo em meio a variações tanto temporais como espaciais. Esta teoria vem sendo cada vez mais utilizada para a análise e simulação de fenômenos complexos em uma grande variedade de escalas [4], devido aos bons resultados obtidos [5], [6], [7], [8], [9] e [10]. A análise *multifractal* é um método robusto utilizado para descrever eventos,

estruturas ou fenômenos, e possui como característica fundamental a auto-similaridade. Esta característica permite que independente da complexidade da forma de um objeto, observando-se mais detalhadamente a sua estrutura, um observador pode encontrar as mesmas formas (ou semelhantes) em escalas diferentes [11].

2. Materiais e métodos

2.1 Aquisição de Dados

O Grupo Fischer [12] disponibilizou à EMBRAPA Instrumentação Agropecuária, 80 amostras de folhas de citros de Valência com o porta-enxerto *Swingle*. Essas folhas já vieram com um pré-diagnóstico realizado por meio de inspeção visual realizada em laboratórios de maneira controlada longe das adversidades que dificultam o diagnóstico em campo, por técnicos especializados (pragueiros), não sendo realizado nenhum teste químico que pudesse confirmar a autenticidade do diagnóstico. Do total de 80 folhas, 20 delas foram diagnosticadas contendo a doença CVC, 20 apresentando deficiência de Manganês, 20 apresentando deficiência de Zinco e 20 delas contendo o HLB.

Foi realizada a aquisição das imagens dessas 80 folhas, utilizando um *scanner* de mesa da marca HP modelo *Scanjet* G4050. A resolução do mesmo foi ajustada para 100 DPI e a dimensão das imagens obtidas foi 400 x 200 *pixels*.

Na Figura 1 pode ser observada a imagem de folhas utilizadas neste trabalho, sendo a imagem (a) de uma folha com CVC, (b) de uma folha com deficiência de Manganês, (c) de uma folha com deficiência de Zinco e (d) de uma folha com HLB.

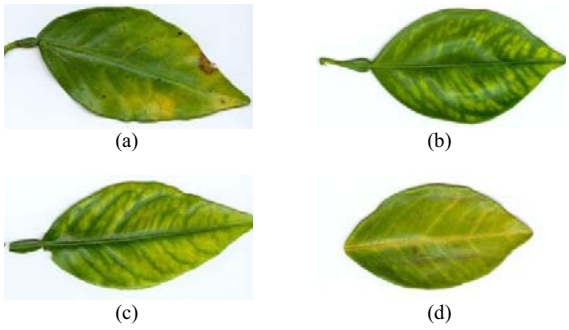


Figura 1 – (a) Imagem de uma folha com CVC, (b) de uma folha com deficiência de Manganês, (c) de uma folha com deficiência de Zinco e (d) de uma folha com HLB.

2.1. Segmentação por cor e Binarização

Como pode ser percebido nas imagens mostradas na Figura 1, em todas as doenças e deficiências nutricionais apresentadas neste artigo, os sintomas presentes nas folhas evidenciam a cor amarelo, e devido a essa característica presente em todas as folhas foi realizada a segmentação por cor [13].

Para a segmentação por cor da folha, foi desenvolvida uma ferramenta baseada em uma Rede Neural Artificial do tipo MLP (*MultiLayer Perceptron*) [14] com algoritmo *backpropagation,* de forma a separar a cor amarelo das demais cores presentes na folha. Foram utilizados como parâmetros de entrada os valores R, G e B do pixel, sendo selecionadas manualmente as áreas de interesse na imagem que pudessem representar as diversas tonalidades existentes para a cor amarela. Para o treinamento da rede foram utilizadas 64 amostras das folhas de citros e para os testes foram utilizadas 16 amostras. Dessa forma foram criadas as classes Amarelo e Fundo. Para os parâmetros referentes à taxa de aprendizagem, *momentum* e quantidade de neurônios na camada neural escondida, foram definidos, respectivamente, os valores 0,3, 0,2 e 3, com base em testes previamente realizados.

Na Figura 2 pode ser observada a imagem das folhas da Figura 1 após a segmentação, sendo a imagem (a) de uma folha com CVC, (b) de uma folha com deficiência de Manganês, (c) de uma folha com deficiência de Zinco e (d) de uma folha com HLB. O preto representa a região amarela da folha.

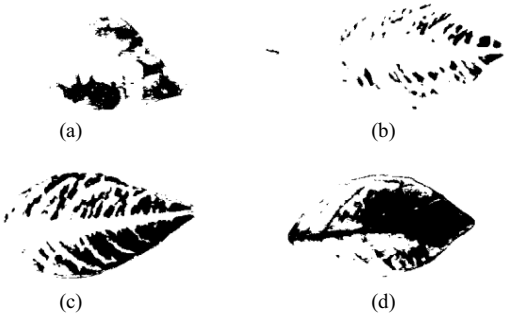


Figura 2 – (a) Imagem de uma folha após a segmentação com CVC, (b) de uma folha com deficiência de Manganês, (c) de uma folha com deficiência de Zinco e (d) de uma folha com HLB.

Em seguida foi aplicado o método *Box-Counting,* para a análise *multifractal* [15].

2.3. Análise multifractal

Como já mencionado na introdução, a análise *multifractal* permite a caracterização das auto-similaridades presentes nas diversas estruturas do objeto analisado [11].

Neste artigo foi utilizada a teoria *multifractal* na tentativa de se caracterizar as estruturas nas folhas amostradas que pudessem evidenciar características inerentes ao CVC, deficiências nutricionais e HLB.

Para a análise *multifractal* das imagens utilizadas neste artigo, foi utilizado o método *Box-Counting* [16], sendo este método eficiente para a análise de sinais binários.

No método *Box-Counting* a imagem é dividida em caixas de tamanho L e a probabilidade de se encontrar uma estrutura naquela região é calculada de acordo com a equação (1):

$$P_i(N) = \frac{Ni(L)}{Nt} \qquad (1)$$

Na equação (1) $P_i(N)$ representa o resultado do cálculo da probabilidade de determinada estrutura em cada uma das caixas em que foi dividida a imagem, $N_i(L)$ é a soma total dos valores de determinada estrutura de uma das caixas da imagem e N_t corresponde o valor da soma dessa estrutura para a imagem inteira.

Os valores obtidos na equação (1) são utilizados para determinar o valor de $\mu_i(q,l)$. Esse valor corresponde à normalização das probabilidades em relação ao intervalo de valores q, chamado de momento estatístico, esse valor permite analisar as estruturas explorando as diferentes regiões de medidas singulares. Para valores de q > 1 a região de medida mais singular é amplificada, para q < 1 a região com menor singularidade é acentuada e para q = 1 a medida original é replicada. A função $\mu_i(q,l)$ é dada pela equação (2):

$$\mu_i(q,l) = \frac{[P_i(l)]^q}{\sum_j [P_j(l)]^q} \qquad (2)$$

Os valores das funções *f(q)* e *α(q)* determinam a dimensão fractal, e são descritos pelas equações (3) e (4):

$$f(q) = \lim_{l \to 0} \frac{\sum_i \mu_i(q,l) x \ln(\mu_i(q,l))}{\ln(l)} \qquad (3)$$

$$\alpha(q) = \lim_{l \to 0} \frac{\sum_i \mu_i(q,l) x \ln(P_i(l))}{\ln(l)} \qquad (4)$$

A Figura 3 mostra o gráfico da relação *alpha(q)* por *f(q)* de uma folha, sendo a imagem (a) de uma folha com CVC, (b) de uma folha com deficiência de Manganês, (c) de uma folha com deficiência de Zinco e (d) de uma folha com HLB.

Na figura 3 o eixo das abscissas é constituído pelos valores de *alpha(q)* e o eixo das ordenadas é constituído pelos valores de *f(q)*.

O formato do gráfico representa uma assinatura para cada um dos padrões estudados.

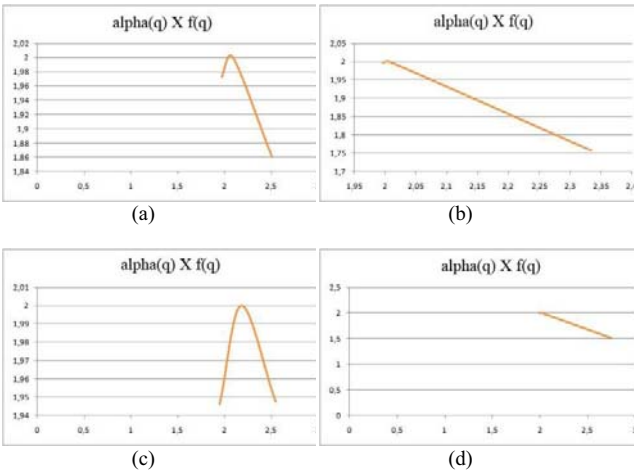


Figura 3 – (a) Imagem de uma folha com CVC, (b) de uma folha com deficiência de Manganês, (c) de uma folha com deficiência de Zinco e (d) de uma folha com HLB.

3. Classificadores

Na tentativa de se tentar identificar padrões para a realização do diagnóstico das folhas, alguns valores obtidos com análise a *multifractal* foram utilizados para compor o vetor de características.

3.1. Vetor de características

O vetor de características foi composto pelos valores das funções *alpha(q)*, *f(q)* e pelos valores das coordenadas de X1, X2, X3, Y1, Y2 e Y3 que representam os lados do triângulo formado pelo comportamento *fractal*.

Além dos valores das funções supracitadas foram utilizados os valores da dimensão fractal D0, D1 e D2 dados pela equação (5):

$$FD = \lim_{r \to 0} \frac{\log(N(r))}{\log(r)} \qquad (5)$$

Na equação (5) *N(r)* corresponde ao número de caixas necessárias para cobrir toda a imagem.

3.2. Classificadores

Neste artigo para a geração do classificador foi utilizada uma rede neural artificial do tipo MLP (*MultiLayer Perceptron)* [14] presente na API (*Application Programming Interface*) do *toolbox* WEKA [17]. Essa rede neural é baseada no algoritmo *backpropagation*.

Com base nos testes realizados durante a classificação, foram definidos os parâmetros referentes à taxa de aprendizagem, *momentum* e quantidade de neurônios na

camada neural escondida, respectivamente, como 0,3, 0,2, e 20. A função de ativação utilizada foi a sigmóide.

4. Resultados e Discussão

Foi realizado o experimento utilizando o vetor de características com os valores obtidos por meio da análise *mutifractal* conforme mencionado na seção 3.1.

Durante a classificação foram utilizados 10 *folds* com validação cruzada, aplicados a todas as imagens das folhas.

Do total das 80 amostras analisadas pelo classificador, ele conseguiu classificar corretamente 60 delas, representando um percentual de 75%.

O valor do erro absoluto médio foi de 0,1628 e o valor do erro quadrático médio foi de 0,3175.

A tabela 1 mostra a matriz de confusão contendo os resultados da classificação de acordo com cada classe.

Tabela 1 – (a) CVC, (b) deficiência de Manganês, (c) deficiência de Zinco e (d) de uma folha com HLB.

A	b	c	d
15	0	1	4
1	18	1	0
2	2	15	1
5	1	2	12

A interpretação dos valores presentes na tabela 1 é dada da seguinte forma:

- Do total de 20 instâncias contendo CVC, 1 instância foi classificada com deficiência de Zinco e 4 instâncias com HLB;
- Do total de 20 instâncias contendo deficiência de Manganês, 1 instância foi classificada com CVC e 1 instância foi classificada com deficiência de Zinco;
- Do total de 20 instâncias contendo deficiência de Zinco, 2 instâncias foram classificadas como CVC, 2 instâncias foram classificadas com deficiência de Manganês e 1 instância foi classificada como HLB;
- Do total de 20 instâncias contendo HLB, 5 instâncias foram classificadas como CVC, 1 instância foi classificada como deficiência de Manganês e 2 instâncias foram classificadas como deficiência de Zinco.

Para o CVC a precisão dada pelo classificador foi de 65,2%, para a deficiência de Manganês foi de 85,7%, para a deficiência de Zinco foi de 78,9% e para o HLB foi de 70,6%.

5. Conclusões

Este trabalho apresenta resultados da aplicação de um classificador baseado em uma rede neural artificial, tendo como parâmetros de entrada os valores obtidos pela análise *multifractal*, aplicados a uma área de extrema importância. Atualmente os valores obtidos por meio da inspeção visual para a detecção do HLB são inferiores a 50%, sendo os sintomas presentes nas folhas contaminadas com HLB confundido muitas vezes com os sintomas de outras doenças e deficiências nutricionais.

Dessa forma, a ferramenta desenvolvida nesse artigo baseada na teoria *multifractal* apresenta resultados satisfatórios se comparada ao método de inspeção visual.

Este trabalho pode ser utilizado como referência para o desenvolvimento de um sistema que possa ser ter aplicação direta no campo.

10. Referências

[1] BOVÉ, J. M. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. **Journal of Plant Pathology**, Pisa, v. 88, p.7-37, 2006.

[2] FUNDECITRUS. Manual Técnico do Greening 2009. Disponível em: <http://www.fundecitrus.com.br/ImageBank/PageFlip/pageflip.aspx?idPage=143>. Acesso em: 29 abr. 2010.

[3] INNIS, M. A.; GELFAND, D. H.; SNINSKY, J. J.; WHITE, T. J. (eds.), PCR Protocols: A guide to methods and applications, Academic Press: San Diego, 1990.

[4] SCHERTZER, D., S. LOVEJOY, 1994: Multifractal Generation of Self-Organized Criticality, in Fractals In the natural and applied sciences Ed. M.M. Novak, Elsevier, North-Holland, 325-339.

[5] MCCARTNEY, H.A. & FITT, B.D.L. Dispersal of foliar fungal plant pathogens: mechanisms, gradients and spatial patterns. In Jones, D.G. (ed.). **The Epidemiology of Plant Diseases**. Dordrecht, Kluwer. 1998. p. 139-160.

[6] MCATEER J.R.T., YOUNG C.A., IRELAND J., and GALLAGHER P.T. 2007. The bursty nature of solar flare x-ray emission. The American Astronomical Society..The Astrophysical Journal, 662:691-700.

[7] POSADAS A., GIMENEZ D., QUIROZ R., and PROTZ R. 2003. Multifractal Characterization Of Soil Pore Systems. **Soil Sci. Soc**. Am. J. 67: 1361-1369.

[8] POSADAS A., ZOROGASTÚA P., QUIROZ R. and LÉON-VELARDE C. 2005. Multifractal characterization of the spatial distribution of ulexite in a Bolivian salt flat.

International Journal of Remote Sensing. Vol. 26 (3): 615–627.

[9] LATKA M., GLAUBIC-LATKA M., LATKA D., WEST B. 2002. The loss of multifractality in migraines. <http://arxiv.org/PS_cache/physics/pdf/0204/0204010v1.pdf>.

[10] IVANOV P. C., AMARAL L. A. N, GOLDBERGER A. L., HAVLIN S., ROSENBLUM M. G., STRUZIK Z. R., and STANLEY H. E. 1999. Multifractality in healthy heartbeat dynamics. Nature 399: 461-465.

[11] MANDELBROT, B. 1983. The Fractal Geometry of Nature. Copyrighted material, USA.

[12] Grupo Fischer. Disponível em: <http://www.citrosuco.com.br/fischer/fischer/sites/fischer/portal_grupo/home/index.html>. Acesso em: 20 fev. 2011.

[13] SIMÕES, A. S. Segmentação de imagens por classificação de cores: uma abordagem neural. DISSERTAÇÃO DE MESTRADO. ESCOLA POLITÉCNICA DA USP. São Paulo. 2000.

[14] WITTEN, I. H.; FRANK, E. Data mining : practical machine learning tools and techniques. 2 ed. San Francisco, 2005.

[15] LOPES, R.; BETROUNI, N. Fractal and multifractal analysis. **Medical Image Analysis**. 13: 634-649.

[16] RUSSEL, D.; HANSON, J.; OTT, E. Dimension of strange attractors. Physical Review Letters 45 (14), 1175-1178.

[17] WEKA. Disponível em: < http://www.cs.waikato.ac.nz/ml/weka/ >. Acesso em: 29 abr. 2010.

11. Agradecimentos

Agradecemos o apoio financeiro do CNPq (processo 578627/2008-6), fornecimento das folhas de citros da empresa Fischer S.A.

Spatiotemporal Gabor filters: a new method for dynamic texture recognition

Wesley Nunes Gonçalves, Bruno Brandoli Machado, Odemir Martinez Bruno
Physics Institute of São Carlos (IFSC)
Institute of Mathematical Sciences and Computing (ICMC)
University of São Paulo (USP)
Av. Trabalhador São-carlense, 400
Cx. Postal 369 - São Carlos - SP - Brasil
wnunes@ursa.ifsc.usp.br, brandoli@icmc.usp.br, bruno@ifsc.usp.br

Abstract

This paper presents a new method for dynamic texture recognition based on spatiotemporal Gabor filters. Dynamic textures have emerged as a new field of investigation that extends the concept of self-similarity of texture image to the spatiotemporal domain. To model a dynamic texture, we convolve the sequence of images to a bank of spatiotemporal Gabor filters. For each response, a feature vector is built by calculating the energy statistic. As far as the authors know, this paper is the first to report an effective method for dynamic texture recognition using spatiotemporal Gabor filters. We evaluate the proposed method on two challenging databases and the experimental results indicate that the proposed method is a robust approach for dynamic texture recognition.

1. Introduction

The vision of animals provides a large amount of information that improves the perception of the world. This information is processed into different dimensions, including color, shape, illumination, and motion. While most of the features provide information about the static world, the motion provides essential information for interaction with external environment. In recent decades, the perception and interpretation of motion have attracted a significant interest in computer vision community [14, 16, 1, 9] motivated by the importance in both scientific and industrial communities. Despite significant advances, the motion characterization is still an open problem.

For modeling image sequences, three classes of motion patterns have been suggested [13]: dynamic texture, activities and events. The main difference between them relies on the temporal and spatial regularity of the motion field. In this work, we aim at modeling dynamic textures, also called temporal textures. They are basically texture in motion which is an extension of image texture to the spatiotemporal domain. Examples of dynamic texture includes real world scenes of fire, flag blowing, sea waves, moving escalator, boiling water, grass, and steam.

Existing methods for dynamic texture can be classified into four categories according to how they model the sequence of images. Due to the efficient estimation of features based on motion (e.g. optical flow), the motion based methods (i) are the most popular ones. These methods model dynamic textures based on a sequence of motion patterns [9, 13, 8]. For modeling dynamic texture at different scales in space and time, the spatiotemporal filtering based methods (ii) use spatiotemporal filters such as wavelet transform [7, 17, 5]. Model based methods (iii) are generally based on linear dynamical systems, which provides a model that can be used in applications of segmentation, synthesis, and classification [6, 3, 15]. Based on properties of moving contour surfaces, spatiotemporal geometric property based methods (iv) extract motion and appearance features from the tangent plane distribution [10]. The reader may consult [4] for a review of dynamic texture methods.

In this paper, we propose a new approach for dynamic texture modeling based on spatiotemporal Gabor filters [12]. As far as the authors know, the present paper is the first one to model dynamic texture using spatiotemporal Gabor filters. These filters are basically built using two parameters: the speed v and direction θ. To model a dynamic texture, we convolve the sequence of images to a bank of spatiotemporal Gabor filter built with different values of speed and direction. For each response, a feature vector is built by calculating the energy statistic.

We evaluate the proposed method by classifying dynamic texture from two challenging databases: dyntex [11] and traffic database [2]. Experimental results in both databases indicate that the proposed method is an effective approach for dynamic texture recognition. For

the dyntex database, filter with low speeds (e.g. 0.1 pixels/frame) achieved better results than high speeds. In fact, the dynamic texture in this database presents low motion patterns. On the other hand, for the traffic database, high speeds (e.g. 1.5 pixels/frame) achieved the best correct classification rate. In this database, vehicles are moving at a speed that matches the filter's speed.

This paper is organized as follows. Section 2 briefly describes spatiotemporal Gabor filters. In Section 3, we present the proposed method for dynamic texture recognition based on spatiotemporal Gabor filters. An analysis of the proposed method with respect to the speed and direction parameters is present in Section 4. Experimental results are given in Section 5, which is followed by the conclusion of this work in Section 6.

2. Spatiotemporal Gabor Filters

Gabor filters are based on the important finding made by Hubel and Wiesel in the beginning of the 1960s. They found that the neurons of the primary visual cortex respond to lines or edges of a certain orientation in different positions of the visual field. Following this discovery, computational models were proposed for modeling the function of this neurons and the Gabor functions proved to be suited for this purpose in many works.

Initially, the researches aimed at studying spatial properties of the receptive field. However, some posterior studies revealed that cortical cells change in time and some of them are inseparable functions of space and time. Therefore, these cells are essentially spatiotemporal filters and they combine information over space and time, which makes a great model for dynamic texture analysis.

In this work, the spatiotemporal receptive field is modeled by a family of 3D Gabor filters [12] described in Equation 1.

$$\begin{aligned} g_{(v,\theta,\varphi)}(x,y,t) = & \frac{\gamma}{2\pi\sigma^2}\exp\left(\frac{-((\overline{x}+v_c t)^2+\gamma^2\overline{y}^2)}{2\sigma^2}\right) \\ & .\cos\left(\frac{2\pi}{\lambda}(\overline{x}+vt)+\varphi\right) \\ & .\frac{1}{\sqrt{2\pi}\tau}\exp\left(\frac{-(t-\mu_t)^2}{2\tau^2}\right) \\ & \overline{x} = x\cos\theta + y\sin\theta \\ & \overline{y} = -x\sin\theta + y\cos\theta \end{aligned} \quad (1)$$

We now discuss the parameters of the spatiotemporal Gabor filter. Some parameters were empirically found based on studies of response in the receptive visual field [12]. The size of the receptive field is determined by the standard deviation σ of the Gaussian factor. The parameter γ is the rate that specifies the ellipticity of the Gaussian envelope in the spatial domain. This parameter is set to $\gamma = 0.5$ for match to the elongated receptive field along $\overline{y}$ axis. The speed v is the phase speed of the cosine factor, which determines the speed of motion. The speed which the center of the spatial Gaussian moves along the $\overline{x}$ axis is specified by the parameter v_c. When $v_c = 0$, the center of the Gaussian envelope is stationary. On the other hand, a moving envelope is obtained when $v_c = v$. Figure 1 presents a moving envelope with $v_c = v = 1$.

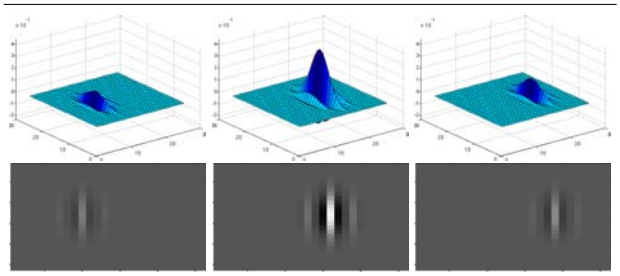


Figure 1. Example of spatiotemporal Gabor filter for $v_c = 1$.

The parameter λ is the wavelength of the cosine factor. It is obtained through the relation $\lambda = \lambda_0\sqrt{1+v^2}$, where the constant $\lambda_0 = 2$. The angle $\theta \in [0, 2\pi]$ determines the direction of motion and the spatial orientation of the filter. The phase offset $\varphi \in [-\pi, \pi]$ determines the symmetry in the spatial domain. The Gaussian distribution with mean μ_t and standard deviation τ is used to model the change in intensities. The mean $\mu_t = 1.75$ and standard deviation $\tau = 2.75$, both parameters are fixed based on the mean duration of most receptive fields.

3. Dynamic Textures Modeling based on Spatiotemporal Gabor Filters

In this section, we describe the proposed method for dynamic texture modeling based on spatiotemporal Gabor filters. Briefly, the sequence of images is convolved with a bank of spatiotemporal Gabor filters and a feature vector is constructed with the energy of the responses as components.

The response $r_{(v,\theta,\varphi)}(x,y,t)$ of a spatiotemporal Gabor filter $g_{(v,\theta,\varphi)}$ to a sequence of images $I(x,y,t)$ is computed by convolution:

$$r_{(v,\theta,\varphi)}(x,y,t) = I(x,y,t) * g_{(v,\theta,\varphi)} \quad (2)$$

Spatiotemporal Gabor filters are phase sensitive because its response to a moving pattern depends on the exact position within the sequence of images. To overcome this drawback, a response that is phase insensitive can be obtained

by:

$$R_{(v,\theta)} = \sqrt{r^2_{(v,\theta,\varphi)}(x,y,t) + r^2_{(v,\theta,\varphi/2)}(x,y,t)} \quad (3)$$

To characterize the Gabor space resulting from the convolution, the energy of the response is computed according to Equation 4.

$$E_{(v,\theta)} = \sum_{x,y,t} R^2_{(v,\theta)}(x,y,t) \quad (4)$$

A central issue in applying spatiotemporal Gabor filters is the determination of the filter parameters that covers the spatiotemporal frequency space, and captures dynamic texture information as much as possible. Each spatiotemporal Gabor filter is determined by two main parameters: the direction θ and the speed of motion v. In order to cover a wide range of dynamic textures, we design a bank of spatiotemporal Gabor filter using a set of values for speed $V = [v_1, v_2, \ldots, v_n]$ and direction $\Theta = [\theta_1, \theta_2, \ldots, \theta_m]$. The feature vector that characterizes the dynamic texture is composed by the energy of the response for each combination of velocities V and direction Θ (Equation 5).

$$\varphi = [E_{(v_1,\theta_1)}, \ldots, E_{(v_n,\theta_m)}] \quad (5)$$

The proposed method is summarized in Figure 2. First, we design a bank of spatiotemporal Gabor filters composed by filters with different directions and speeds. Then, the sequence of images is convolved with the bank of filters. For each convolved sequence of images, we calculate the energy to compose a feature vector.

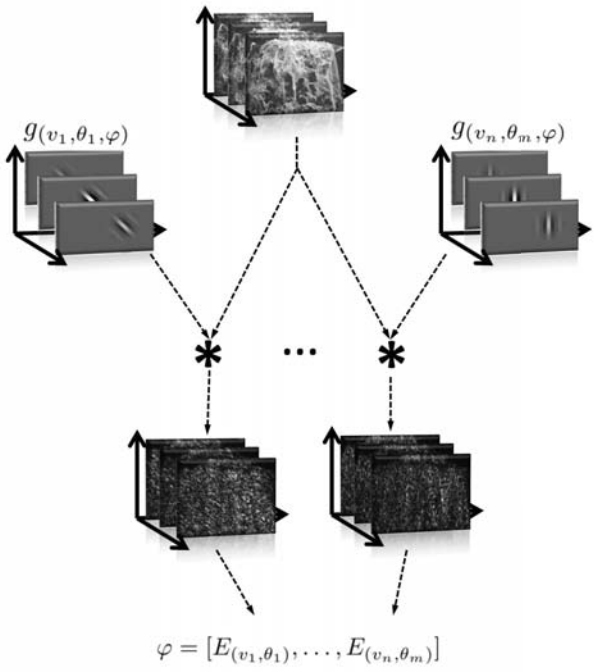


Figure 2. Proposed method considers the following steps: (i) Design a bank of spatiotemporal Gabor filters using different values of speed v and direction θ; (ii) Convolve the sequence of images with the bank of filters; (iii) Calculate the energy for each response to compose the feature vector.

4. Response Analysis of Spatiotemporal Gabor Filters

Here, we analyze the speed and direction properties of the spatiotemporal Gabor filters in synthetic sequence of images. In Figure 3, we present the response of spatiotemporal Gabor filters to bars moving at the same speed but in different direction θ. The filters and the moving bars have preference for the same speed $v = 1$. The response has the highest magnitude when the direction of the filter θ matches the direction of the moving bar. For instance, when $\theta = 0$, a vertical bar moving rightwards evokes higher response than bars with other direction of movement.

The speed property is evaluated in Figure 4. We analyze the response of spatiotemporal Gabor filters to edges drifting rightward at different speeds. The filters and the synthetic sequence of images have for the same preference direction $\theta = 0$. The highest response is achieved by filters which the speed matches the speed of the moving edge.

In Figure 5(a), we plot the response of filters to a moving bar with direction $\theta = 0$ at a speed $v = 1$. The response reaches its maximum value to a filter with direction $\theta = 0$, which matches to the direction of the moving bar. As we can see, the filter with moving envelope ($v_c = v$) achieved higher response than a filter with stationary envelope ($v_c = 0$). The plot for the speed parameter is shown in Figure 5(b). We convolved filters to an edge drifting rightward in direction $\theta = 0$ at a speed of $v = 2$. The maximum response is achieved for the filter whose speed matches to stimulus' speed. Again, we can conclude that filters with the moving envelope are more selective for both direction and speed than filters with stationary envelope.

5. Experimental Results

In this section, we present the experimental results using two databases: (i) dyntex database and (ii) traffic video database. The dyntex database consists of 50 dynamic texture classes each containing 10 samples collected from Dyntex database [11]. The videos are at least 250 frames long with dimension of 400×300 pixels. Figure 6 shows

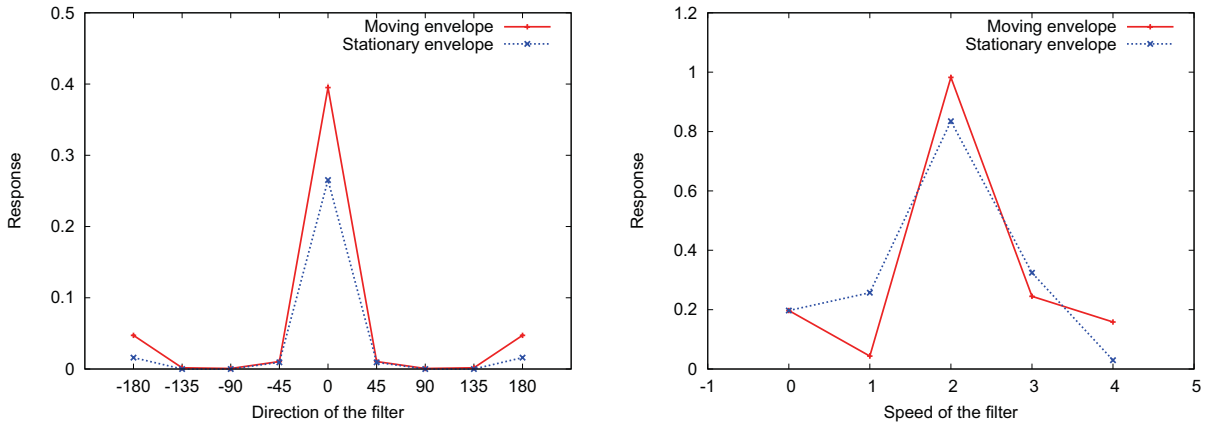


Figure 5. (a) Response for filters of different direction θ to a moving bar with $\theta = 0$. (b) Response for filters of different speeds v to a edge moving at speed $v = 2$.

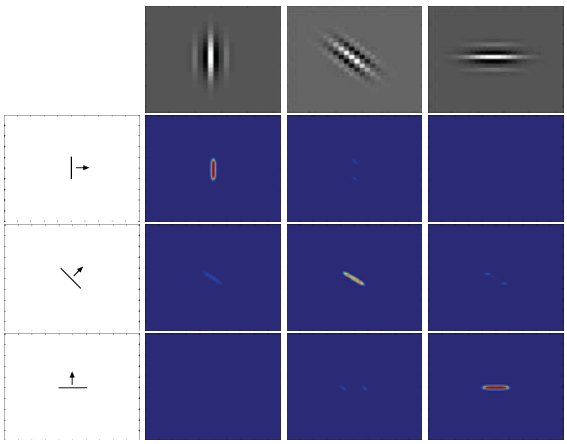


Figure 3. Response of spatiotemporal Gabor filters to bars moving in different direction θ. First row corresponds to the filters. Second, third and fourth rows corresponds to the response of a bar moving in direction $\theta = 0, \frac{\pi}{4}, \frac{\pi}{2}$, respectively.

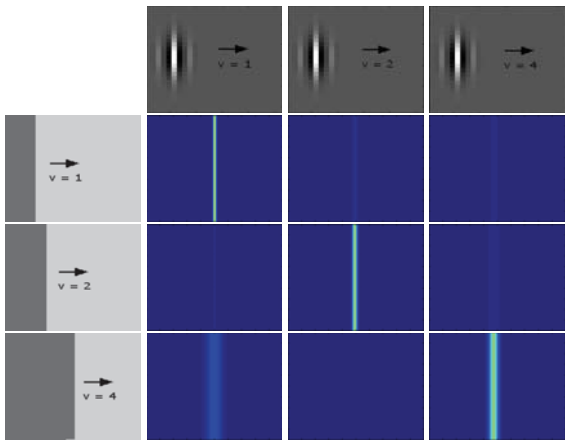


Figure 4. Response of spatiotemporal Gabor filters to edges moving at different speed v. First, second, and third rows corresponds to the response of a edges moving at speed $v = 1, 2, 4$, respectively.

examples of dynamic textures from the first database. The second database, collected from traffic database [2], consists of 254 videos divided into three classes: light, medium, and heavy traffic. Videos had 42 to 52 frames with a resolution of 320×240 pixels. The variety of traffic patterns and weather conditions are shown in Figure 7. All the experiments used a k-nearest neighbor classifier with $k = 1$ in a scheme 10-fold cross validation.

Now, we discuss the influence of the direction and speed parameters in the dynamic texture recognition. Table 1 shows the average and standard deviation of correct classification rate for the traffic database. Columns present the direction parameter evaluation using 4-directions, which is a combination of filters with $\theta = [0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}]$, and 8-directions which is a combination of filters with $\theta = [0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}]$. Rows present the speed evaluation using combination of speed with step of 0.25. As we can see, the bank of filter composed by filters with 8-directions outperformed the bank of filter with 4-directions

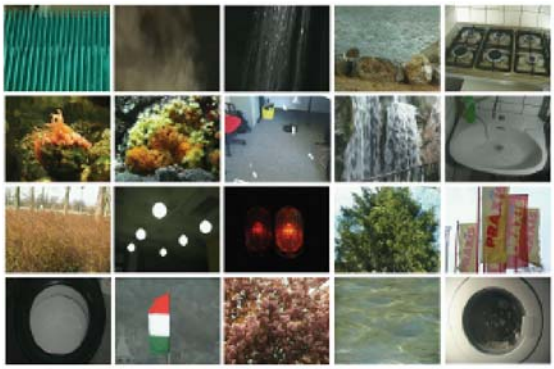


Figure 6. Examples of dynamic textures from dyntex database [11]. The database is composed by 50 dynamic texture classes each containing 10 samples.

Figure 7. Examples of dynamic textures from traffic database [2]. The database consists of 254 videos split into three classes of light, medium and heavy traffic.

for all combination of speeds. However, very little improvement can be appreciated as the combination of direction is increased. The improvement in correct classification rate was on average less than 1% when direction combination rises from 4 to 8. This is because the cars on the highway move always on the same direction, which can be modeled by 4-direction. With respect to the speed parameter, the best results were achieved for high speeds, such as 1.5 pixels/frame and 2.0 pixel/frame. A correct classification rate of 91.50% was achieved by a bank of filter composed by 8-directions and speeds $[0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0]$. The high speeds match to speed of the cars in the sequence

Speed Combination	4-directions	8-directions
$[0.50 - 0.75]$	90.06(5.53)	90.18(5.62)
$[0.50 - 1.00]$	90.17(5.70)	90.57(5.15)
$[0.50 - 1.25]$	89.07(5.91)	89.90(5.06)
$[0.50 - 1.50]$	89.68(5.99)	89.74(5.80)
$[0.50 - 1.75]$	89.69(6.18)	90.56(5.58)
$[0.50 - 2.00]$	89.45(5.77)	**91.50(5.20)**
$[0.50 - 2.25]$	90.24(5.51)	90.87(5.07)
$[0.50 - 2.50]$	89.33(5.60)	90.75(5.26)
$[0.50 - 2.75]$	89.76(5.44)	90.35(5.44)
$[0.50 - 3.00]$	89.60(5.16)	90.63(5.21)

Table 1. Correct classification rate and standard deviation for different combinations of speed and direction on the traffic database.

Speed Combination	4-directions	8-directions
$[0.1 - 0.2]$	92.50(3.40)	94.92(3.09)
$[0.1 - 0.5]$	96.00(2.40)	96.82(2.41)
$[0.1 - 1.0]$	96.56(2.11)	98.02(1.65)
$[0.1 - 1.5]$	96.92(2.32)	**98.60(1.60)**
$[0.1 - 2.0]$	97.24(2.18)	97.84(1.92)
$[0.1 - 2.5]$	96.92(2.49)	97.34(2.36)
$[0.1 - 3.0]$	96.37(2.98)	96.94(2.73)

Table 2. Correct classification rate and standard deviation for different combinations of speed and direction on the dyntex database.

of images and then the traffic jam can be modeled using these parameters.

In Table 2, we present the experimental results obtained on the dyntex database. The same combination of directions of the early experiment was used to evaluate the proposed method. However, as the dynamic textures in this database present low speeds, the combination of speed started in $v = 0.1$ pixels/frame and taken step of 0.1. As the previous results, the 8-direction bank of filters achieved higher values of correct classification rate compared to the 4-direction bank of filters. In this case, a correct classification rate of 98.60% was obtained, which clearly shows the effectiveness of the proposed method on the dynamic texture recognition.

6. Conclusion

In this paper, we proposed a new method for dynamic texture recognition based on spatiotemporal Gabor filters. First, it convolves a sequence of images to a bank of filters

and then extracts energy statistic from each response. Basically, the spatiotemporal Gabor filters are built using speed v and direction θ. The bank of filter is composed by filters with different values of speed and direction.

Promising results considering different combinations of v and θ were achieved on two important databases: dyntex and traffic database. On the traffic database, our method achieved a correct classification rate of 91.50% using combination of high speeds. On the other hand, a correct classification rate of 98.60% was obtained on the dyntex database using a combination of low speeds.

Acknowledgments. WNG was supported by FAPESP grants 2010/08614-0. BBM was supported by CNPq. OMB was supported by CNPq grants 306628/2007-4 and 484474/2007-3.

References

[1] C. Cedras and M. Shah. Motion-based recognition: A survey. *Image and Vision Computing*, 13(2):129–155, mar 1995.

[2] A. B. Chan and N. Vasconcelos. Classification and retrieval of traffic video using auto-regressive stochastic processes. In *IEEE Intelligent Vehicles Symposium*, pages 771–776, June 2005.

[3] A. B. Chan and N. Vasconcelos. Classifying video with kernel dynamic textures. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–6, 2007.

[4] D. Chetverikov and R. Péteri. A brief survey of dynamic texture description and recognition. In *Computer Recognition Systems, Proceedings of the 4th International Conference on Computer Recognition Systems*, volume 30 of *Advances in Soft Computing*, pages 17–26. Springer, 2005.

[5] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In *ICCCN '05: Proceedings of the 14th International Conference on Computer Communications and Networks*, pages 65–72, Washington, DC, USA, 2005. IEEE Computer Society.

[6] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto. Dynamic textures. *International Journal of Computer Vision*, 51(2):91–109, February 2003.

[7] S. Dubois, R. Péteri, and M. Ménard. A comparison of wavelet based spatio-temporal decomposition methods for dynamic texture recognition. In *IbPRIA '09: Proceedings of the 4th Iberian Conference on Pattern Recognition and Image Analysis*, pages 314–321, Berlin, Heidelberg, 2009. Springer-Verlag.

[8] R. Fablet and P. Bouthemy. Motion recognition using nonparametric image motion models estimated from temporal and multiscale cooccurrence statistics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(12):1619–1624, 2003.

[9] S. Fazekas and D. Chetverikov. Analysis and performance evaluation of optical flow features for dynamic texture recognition. *SP:IC*, 22(7-8):680–691, Aug. 2007.

[10] M. Fujii, T. Horikoshi, K. Otsuka, and S. Suzuki. Feature extraction of temporal texture based on spatiotemporal motion trajectory. In *ICPR*, pages Vol II: 1047–1051, 1998.

[11] R. Peteri, S. Fazekas, and M. Huiskes. Dyntex: A comprehensive database of dynamic textures.

[12] N. Petkov and E. Subramanian. Motion detection, noise reduction, texture suppression, and contour enhancement by spatiotemporal gabor filters with surround inhibition. *Biological Cybernetics*, 97:423–439, March 2008.

[13] R. Polana and R. C. Nelson. Temporal texture and activity recognition. In *Motion-Based Recognition*, page Chapter 5, 1997.

[14] H. Qian, Y. Mao, W. Xiang, and Z. Wang. Recognition of human activities using svm multi-class classifier. *Pattern Recognition Letters*, In Press, Corrected Proof:–, 2009.

[15] M. Szummer and R. W. Picard. Temporal texture modeling. In *ICIP*, pages III: 823–826, 1996.

[16] S. Wu and Y. F. Li. Motion trajectory reproduction from generalized signature description. *Pattern Recognition*, 43(1):204–221, 2010.

[17] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 819–826, 2004.

Segmentação do Músculo Peitoral em Imagens Mamográficas Através do Recorte Automático de ROI e da Binarização de Otsu

José Rogério Bezerra Barbosa Filho
Programa de Pós-Graduação em Informática, Universidade Federal da Paraíba
jose.rogerio.filho@gmail.com

Leonardo Vidal Batista
Programa de Pós-Graduação em Informática, Universidade Federal da Paraíba
leonardo@di.ufpb.br

Homero Schiabel
Programa de Pós-Graduação em Engenharia Elétrica
Universidade de São Paulo
homero@sc.usp.br

Luciana Buffa Verçosa
Programa de Pós-Graduação em Engenharia Elétrica
Universidade de São Paulo
dralubv@hotmail.com

Abstract

Breast cancer is the most common type of cancer in the female population and a serious global health problem. Mammography is the radiograph exam of the breast that allows the detection of cancer even in its early stages. Many Computer-Aided Detection and Diagnosis systems try to help physicians by either enhancing mammograms images or pointing to radiographic findings. This work presents a method for pectoral muscle segmentation in mammograms based on the extraction of a ROI containing part of the muscle and the application of Otsu's binarization method on this ROI. The results show an exact segmentation in 32% of the images and at least an adequate segmentation for 89% of images.

1. Introdução

O câncer de mama é o tipo mais comum de tumor que atinge a população feminina mundial. Estimativas sugerem que 410.000 mortes foram ocasionadas por essa doença no ano de 2002 [1]. Segundo o Instituto Nacional de Câncer (INCA), cerca de 49.240 novos casos de câncer de mama foram esperados para 2010, resultando em uma proporção de 49 casos para cada 100,000 mulheres [2].

A mamografia, o exame radiográfico da mama, é considerada a técnica mais eficiente para identificar lesões quando estas não são palpáveis ou quando as lesões são estruturas da mama modificadas [3]. Esse exame permite ao médico identificar diferenças de densidade nos tecidos da mama, característica fundamental para um diagnóstico correto.

Sistemas computadorizados de auxílio à detecção e diagnóstico (Computer-Aided Detection and Diagnosis, CADDx) têm como objetivo ajudar ao médico na detecção de lesões mesmo em seus primeiros estágios. Essas ferramentas chamam a atenção do profissional da saúde para áreas que podem conter achados radiográficos. Esse importante auxílio é razão suficiente para que estudos sejam feitos com o objetivo de melhorar a eficiência desses sistemas [4].

Uma importante etapa dos sistemas CADDx é a segmentação. Especificamente no caso da mamografia, a etapa de segmentação pode ser utilizada para reduzir a área da imagem a ser processada. Uma segmentação de boa qualidade é necessária para que, em etapas posteriores do CADDx, apenas estruturas de interesse sejam processadas. Eliminar partes da imagem que não trazem informação à detecção e ao diagnóstico como, por exemplo, o fundo da imagem, aumenta a eficiência e sensibilidade do sistema.

Esse trabalho propõe um método de segmentação do músculo peitoral em mamografias de incidência médio-lateral. O recorte automático de uma região de interesse (ROI) contendo o músculo peitoral, presente em outros trabalhos da literatura e mais detalhados nas sessões posteriores, é utilizado em conjunto com o método de binarização de Otsu [5] para obtenção de um limiar que será utilizado na segmentação da imagem. Por fim, a técnica de delimitação de contorno definida por Nunes *et al*. [6] é utilizada para encontrar os limites da área segmentada.

2. Fundamentação Teórica

Nessa sessão serão apresentados os conceitos presentes em trabalhos anteriores que foram utilizados no método proposto: o recorte automático da ROI onde se encontra presente o músculo peitoral, o algoritmo de binarização de Otsu e a delimitação de contorno de Nunes.

2.1. Recorte da ROI Contendo o Músculo Peitoral

A técnica de segmentação do músculo peitoral de Karssemeijer [7] tornou-se bastante recorrente na literatura. No trabalho de Karssemeijer, é feito um recorte

retangular numa área da mamografia que sabe-se conter parte ou todo o músculo peitoral. Em seguida, a transformada de Hough é aplicada nessa ROI para obtenção de uma reta que melhor delimita o músculo peitoral na imagem. O recorte automático de uma ROI contendo o músculo peitoral também foi utilizado em trabalhos posteriores como os artigos de Ferrari *et al* [8][9], Kwok *et al* [10] e Tzikopoulos *et al* [11].

No presente trabalho, utiliza-se o recorte automático da região de interesse presente dos trabalhos de Ferrari *et al.* [8][9]. Nessa abordagem, uma vez conhecido o contorno da mama, são definidos três pontos principais. O ponto A é o ponto superior inicial da mama. O ponto B está localizado na parte superior e marca o limite entre a mama e o fundo da imagem. O ponto C está localizado na parte inferior da imagem e, semelhante ao ponto B, marca o limite entre a mama e o fundo da imagem. Define-se também o ponto D, que é o ponto médio entre o ponto A e o ponto C. Dessa forma, a região de interesse que contém o músculo peitoral é um retângulo definido pelos pontos A, B, D e E como visto na Figura 1.

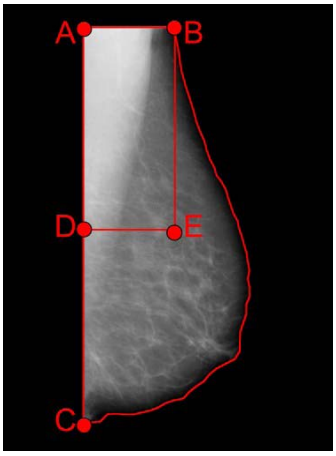


Figura 1. Delimitação da mama e da região de interesse contendo o músculo peitoral.

A região definida pelos pontos A, B, D e E pode não conter toda a extensão do músculo peitoral, mas com certeza englobará parte dele.

2.2. Binarização de Otsu

O método de Otsu [5] é considerado um método eficiente de binarização de imagens. O método assume que a imagem está dividida em duas classes, objeto e fundo, e através de múltiplas iterações tenta encontrar o limiar, nível de cinza que divide essas duas classes, com a máxima variância inter-classe. É um método automático que não necessita de parâmetros. A Figura 2 mostra uma imagem binarizada pelo método de Otsu.

Figura 2. Imagem binarizada pelo método de Otsu

2.3. Definição de Contorno

O trabalho de Nunes *et al.* [5], além da proposta principal de apresentar um aumento de contraste em imagens de mamas densas, apresenta um método para obtenção do contorno da mama.

Primeiramente, binariza-se a imagem para que a área contendo a mama diferencie-se do fundo da imagem (Figura 3.b). Em seguida localiza-se o ponto no início da mama e na metade de sua altura, o mesmo ponto D do método de Ferrari *et al.* [8]. A partir desse ponto traçam-se retas variando o coeficiente angular de 90º a -90º ou de 90 º a 270 º, dependendo da orientação da imagem (Figura 3.c). O ponto em que a reta passa de um pixel com valor 1 (mama) para um pixel com valor 0 (fundo) é considerado um dos vértices do contorno da mama. Finalmente, juntam-se todos os vértices encontrados em um polígono e é definido assim o contorno da mama (Figura 3.d). A Figura 3 ilustra as três etapas citadas acima.

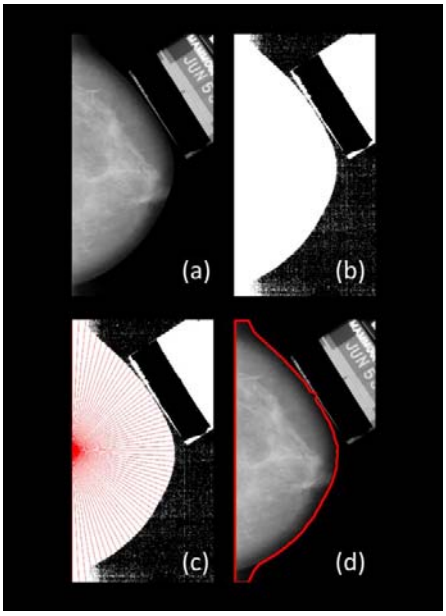


Figura 3. Etapas do método de delimitação do contorno da mama.

3. Materiais e Métodos

Nessa sessão será definido o método proposto bem como as ferramentas utilizadas para o trabalho, incluindo o banco de imagens utilizado e o protocolo de testes.

3.1. Plataforma de Desenvolvimento

O ImageJ é um software de código aberto voltado para análise e processamento de imagens. Por ser totalmente desenvolvido na linguagem Java pelo United States National Institutes of Health, o ImageJ é portável para quaisquer sistemas operacionais com Máquina Virtual Java [12]. O ImageJ fornece uma vasta quantidade de algoritmos para o processamento de imagens. Além disto, com sua arquitetura aberta, as funcionalidades do ImageJ podem ser facilmente expandidas através do desenvolvimento e incorporação de plugins, também escritos em Java. Por estes motivos, o presente trabalho foi produzido na forma de um plugin para o ImageJ.

3.2. Delimitação do Contorno da Mama

O presente trabalho propõe um novo método de segmentação do músculo peitoral utilizando uma associação de algumas técnicas descritas na sessão anterior. Primeiramente, utiliza-se o método de Nunes *et al.* [6] para obtenção do contorno da mama, informação necessária para as próximas etapas da segmentação. Para evitar que o usuário necessite entrar com o valor do limiar para binarizar a imagem, é aplicado um aumento de contraste logarítmico, encontrado anteriormente no trabalho de Ferrari *et al.* [13], seguido do método de Otsu [5]. O aumento de contraste diferencia ainda mais o fundo da imagem da mama e melhora o resultado da binarização de Otsu, fornecendo um limiar automático para o método de Nunes *et al.* [6]. A fórmula do aumento de contraste logarítmico pode ser vista na Equação 1 onde $I_{x,y}$ é o pixel da imagem original e $I'_{x,y}$ é o pixel da imagem com contraste.

$$I'_{x,y} = \log_2(1 + I_{x,y}) \qquad (1)$$

Apesar do resultado da delimitação do contorno da mama ter sido bem-sucedido para todas as imagens testadas, um deslocamento no contorno encontrado foi feito para que os pixels menos brilhantes presentes na interface mama-ar não atrapalhassem os cálculos envolvendo a binarização da imagem nas etapas posteriores. Testes empíricos mostraram que um retraimento de 100 pixels no contorno da mama elimina os pixels que estão na fronteira com o fundo da imagem. A Figura 4 mostra o contorno de um mamograma sem a redução (4.a) e com a redução de 100 pixels (4.b). As imagens possuem largura padrão, como explicado na sessão 3.4, de 600 pixels.

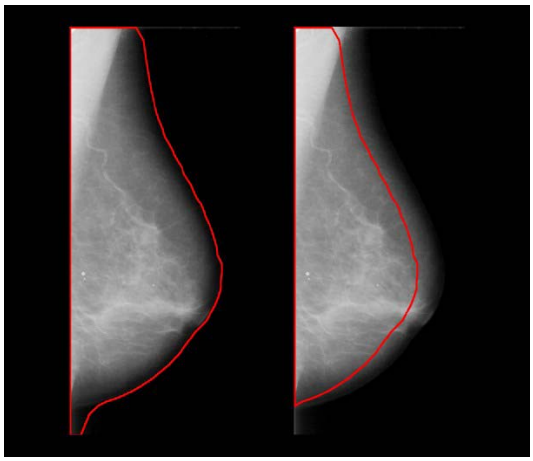


Figura 4. Contorno de uma mama sem a redução (a) e com a redução em 100 pixels (b).

3.3. Delimitação do Contorno do Músculo

Com a informação do contorno da mama, recorta-se a ROI contendo o músculo peitoral como descrito nos trabalhos de Ferrari *et al.* [8][9]. Aplica-se o filtro da mediana de dimensões 5x5 para suavizar a imagem e diminuir o ruído e em seguida aplica-se o método de Otsu [5] na ROI, obtendo o limiar que idealmente divide a imagem em duas classes, uma contendo os pixels pertencentes ao músculo peitoral, e a outra contendo os demais pixels. Com esse limiar, binariza-se a imagem original.

Uma vez com a imagem em sua forma binária, o músculo peitoral deve aparecer como uma única região branca e contínua. Mesmo que outras regiões como fibras ou outros artefatos mais densos também sejam detectados na binarização, é suficiente que a região do músculo não se mescle com outras regiões para que o resultado final seja bem-sucedido.

Após a binarização, é aplicada novamente a técnica de Nunes *et al.* [6] com o ponto inicial do algoritmo definido no ponto A do mamograma ao invés do ponto D. O motivo disso é que o ponto A é mais indicado para se obter o contorno do músculo peitoral devido a sua localização na imagem. Após a aplicação do método de Nunes, a informação dos vértices que compõem a região do músculo torna-se conhecida e basta juntá-los na forma de um polígono.

A Figura 5 ilustra as etapas citadas acima. A partir da imagem original (5.a) é recortada uma ROI contendo o músculo peitoral (5.b). É aplicado o método de Otsu na ROI (5.b) e o limiar obtido é usado na binarização da imagem (5.c). O método de Nunes et *al.* [6], é aplicado na imagem binarizada (5.d) resultando no contorno do músculo (5.e).

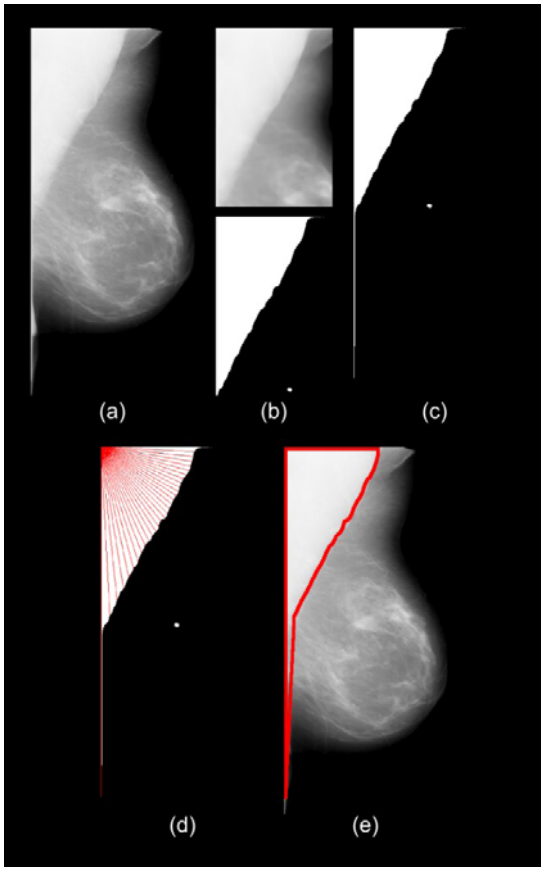


Figura 5. Etapas do método proposto pelo trabalho.

Em síntese, uma vez de posse da informação do contorno da mama, o método proposto nesse trabalho pode ser resumido em 5 etapas sendo elas:

1. Recorte da ROI contendo o músculo peitoral
2. Aplicação do filtro da mediana de 5x5 na ROI
3. Aplicação do método de Otsu na ROI e obtenção do limiar
4. Utilização do limiar encontrado no passo anterior para binarizar a imagem
5. Aplicação do método de Nunes *et al.* [6] na imagem binarizada para obtenção dos limites

3.4. Banco de Imagens

Foram utilizadas 40 imagens do volume BRCP_MASS_0 de incidência médio-lateral do banco de imagens mamográficas Digital Database for Screening Mammography (DDSM) [14]. As imagens mamográficas do DDSM possuem resolução de 16 bits e regiões de interesse contendo achados radiográficos delimitados por no mínimo dois especialistas, motivo pelo qual é amplamente usado na literatura. Uma vez delimitado o contorno da mama nas imagens, as imagens eram recortadas e redimensionadas para uma largura de 600 pixels a fim de reduzir o tempo de processamento dos testes.

3.5. Metodologia de Testes

Seguindo a mesma atribuição de notas de Saltanat *et al.* [15], um especialista (Luciana Buffa Verçosa) mastologista examinou o resultado de cada segmentação e atribuiu notas às imagens segmentadas variando de "Exata" até "Inadequada". Os resultados mostram a porcentagem de segmentações referentes a cada nota. A Tabela 1 mostra o sistema de notas traduzido do trabalho de Saltanat *et al.* [15].

Tabela 1. Sistemas de notas utilizadas

Nota	Descrição
Exata	A linha segmentada delimita a margem peitoral de forma exata. Qualquer desvio a partir da margem são imperceptíveis ou insignificante.
Muito Bom	A linha segmentada delimita a margem peitoral exatamente por pelo menos metade do seu comprimento e de forma adequada para a outra metade.
Adequada	A linha segmentada delimita a margem peitoral de forma inexata, mas com precisão suficiente para o fim pretendido.
Insuficiente	A linha segmentada delimita a margem peitoral inadequada por pelo menos metade do seu comprimento.
Inadequada	A linha segmentada está ausente ou é insuficiente para a localização da margem peitoral para o fim pretendido.

4. Resultados

O método proposto conseguiu segmentar com sucesso a maioria das imagens possuindo 32% de imagens segmentadas sob a classificação "Exata" e 42% imagens na classificação "Muito Bom". Considerando que imagens cuja nota foi "Adequada" ou superior são suficientes para prover uma redução na área processada em métodos de detecção de estruturas e extração de atributos, pode-se dizer que o sistema obteve 89% de segmentações corretas. A Figura 6 mostra as porcentagens das classificações obtidas na segmentação das imagens e a Figura 7 mostra alguns exemplos de segmentações.

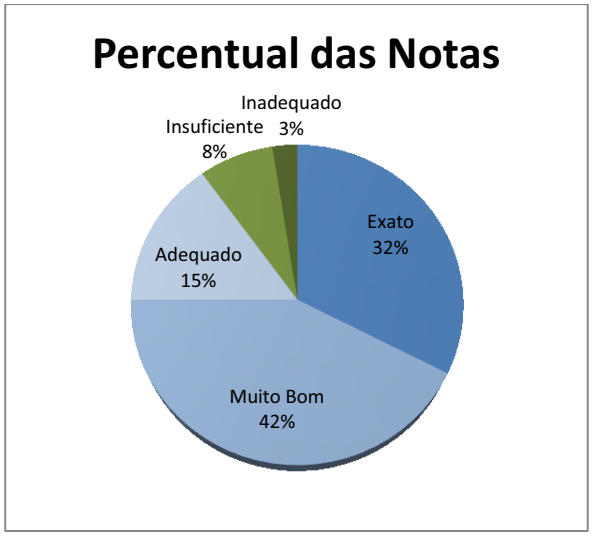


Figura 6. Porcentagem das notas referentes às segmentações.

5. Discussões

Durante a fase de implementação do método proposto, percebeu-se que o resultado final dependia principalmente da etapa de binarização da imagem. O limiar obtido através da aplicação do método de Otsu na ROI contendo o músculo peitoral deve separar, na etapa de binarização, o músculo peitoral do resto da imagem para que o método de Nunes *et al.* [6] obtenha adequadamente os contornos do músculo. Por isso a redução no contorno da mama é necessária: eliminando os pixels da borda da mama mais escuros permite que um limiar mais adequado seja obtido.

Uma comparação direta com os resultados obtidos por Saltanat *et al.* [15] pode não ser justa devido à utilização de bancos de imagens e especialistas diferentes. O presente trabalho apresenta uma segmentação correta de 89% das imagens enquanto o trabalho de Saltanat apresenta 98% se feito uma média entre os resultados obtidos por ambos os especialistas utilizados na avaliação do trabalho citado. Apesar de possuir resultados comparáveis visualmente com o trabalho de Saltanat, a aplicação de notas às imagens foi feito de forma mais rígida neste trabalho, considerando imagens inadequadas para uso aquelas cuja segmentação englobasse qualquer tecido mamário, mesmo que em pequena escala.

6. Conclusão

O trabalho expõe um método para a segmentação do músculo peitoral baseado processamentos feitos no domínio do espaço na imagem. A principal etapa do método consiste na binarização da imagem através de um limiar obtido pela aplicação do método de Otsu numa ROI que contém o músculo peitoral. Os resultados obtidos mostram que o método segmentou corretamente o músculo peitoral na maioria dos casos. Pode-se concluir, portanto, que o método proposto pode ser usado na construção de sistemas CADDx.

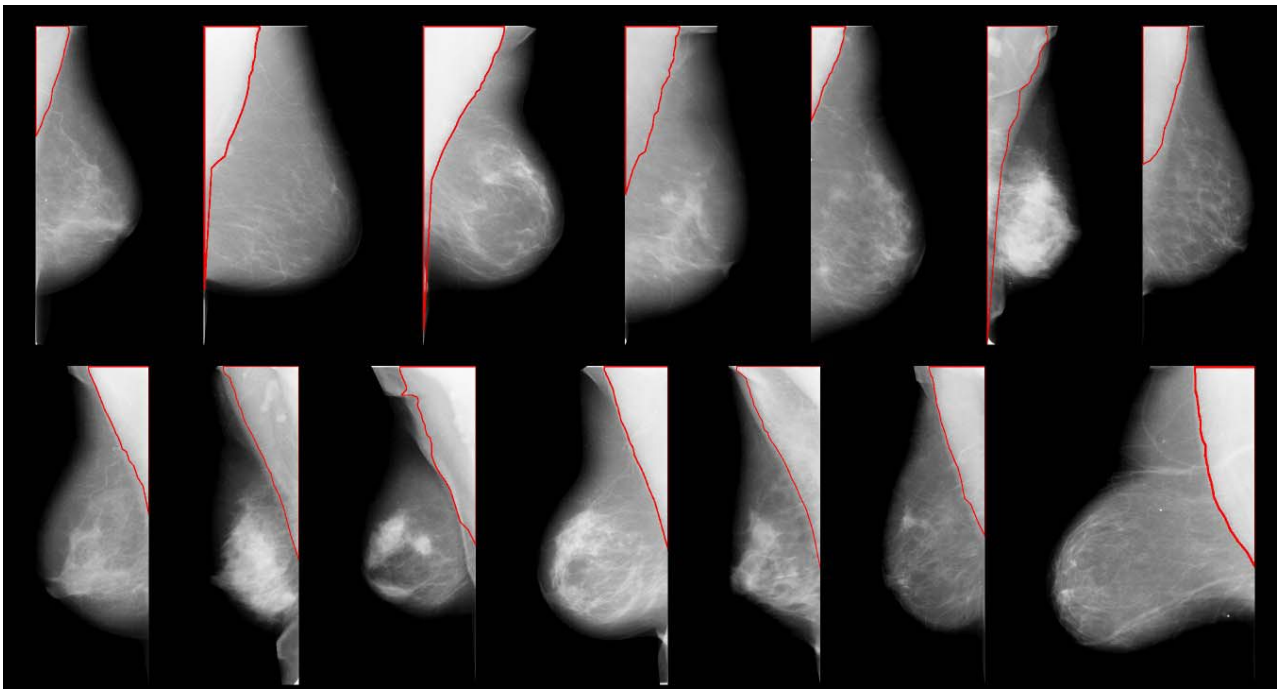


Figura 7. Exemplos de imagens segmentadas

O método pode ser aprimorado, com melhorias na obtenção do limiar usado na binarização da imagem. Os resultados podem ser mais precisamente avaliados calculando-se a taxa de pixels Verdadeiros Positivos, Falsos Positivos, Verdadeiros Negativos e Falsos Negativos sendo necessária uma delimitação prévia do especialista para esse tipo de avaliação. Tal abordagem de exibição dos resultados faz parte dos trabalhos futuros envolvendo a evolução do corrente trabalho.

Por fim, o método apresentado está sendo utilizado em um trabalho posterior dos mesmos autores como etapa de segmentação em um sistema de detecção automática de massas em mamografias.

10. Referências

[1] P. Boyle, B. Levin, "World Cancer Report", *International Agency for Research on Cancer*, Lyon, França, 2008, p. 412.

[2] Ministério da Saúde, Secretaria de Atenção à Saúde, Instituto Nacional de Câncer, Coordenação de Prevenção e Vigilância de Câncer. "Estimativas 2010: Incidência de Câncer no Brasil", Rio de Janeiro, Brasil, 2009, p. 29.

[3] G. M. Swanson, "Breast Cancer in the 1990's", *Journal of American Medical Women's Association*, v. 47, 1992, p. 140-148.

[4] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging", *British Journal of Radiology*, n. 78, 2005, p. 3-19.

[5] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms", *IEEE Transactions on Systems, Man, And Cybernetics*, v. 9, n. 1, 1979, p. 62-66.

[6] F. L. S. Nunes, H. Schiabel, R. H. Benatti, R. C. Stamato, M. C. Escarpinati, C. E. Goes, "A method to contrast enhancement of digital dense breast images aimed to detect clustered microcalcifications", Proceedings of International Conference On Image Processing, v. 1, Grécia, 2001, p. 305-308.

[7] N. Karssemeijer, M. Thijssen, J. Hendriks, L. van Erning, "Automated classification of parenchymal patterns in mammograms", Nijmegen, Physics in Medicine and Biology, v. 43, Inglaterra, 1998, p. 365-378.

[8] R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels, A. F. Frère, "Segmentação de Mamogramas: Identificação da borda da mama, músculo peitoral e disco glandular", Anais do XVII Congresso Brasileiro de Engenharia Biomédica, 2000.

[9] R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels, R. A. Borges, A. F. Frère, "Automatic Identification of the Pectoral Muscle in Mammograms", IEEE Transactions on Medical Imaging, v. 23, n. 2, 2004, p. 232-245.

[10] S. M. Kwok, R. Chandrasekhar, M. T. Rickard, "Automatic Pectoral Muscle Segmentation on Mediolateral Oblique View Mammograms", IEEE Transactions on Medical Imaging, v. 23, n. 9, 2004, p. 1129-1140.

[11] S. Tzikopoulos, H. Georgiou, M. Mavroforakis, N. Dimitropoulost, S. Theodoridis, "A Fully Automated Complete Segmentation Scheme for Mammograms", Anais da 16th International Conference on Digital Signal Processing, Grécia, 2009.

[12] United States National Institutes of Health. "ImageJ", Estados Unidos. Disponível em http://rsbweb.nih.gov/ij/. Acesso em 15 de junho de 2010.

[13] R. J. Ferrari, "Detecção de Assimetrias entre Mamogramas", Tese de Doutorado, São Carlos, 2002.

[14] M. Heat, K. Bowyer, D. Kopans, R. Moore, P. Kegelmeyer, "The Digital Database for Screening Mammography", Proceedings of the 5th International Workshop on Digital Mammography, Toronto, 2000, p. 212-218.

[15] N. Saltanat, M. A. Hossain, M. A. Alam, "An Efficient Pixel Value Based Mapping Scheme to Delineate Pectoral Muscle from Mammograms", Anais da Fifth International Conference on Bio-Inspired Computing: Theories and Applications, China, 2010, p. 1510-1517.

Segmentação Não-Supervisionada de Imagens em Duas Regiões Baseada no Método Competição entre Regiões Fuzzy utilizando Distribuições de Probabilidade

Vinicius R. P. Borges[1], Celia A. Zorzo Barcelos[2], Marcos Aurelio Batista[3], Karla B. Freitas[4]
Faculdade de Computação[1], Faculdade de Matemática[2] [4], Departamento de Ciência da Computação[3]
Universidade Federal de Uberlândia[1] [2] [4], Universidade Federal de Goiás[3]
viniciusrpb@mestrado.ufu.br, celiazb@ufu.br, marcos@catalao.ufg.br, karla@pontal.ufu.br

Resumo

In this paper, we propose an unsupervised variational two-phase image segmentation model based on Fuzzy Region Competition that uses probability distributions to design image regions and to guide region competition by the domain points through log-likelihood test. The functional involved in the proposed model is convex related to the fuzzy membership function, so global solutions can be computed and stable segmentation algorithms can be derived from the refered functional. The experiments performed in the proposed model showed noise robustness and presented good segmentation results when dealing with natural images.

1. Introdução

A segmentação de imagens é uma área fundamental em processamento de imagens e visão computacional. Apesar dos intensos estudos nas últimas décadas, a segmentação de imagens é ainda um importante tópico de pesquisa com muitos problemas em aberto a serem resolvidos. Seu objetivo é particionar o domínio de uma imagem em seus sub-domínios constituintes. Existe uma grande variedade de algoritmos para segmentação de imagens e pesquisadores tem concentrado esforços no desenvolvimento de métodos eficientes e consistentes. Técnicas que incorporam princípios variacionais surgem como ferramentas eficazes para resolver diversos problemas relacionados. Em uma abordagem variacional, a segmentação é obtida ao minimizar um funcional de energia formulado sob um critério de homogeneidade das regiões da imagem.

Com a popularização do método de evolução de curvas dinâmicas *Contornos Ativos* [6], e do funcional *Mumford-Shah* [12], útil na remoção de ruídos e na segmentação de imagens, diversos métodos variacionais baseados em região e em contorno tem sido propostos com o intuito de se obter algoritmos mais robustos e eficazes. *Chan* e *Vese* [4] exploraram o caso constante do funcional *Mumford-Shah*, juntamente com métodos Level Sets [13] a fim de se obter um tipo de contorno ativo que não utiliza operador gradiente. *Zhu* e *Yuille* [16] desenvolveram o método *Competição entre Regiões*, onde as regiões da imagem competem entre si pelos pontos do domínio mediante análise estatística. O trabalho de *Zhu* e *Yuille* [16] foi um dos primeiros a estabelecer uma relação entre o caso constante do método *Mumford-Shah* e o modelo de segmentação baseado no princípio máximo *a posteriori* de *S. Geman* e *D. Geman* [5]. Assim, a técnica *Competição entre Regiões* pode ser vista como uma generalização do caso constante de *Mumford-Shah*, onde as aproximações constantes das regiões da imagem são substituídas por distribuições de probabilidade [2].

Posteriormente, outros trabalhos foram desenvolvidos baseados na idéia apresentada por *Zhu* e *Yuille* [16]. *Rousson* e *Deriche* [15] generalizaram o modelo *Chan-Vese* [4] ao formular um princípio de probabilidade de segmentação máximo *a posteriori* para segmentação de imagens em duas regiões, onde as funções de densidade de probabilidade de cada região fornecem as informações para a evolução de uma curva dinâmica *level set* pelo domínio. *Brox* e *Cremers* [2] estudaram as relações existentes entre a estatística Bayesiana e o funcional *Mumford-Shah* suave, onde um modelo Bayesiano com distribuições de densidade de probabilidade estimadas localmente são utilizadas para aproximar o funcional *Mumford-Shah*. *Paragios* e *Deriche* propuseram o método supervisionado *Regiões Geodésicas Ativas* [14] para segmentar imagens texturizadas ao determinar uma força proveniente das energias da suavização, da região e do contorno para movimentar uma curva dinâmica *level set* em direção às bordas do objeto de interesse da imagem. *Kim et. al* [7] propuseram uma técnica estatística de segmentação de imagens em

várias regiões formulada juntamente com *level sets*, onde o movimento desta curva ocorre mediante critérios estatísticos calculados com base na estimação das funções de densidade de probabilidade não-paramétricas. Todos estes métodos variacionais para segmentação de imagens citados calculam soluções locais e apresentam um processo de convergência lento, pois são baseados em técnicas de evolução de curvas, que são numericamente resolvidas por meio das Equações de Euler-Lagrange associadas com esquema gradiente descendente.

Mory e *Ardon* [10] propuseram uma extensão da técnica *Competição entre Regiões* para uma abordagem $fuzzy$ ($soft$), denominada *Competição entre Regiões Fuzzy*, onde as regiões da imagem competem entre si pelos pontos do domínio mediante algum critério de homogeneidade de região, podendo ser proveniente de princípios Bayesianos [16], da técnica *Mumford-Shah* constante por partes ou suave por partes [12], além de possuírem uma abordagem supervisionada ou não-supervisionada. O funcional de energia do método *Competição entre Regiões Fuzzy* é convexo com relação à função de pertinência, logo, soluções globais são obtidas nos processos de segmentação das imagens. Posteriormente, *Mory* e *Ardon* [11], e *Li* e *Ng* [8] estudaram o uso de funções de densidade de probabilidade não-paramétricas no modelo *Competição entre Regiões Fuzzy*, onde estas funções de densidade desconhecidas são estimadas via *densidade de Kernel* enquanto ocorre o processo de segmentação.

Também em [10], foi apresentado um modelo de segmentação supervisionado baseado no método *Competição entre Regiões Fuzzy* [10], onde assume-se que cada região da imagem é aproximada por uma distribuição de probabilidade e que os valores de seus parâmetros são conhecidos *a priori*. Apesar deste modelo supervisionado apresentar bons resultados, existe o incoveniente no que se refere a necessidade de se utilizar um agente externo no processo de segmentação, para amostrar cada região da imagem, o que na prática pode limitar a aplicabilidade da técnica. Além disso, a estimação dos parâmetros de amostras obtidas *a priori* das regiões da imagem pode resultar em valores que não representam significativamente toda a região.

Diferentemente de [10], neste trabalho, propõe-se um modelo não-supervisionado para segmentar imagens em duas regiões baseado no método *Competição entre Regiões Fuzzy* que utiliza a formulação estatística descrita na técnica *Competição entre Regiões*. Os parâmetros das distribuições de probabilidade são otimizados durante o processo de segmentação da imagem, sem qualquer informação obtida *a priori* das regiões a serem identificadas na imagem. Este artigo está organizado da seguinte maneira: a Seção 2 apresenta o funcional de energia do método *Competição entre Regiões Fuzzy* e como foi construído seu funcional convexo. A Seção 3 descreve o novo modelo proposto, que utiliza a mesma idéia do caso supervisionado *Competição entre Regiões Fuzzy*, mas que foi formulado sob uma abordagem não-supervisionada. A Seção 4 descreve os experimentos realizados e discute seus resultados. A Seção 5 apresenta as conclusões sobre o modelo proposto.

2. Competição entre Regiões Fuzzy

Com o sucesso da técnica estatística *Competição entre Regiões* [16], diversos métodos variacionais baseados em região e que utilizam princípios estatísticos tem sido propostos. Uma formulação variacional para genérica para a segmentação de uma imagem I em duas regiões é dada pelo seguinte problema de minimização:

$$\min_{\Omega_1,\boldsymbol{\alpha}} \left\{ F_{2R}(\Omega_1,\boldsymbol{\alpha}) = \int_\Gamma g(\Gamma(s))ds + \int_{\Omega_1} r_1(\alpha_1,x)dx + \int_{\Omega\setminus\Omega_1} r_2(\alpha_2,x)dx \right\} \quad (1)$$

onde $\Omega \subset \mathbb{R}^m$ é o domínio da imagem, $\Omega_1 \subset \Omega$ representa o objeto da imagem, $\Omega_1\setminus\Omega$ representa o fundo da imagem e $\Gamma = \partial\Omega_1$ é a fronteira do domínio Ω_1. A função g é um potencial de bordas definido como uma função positiva e decrescente com relação ao gradiente da imagem. As funções erro $r_i : \Omega \to \mathbb{R}$ possuem informações relativas à intensidade de cada região, que são modeladas utilizando-se análise estatística sobre o conjunto de parâmetros $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ e podem ser escalares [16], constantes [4] ou funções que variam no espaço [12]. Caso os parâmetros ótimos $\boldsymbol{\alpha}$ sejam conhecidos *a priori*, o problema (1) é supervisionado, caso contrário, a segmentação é não-supervisionada, onde $\boldsymbol{\alpha}$ deve ser otimizado.

O problema de minimização (1) pode ser resolvido utilizando-se integrais de linha, ou através de integrais sobre todo o domínio Ω, mediante o uso de funções características [10] [4]. Nestes casos, a solução numérica é obtida por meio de métodos *Level Sets* [13] utilizando-se Equações de Euler-Lagrange associadas com um esquema gradiente descendente para a evolução da curva *level set*. No entanto, esta estratégia é mal-posta na função *level set* e é limitada no sentido de que o espaço de otimização é não-convexo, resultando em modelos de segmentação que calculam soluções locais dependentes das condições iniciais. Estes contratempos motivaram o desenvolvimento de esquemas convexos para minimização global de funcionais de energia regularizados por variação total.

O funcional (1) pode ser reescrito por meio de integrais sobre todo o domínio utilizando funções características, tornando-se

$$\min_{\chi,\boldsymbol{\alpha}} \left\{ \hat{F}_\chi(\chi,\boldsymbol{\alpha}) = \int_\Omega g(x)|\nabla\chi(x)| + \int_\Omega \chi(x) r_1(\alpha_1,x)dx + \int_\Omega (1-\chi(x)) r_2(\alpha_2,x)dx \right\} \quad , \quad (2)$$

onde χ é uma função característica de um conjunto $\Omega_1 \subset \Omega$, tal que $\chi(x) = 1$ se $x \in \Omega_1$, ou $\chi(x) = 0$, caso contrário. Visto que a formulação (2) ainda é não-convexa, *Mory* e *Ardon* [10] propuseram substituir a função característica χ por uma função de pertinência $fuzzy$ u pertencente ao conjunto convexo $BV_{[0,1]}(\Omega)$ (conjunto das funções com variação limitada em $[0,1]$). Desta maneira, o problema de minimização (2) é reescrito como

$$\min_{u\in BV(\Omega)_{[0,1]},\boldsymbol{\alpha}} \left\{ F_{CRF}(u,\boldsymbol{\alpha}) = \int_\Omega g|\nabla u| + \int_\Omega u(x) r_1(\alpha_1,x)dx + \int_\Omega (1-u(x)) r_2(\alpha_2,x)dx \right\} \quad , \quad (3)$$

obtendo-se, portanto, o método *Competição entre Regiões Fuzzy* (CRF), onde o primeiro termo do funcional de energia da equação (3) é responsável pela suavização na variável u, enquanto que os termos restantes se referem ao custo de atribuir cada ponto do domínio para as suas respectivas regiões. Nesta forma, o problema de minimização (3) é convexo na função u e em um caso de segmentação supervisionada (quando $\boldsymbol{\alpha}$ é mantido fixo), soluções globais ótimas são obtidas ao minimizar o funcional F_{CRF} em relação a u. Em problemas de segmentação não-supervisionada, a função u e o conjunto de parâmetros $\boldsymbol{\alpha}$ são otimizados alternadamente em sucessivas etapas. Neste caso, a convexidade em u não garante a obtenção de soluções ótimas, mas modelos de segmentação estáveis e fracamente dependente das condições iniciais podem ser derivados do problema de minimização (3).

Como foi dito anteriormente, as funções erro r_i podem assumir valores escalares, constantes, ou funções que variam no espaço. *Mory* e *Ardon* [10] utilizaram uma função erro derivada de um critério MDL (*Mimimum Description Length*) e que foi baseada na proposta de *Zhu et. al* [16], dada por

$$r_i(\alpha_i,x) = -\lambda log(P_i(I(x)|\alpha_i)) \qquad \forall x \in \Omega, \quad (4)$$

onde λ é um parâmetro que balanceia a suavização e a regularização no problema de minimização (3).

Mory e *Ardon* desenvolveram em [10] um modelo de segmentação supervisionado que utiliza a função erro (4), onde deve-se amostrar cada região da imagem, calcular os respectivos parâmetros com base nas informações das amostras e determinar as distribuições de probabilidade para estas regiões, que serão utilizadas no processo de competição entre as regiões mediante teste de log-verossimilhança para cada ponto do domínio.

Apesar de que bons resultados podem ser obtidos a partir do modelo *Competição entre Regiões Fuzzy* supervisionado, descrito em [10], deve-se ressaltar que o processo de segmentação não é totalmente automático, pois um agente externo é necessário para amostrar as regiões do objeto e do fundo da imagem para determinar os parâmetros ótimos das respectivas distribuições de probabilidade. Além disso, a estimação dos parâmetros a partir de amostras de uma região da imagem pode resultar em valores que não representam significativamente toda a região, o que pode acarretar em resultados insatisfatórios na segmentação final.

Alternativamente ao modelo supervisionado, foram apresentados em [10] modelos não-supervisionados para segmentação de imagens baseados no método *Competição entre Regiões Fuzzy*. Nestes modelos, funções erro baseadas no funcional *Mumford-Shah* suave por partes e constantes por partes [12] foram introduzidas no funcional (3), mas que diferentemente do modelo supervisionado, têm seus parâmetros otimizados simultaneamente à minimização da função u. Apesar de apresentarem bons resultados em imagens ruidosas e em imagen com regiões homogêneas, estes modelos não são indicados para segmentar imagens naturais ou imagens que contém algum tipo de textura, pois as regiões podem ser representadas apenas por uma constante ou por uma aproximação suave para a referida região.

Assim, na Seção seguinte, será apresentado um modelo de segmentação que utiliza análise estatística em sua formulação e que otimiza os parâmetros das distribuições de probabilidade durante o processo de segmentação, caracterizando o modelo proposto como não-supervisionado.

3. Modelo Proposto

Na Seção anterior, foi dada uma motivação para o desenvolvimento deste modelo e a função erro estatística para caracterizar as regiões da imagem. Substituindo-se a função erro (4) no funcional da equação (3), o problema de minimização obtido é

$$\min_{u\in BV(\Omega)_{[0,1]},\boldsymbol{\alpha}} \left\{ F_{ECRF}(u,\boldsymbol{\alpha}) = \int_\Omega g|\nabla u| - \lambda \int_\Omega u(x) log(P_1(I(x)|\alpha_1))dx - \lambda \int_\Omega (1-u(x)) log(P_2(I(x)|\alpha_2))dx \right\} \quad , \quad (5)$$

onde assume-se que as distribuições de probabilidade $P_i(I|\alpha_i)$ são previamente conhecidas. Neste trabalho, utilizou-se a distribuição de probabilidade Gaussiana, cujos parâmetros são a média e o desvio-padrão no caso unidimensional, ou a média e a matriz de co-variância no caso d-dimensional (imagens com d canais de cores) [16] [15]. O problema de minimização (5) pode ser resolvido de acordo com o algoritmo abaixo:

1. Inicialize a função de pertinência u aleatoriamente.
2. Enquanto u não convergir, faça
   (a) Mantendo-se u fixo, calcule valores ótimos para os parâmetros α_1 e α_2. Determine as distribuições de probabilidade $P_1(I(x)|\alpha_1)$ e $P_2(I(x)|\alpha_2)$ utilizando-se dos respectivos parâmetros calculados;
   (b) Mantendo-se α_1 e α_2 fixos, atualize a função de pertinência u minimizando-se o funcional (5) em relação a u;

As Equações de Euler-Lagrange em relação à equação (5) podem ser tipicamente utilizadas para o cálculos do parâmetros α_1 e α_2 (passo (a)) e para atualizar a função de pertinência u (passo (b)). No entanto, a minimização por esta estratégia não tira nenhuma vantagem da convexidade do funcional presente em (5). Alternativamente, a minimização deste problema pode ser feita eficientemente pelo esquema de projeção $dual$ de *Chambolle* [3]. Para utilizar este esquema de minimização, utilizaremos a estratégia descrita por *Bresson et. al* [1], onde o problema de segmentação (3) é aproximado por

$$\min_{(u,v)\in BV(\Omega)_{[0,1]}} \left\{ F_{ap}(u,v) = \int_\Omega g|\nabla u| + \frac{1}{2\theta}\int_\Omega |u-v|^2 + \int_\Omega rv + \rho\nu(v) \right\} \quad , \tag{6}$$

dado que v é uma função auxiliar, $r = r_1 - r_2$ é a função de competição entre regiões, $\nu(\varepsilon) = max(0, |2\varepsilon - 1| - 1)$, $\rho > \frac{1}{2}|r|_\infty$ e θ é um valor suficientemente pequeno para que o par minimizador (u^*, v^*) seja quase idêntico em relação a norma L^2. Como a função u é dependente apenas nos dois primeiros termos do funcional da equação (6), o algoritmo de projeção $dual$ de *Chambolle* pode ser utilizado para a minimização de u. Baseado nisso, a estratégia de minimização do funcional (6) consiste em executar sucessivas etapas de minimização alternadamente para u e v como se segue:

1. Mantendo-se u fixo, calcula-se v:

$$v = max\{min\{u - \theta r, 1\}, 0\} \quad , \tag{7}$$

   sendo que $r = \lambda log(P_2/P_1)$.

2. Mantendo-se v fixo, calcula-se u:

$$u = v - \theta div\mathbf{p} \quad , \tag{8}$$

   onde o vetor $\boldsymbol{p} = (p^1, p^2) \in \mathcal{C}_c^1(\Omega, \mathbb{R}^m)$ pode ser calculado pelo algoritmo de ponto fixo, iterando-se para $n \geq 0$,

$$\mathbf{p}^{n+1} = \frac{\mathbf{p}^n + \tau\nabla(div\mathbf{p}^n - v/\theta)}{1 + \tau|\nabla(div\mathbf{p}^n - v/\theta)|/g} \quad , \tag{9}$$

   em que $\boldsymbol{p}^0 = \mathbf{0}$, $\tau < 1/8$ garante a estabilidade numérica deste esquema e g é a mesma função apresentada na equação (5), dada por:

$$g(|\nabla I|) = \frac{1}{1 + \beta|\nabla I|^2} \quad , \tag{10}$$

   onde β é uma constante positiva responsável por controlar a influência do gradiente da imagem.

O esquema de projeção $dual$ de *Chambolle* apresenta estabilidade numérica no processo de minimização e é extremamente vantajoso para o processo de convergência.

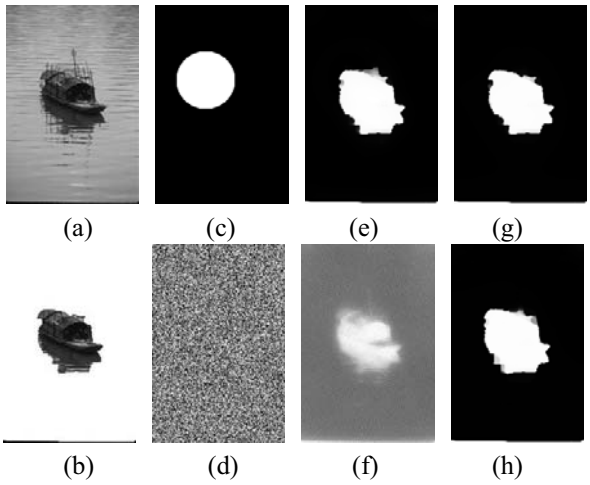
Na próxima seção, serão apresentados alguns experimentos realizados utilizando o modelo proposto.

4. Resultados

Nesta seção, o modelo proposto foi avaliado utilizando-se de várias imagens da base de dados *Berkeley Segmentation Dataset* [9]. Como este modelo segmenta uma imagem em duas regiões, foram escolhidas da referida base de dados apenas imagens constituídas de objeto e fundo, e cinco delas são mostradas neste trabalho. Os parâmetros utilizados foram $\theta = 0.15$ (presente nas equações (9) e (8)), $\tau = 0.1$ (equação 9) e 10 foi número de iterações no algoritmo de ponto fixo de *Chambolle*. Já os parâmetros, λ (equação 7) e β equação (10) devem ser fixados de acordo com a imagem a ser segmentada.

O primeiro experimento tem como objetivo verificar a sensibilidade do modelo em relação às condições iniciais. Deste modo, a função de pertinência u foi iniciada de duas maneiras distintas. A Figura 1(a) ilustra a imagem original. A Figura 1(b) mostra a segmentação final. As Figuras 1(c-d) e 1(e-f) mostram a função de pertinência u nos estados inicial e intermediário de tempo, respectivamente, para cada inicialização distinta da função de pertinência u. As Figuras 1(g-h) apresentam as funções de pertinência u finais. Por estes resultados, pode-se perceber que o modelo proposto é fracamente dependente das condições iniciais, pois a partir de diferentes inicializações, um resultado final semelhante foi obtido. Os parâmetros utilizados foram: $\lambda = 0.25$ e $\beta = 0.0001$.

A Figura 2 ilustra um exemplo onde deseja-se segmentar o elefante. A Figura 2(a) ilustra a imagem original. As

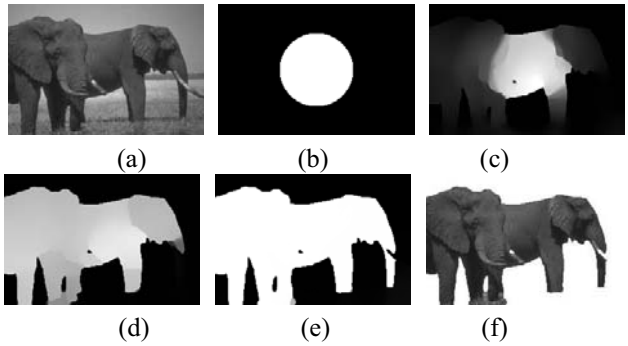


(a) (c) (e) (g)

(b) (d) (f) (h)

Figura 1. Imagem de um barco 481×321 **[9]: (a) Imagem original; (b) Imagem segmentada; (c-d) Duas funções** u **iniciais e distintas; (e-f) Respectivas** u **em estados intermediários; (g-h) Respectivas** u **finais.**

Figuras 2(b), 2(c-d) e 2(e) mostram a função de pertinência nos estados inicial, intermediários e final, respectivamente. A Figura 2(f) apresenta o elefante segmentado da imagem. Os parâmetros utilizados foram: $\lambda = 0.075$ e $\beta = 0.001$.

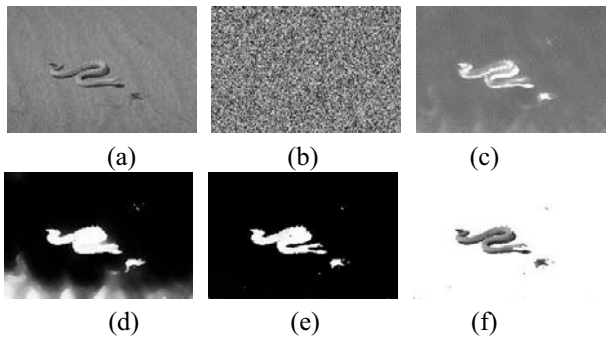


(a) (b) (c)

(d) (e) (f)

Figura 2. Imagem de um elefante 321×481 **[9]: (a) Imagem original ; (b)** u **inicial ; (c-d)** u **em estados intermediários (**50 **e** 200 **iterações no tempo, respectivamente) ; (e)** u **final (após** 800 **iterações) (f) Imagem segmentada.**

No experimento mostrado pela Figura 3, o objetivo é segmentar a cobra da imagem. A Figura 3(a) ilustra a imagem original. As Figuras 3(b), 3(c-d) e 3(e) mostram a função de pertinência nos estados inicial, intermediários e final de tempo, respectivamente. A Figura 3(f) apresenta a imagem segmentada final contendo apenas a cobra. Os parâmetros utilizados foram: $\lambda = 0.4$ e $\beta = 0.005$.



(a) (b) (c)

(d) (e) (f)

Figura 3. Imagem de uma cobra 321×481 **[9]: (a) Imagem original; (b)** u **inicial; (c-d)** u **em estados intermediários (**20 **e** 60 **iterações no tempo, respectivamente); (e)** u **em estado estacionário (após** 700 **iterações); (f) Imagem segmentada.**

O experimento representado pela Figura 4 tem como objetivo avaliar a robustez do método em relação à ruidos. O que se espera é que o modelo proposto identifique corretamente as estruturas geométricas da imagem corrompida por ruído gaussiano. A Figura 4(a) apresenta a imagem original. As Figuras 4(b), 4(c) e 4(d) mostram a função de pertinência nos estados inicial, intermediário e final de tempo, respectivamente. Os parâmetros utilizados foram: $\lambda = 0.85$ e $\beta = 0$.

(a) (b) (c) (d)

Figura 4. Imagem ruidosa 300×300**: (a) Imagem original; (b)** u **inicial; (c)** u **intermediária (**50 **iterações no tempo); (d)** u **final (após** 400 **iterações).**

Na Figura 5, deseja-se segmentar o avião da imagem. A Figura 5(a) ilustra a imagem original. As Figuras 5(b), 5(c-d) e 5(e) mostram a função de pertinência nos estados inicial, intermediários e final de tempo, respectivamente. A Figura 5(f) apresenta a imagem segmentada final contendo

apenas o avião. Os parâmetros utilizados foram: $\lambda = 0.05$ e $\beta = 0.00075$.

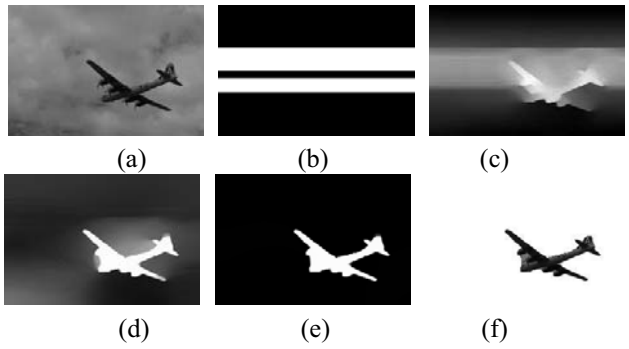


Figura 5. Imagem de um avião 321×481 **[9]: (a) Imagem original; (b)** u **inicial; (c-d)** u **intermediária (**50 **e** 110 **iterações no tempo, respectivamente); (e)** u **final (após** 400 **iterações); (f) Imagem segmentada.**

5. Conclusão

Neste artigo, foi proposto um modelo variacional não-supervisionado para segmentação de imagens em duas regiões baseado no método *Competição entre Regiões Fuzzy* por uma abordagem estatística. No modelo proposto, são utilizados distribuições de probabilidade para descrever as regiões da imagem em termos de suas intensidades e a competição entre as regiões ocorre mediante teste de log-verossimilhança. Além disso, o funcional do modelo proposto é convexo em relação à função de pertinência, e devido a este fato, utilizou-se um eficiente esquema de minimização que fornece soluções numéricas estáveis e que é pouco sensível em relação às condições iniciais. Os resultados mostraram que o modelo proposto é robusto em relação à presença de ruídos nas imagens e que apresenta bons resultados em imagens naturais caracterizadas por possuírem textura.

6. Agradecimentos

Gostaríamos de agradecer ao Conselho Nacional de Pesquisa e Desenvolvimento (CNPq) pelo suporte financeiro (Proc. # 551852/2010-0).

Referências

[1] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28:151–167, 2007.

[2] T. Brox and D. Cremers. On the statistical interpretation of the piecewise smooth mumford-shah functional. *Proc. International Conference on Scale Space and Variational Methods in Computer Vision*, 4485:203–213, May 2007.

[3] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20:89–97, 2004.

[4] T. Chan and L. Vese. Active contour model without edges. *IEEE Transactions on Image Processing*, 102:266–277, 2001.

[5] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[6] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[7] J. Kim, I. Fisher, J.W., A. Yezzi, M. Cetin, and A. Willsky. A nonparametric statistical method for image segmentation using information theory and curve evolution. *Image Processing, IEEE Transactions on*, 14(10):1486 –1502, 2005.

[8] F. Li and M. K. Ng. Kernel density estimation based multiphase fuzzy region competition method for texture image segmentation. *Commun. Comput. Phys.*, 8:623–641, 2010.

[9] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2:416–423, 2001.

[10] B. Mory and R. Ardon. Fuzzy region competition: A convex two-phase segmentation framework. *Scale Space and Variational Methods in Computer Vision*, 4485:214–226, 2007.

[11] B. Mory, R. Ardon, and J.-P. Thiran. Variational segmentation using fuzzy region competition and local non-parametric probability density functions. *IEEE 11th International Conference on Computer Vision. ICCV*, pages 1–8, 2007.

[12] D. Mumford and J. Shah. Optimal approximations by piecewise smooth. functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 2:577–685, 1989.

[13] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, 1988.

[14] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46(3):223–247, 2002.

[15] M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. *Motion and Video Computing, IEEE Workshop on*, pages 56–61, 2002.

[16] S. C. Zhu and A. L. Yuille. Region competition: unifying snakes, region growing, energy/bayes/mdl for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884–900, 1996.

Sistema de Visão Computacional para Classificação de Grãos de Café por Cor e Forma

Pedro Ivo de Castro Oyama[1]; Lúcio André de Castro Jorge[2]; Carlos Cesar Gomes[3]

1- Bolsista, Embrapa Instrumentação, São Carlos – SP, pedro.oyama@gmail.com
2- Pesquisador, Embrapa Instrumentação, São Carlos – SP, lucio@cnpdia.embrapa.br
3- Analista de qualidade, Cooperativa de Café Cooxupé, Guaxupé – MG, cesar@cooxupe.com.br

Abstract

The present work aims to develop a vision system for coffee beans classification of small samples to be used in the lab of coffee quality analysis of the COOXUPÉ, cooperative of coffee producers from Guaxupé, Brasil, in order to replace the current visual inspection made by human specialist, which is slow and is often unable to fulfill the demands of the coffee industry. Moreover, for being made by humans, these methods are more susceptible to failure. The system is a machine vision that uses image processing algorithms to extract color and shape beans attributes from the image samples. An artificial neural network Multilayer Perceptron was used to recognize patterns of color and shape from images of beans. It was possible to accurately classify coffee beans but an improvement in the illumination system is necessary.

1. Introdução

Com o aumento da produção e da demanda pelo produto de qualidade, cada vez mais o número de análises rápidas e precisas dos atributos físicos e fisiológicos dos grãos tem se tornado o gargalo na tomada de decisões rápidas em tempo compatível com o processo produtivo [1].

A análise dos atributos físicos de um lote de grãos é baseada principalmente em caracteres morfológicos das frações componentes da amostra. Dentre os atributos físicos do lote destacam-se as dimensões, a forma, a presença de impurezas tais como restos vegetais, pedras, partículas de solo, frações de grãos menores que sua metade e etc. [5]. A análise física em grãos é um procedimento moroso e dependente da interpretação do analista, pois hoje é feita manualmente.

Os atributos fisiológicos de um lote de sementes devem indicar a sua capacidade de germinar e o seu vigor. No caso do grão, servem para identificar principalmente defeitos baseados na cor da amostra.

Vários testes para avaliação dos atributos físicos e fisiológicos de um lote de sementes são relatados e descritos em [2], [3] e [4] e podem ser aplicados aos grãos.

Com o avanço na área de visão computacional, com a criação de sensores mais rápidos e processamentos mais eficientes, podem ser encontrados trabalhos de classificação de sementes de soja com base na forma [6] e a análise de cor para a separação de grãos em máquinas utilizadas pela indústria alimentícia [7], dentre outros.

Em geral, as aplicações aparecem em máquinas e sistemas para aplicação no processo de produção de produtoras de sementes. No laboratório, a análise de amostras ainda continua sendo feita de forma subjetiva e manual.

Nesse contexto, está sendo desenvolvida uma máquina para a classificação automática de grãos de café através de imagens digitais e neste trabalho serão apresentados os primeiros resultados da classificação feita através da aplicação de algoritmos de processamento de imagem numa amostra de grãos de café com base nos atributos de cor e forma.

2. Metodologia Utilizada

A metodologia utilizada neste trabalho é dividida em: aquisição de imagens da amostra; aplicação de algoritmos de processamento de imagem para detecção dos objetos e extração de suas características; classificação por padrões de cor; definição de descritores de forma; e, finalmente, classificação por padrões de forma.

2.1 Amostragem e Classificação

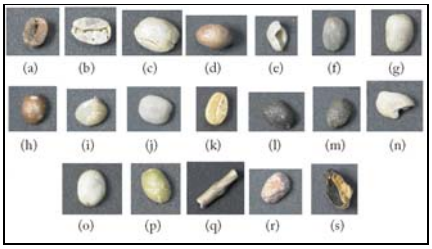


Figura 1. Classes de grãos e impurezas. (a) ardido, (b) brocado, (c) cabeça, (d) côco, (e) concha, (f) fava escura, (g) marago, (h) melado, (i) película, (j) perfeito, (k) pergaminho, (l) preto, (m) preto-verde, (n) quebrado, (o) secador, (p) verde, (q) paus, (r) pedras e (s) cascas.

Foram utilizadas amostras de grãos de diferentes peneiras fornecidas pela Cooperativa Regional de Cafeicultores de Guaxupé Ltda. (Cooxupé), juntamente com os grãos com defeitos e impurezas já separados e classificados por especialistas da cooperativa.

As classes de defeitos e de grãos sadios são apresentadas na Figura 1. As impurezas de uma amostra são paus, pedras e cascas.

2.2 Aquisição de imagens

Para a aquisição de imagens das amostras de grãos de café com iluminação padrão, foi desenvolvida uma máquina de classificação (Figura 2), composta de uma bandeja de vidro – onde são depositados os grãos sob condições controladas de luz. A bandeja é iluminada por duas lâmpadas fluorescentes de 20W por cima e duas por baixo, duas câmeras de vídeo posicionadas cerca de 60cm acima e abaixo da amostra. A câmera de vídeo utilizada é do tipo IP e é conectada ao computador que executa o software através de uma interface *Gigabit Ethernet*. As imagens são capturadas diretamente pelo software e são salvas no formato JPG com resolução de 1280x1024 *pixels*.

(a) (b)

Figura 2: (a) Representação gráfica da máquina de classificar café e (b) Foto da primeira versão construída

Na Figura 3 pode ser observada a tela de captura da imagem, onde são possíveis ajustes de brilho, contraste e iluminação da amostra.

Figura 3: Interface de controle de aquisição

2.3 Processamento das Imagens

O software Qualicafé desenvolvido foi dividido em módulos: pré-processamento, classificação por cor, por forma e por tipo de grão, conforme descrito no diagrama da Figura 4.

Todos os processos de classificação se baseiam numa rede neural do tipo *Multilayer Perceptron* (MLP). Foi utilizada a API Weka, com algoritmos de aprendizado de máquina implementados em Java e distribuído gratuitamente sob a licença GNU [8].

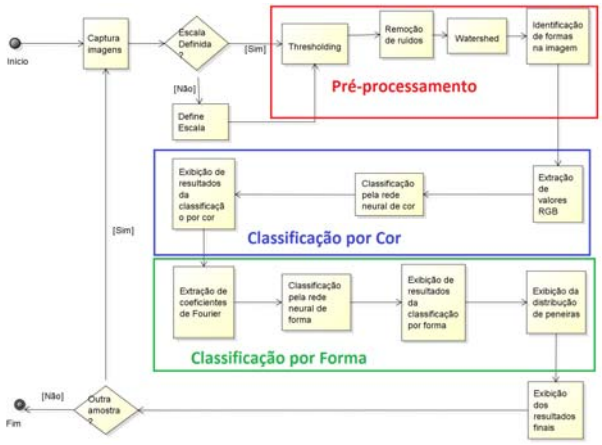


Figura 4: Diagrama de blocos do Qualicafé

As duas MLPs presentes neste trabalho utilizam o algoritmo de *backpropagation*, com taxa de aprendizado de 0,3, momento de 0,2, limiar de validação de 20 e 1000 épocas como condição de parada. Os números de neurônios das camadas de entrada e de saída são iguais, respectivamente, ao número de atributos e ao número de classes definidos e ambas MLPs apresentam apenas uma camada escondida, com n neurônios cada, sendo n = (número de atributos + número de classes) / 2.

2.4 Pré-processamento

A primeira etapa de processamento é a identificação de grãos e impurezas presentes na amostra. Esta identificação é feita através do contorno de cada objeto encontrado na imagem. Para a identificação dos contornos, primeiramente é feita a segmentação da imagem pela aplicação de um simples *Thresholding*, algoritmo que classifica cada pixel como fundo ou objeto, através da comparação do valor do pixel em escala de cinza com um valor adotado como limiar - o *threshold*. Com a separação do fundo dos objetos, a imagem é então binarizada. Utiliza-se os processos morfológicos de erosão e dilatação, a fim de se eliminar possíveis ruídos. Visando separar grãos que estejam encostados um no outro,

podendo gerar problemas na contagem de grãos, é executado o algoritmo de *watershed*. Para cada *pixel* 'objeto' da imagem é calculada a sua distância ao *pixel* 'fundo' mais próximo. Então são definidos os pontos com as distâncias máximas locais e cada um desses pontos é dilatado até que se chegue à borda do objeto ou à borda da região de outro ponto sendo dilatado. Dessa forma, cada uma das regiões provenientes de um ponto dilatado define um objeto diferente. A partir da imagem binarizada é finalmente executado um algoritmo para rotular as diferentes formas e se extrair os *pixels* de seus contornos.

Na Figura 5 pode ser visualizada uma imagem de saída do processo de identificação de grãos. Nesta imagem, os contornos dos grãos são pintados de verde e numerados sequencialmente. A rotulação é armazenada juntamente com os *pixels* localizados nos seu interior.

Para se trabalhar com as dimensões reais dos objetos durante a análise por forma, foi desenvolvida uma ferramenta para a indicação da escala da imagem, através da qual se desenha uma linha sobre a imagem e informa-se a distância real que ela representa.

Figura 5. Objetos identificados e rotulados.

2.5 Classificação por cor

A análise de cor dos grãos de café é feita para se distinguir grãos bons de grãos com certos tipos de defeito, como: grãos verdes, pretos, pretos-verdes, ardidos, etc. Essa análise se baseia no reconhecimento de padrões de cor presentes nos grãos, visto que certos defeitos são identificados por uma cor característica do grão, ou de uma parte dele, como por exemplo o tom mais esverdeado que é comum nos grãos-verdes e manchas marrons nos grãos melados. Da mesma forma, as sementes perfeitas também têm uma coloração característica que as definem.

Sendo assim, cada um dos defeitos passíveis de serem identificados pela análise de cor foi associado a uma classe, que juntamente com a classe de grão perfeito formam o conjunto de padrões a serem reconhecidos pela rede neural. As classes são: ardido, brocado, casca, côco, fava escura, melado, pau, pedra, película, perfeito, pergaminho, preto, preto-verde, secador e verde (todas ilustradas na Figura1). A MLP foi definida com três atributos de entrada: as componentes de cor R, G e B.

Para cada uma das classes, foram capturadas imagens de amostras contendo grãos da classe em questão para execução do treinamento da rede. Foram utilizadas janelas selecionadas manualmente sobre a imagem, cujas áreas apresentavam a coloração característica do defeito do grão (Figura 6), ou um padrão de saudável. Para cada um dos *pixels* que compõem essas áreas foi determinado o valor médio de R,G e B, com base nos *pixels* vizinhos. Esses valores foram então utilizados como entrada para a MLP no seu treinamento. Para cada classe foram coletados cerca de 8.000 *pixels*.

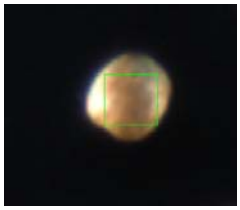


Figura 6. Seleção de amostra em um grão para treinamento da MLP.

A classificação dos grãos de uma amostra é feita se passando como entrada para a MLP treinada os valores médios de R, G e B de cada um dos *pixels* do grão com seus vizinhos. Conforme os *pixels* são classificados eles são pintados na imagem da amostra com uma cor diferente para cada classe, de modo que os padrões de cor que constituem o grão sejam facilmente identificados, como exibido na Figura 7. Quando todos os *pixels* de um grão tiverem sido classificados, a classe atribuída ao maior número de *pixels* é definida como a classe do grão.

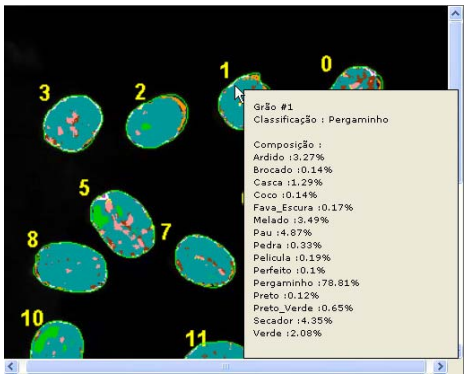


Figura 7. Resultado de uma classificação por cor.

2.6 Descritores de forma utilizados

Para que a MLP possa realizar o reconhecimento de padrões de forma, foram utilizados alguns descritores de forma. Os descritores utilizados foram:

Assinaturas: a distância de um ponto da borda até o seu centro de massa. Os valores foram normalizados pelo maior valor, de forma que cada um variasse entre 0 e 1.

Foram utilizados 36 valores, separados entre si de 10°. Juntamente com as assinaturas também foram utilizados o comprimento e a largura da forma, em milímetros.

Descritores de Fourier: Descritores de Fourier representam a borda de uma região como uma função periódica que pode ser expandia em uma série de Fourier. Uma característica típica dos descritores de Fourier é que geralmente a forma geral da figura é relativamente bem definida a partir de alguns dos termos de menor ordem da expansão. Para este projeto, optou-se por usar os 40 primeiros coeficientes para se definir as formas.

Na Figura 8 podem ser observados exemplos de assinatura e descritores de Fourier para cada tipo de contorno estudado.

A comparação dos modelos consistiu no treinando de duas MLPs para o reconhecimento de grãos perfeitos, grãos quebrados, paus e pedras, cada uma utilizando-se de um dos modelos. Então as MLPs foram utilizadas na classificação de amostras com grãos conhecidos, para que finalmente pudessem ser calculadas suas taxas de acerto.

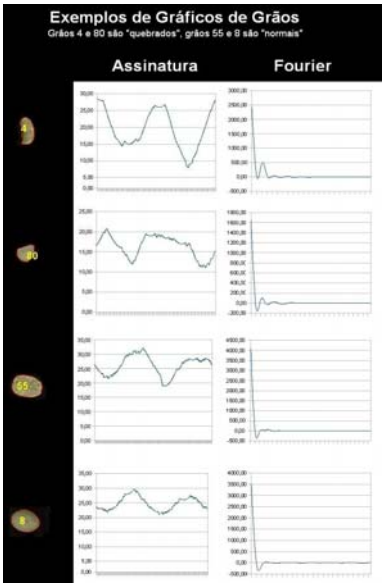


Figura 8: Assinatura e Descritor de Fourier

A MLP treinada com assinaturas, largura e comprimento teve uma taxa de acerto de cerca de 80%, enquanto que para a treinada com descritores de Fourier, a taxa foi de cerca de 60%. Portanto o primeiro modelo foi o adotado.

2.7 Classificação por forma

As impurezas de uma amostra de grãos de café podem ser facilmente reconhecidas através de suas formas, que estão fora do padrão que as sementes apresentam. Como exemplo, pode-se citar os paus, que apresentam um contorno fino e comprido, ou as pedras, que são pequenas e têm um contorno irregular. A análise de forma também é capaz de indicar grãos com defeitos, como no caso de grãos quebrados, que naturalmente são menores e não possuem o característico formato arredondado de uma semente perfeita. Seguindo essas premissas foi utilizada uma MLP para reconhecer os padrões de forma dos objetos presentes na imagem de uma amostra e indicar de qual classe de grão ou impureza esses padrões são característicos.

Essa MLP foi configurada com 38 atributos, dos quais 36 são assinaturas, uma é a largura e a outra o comprimento (ambos em milímetros) do grão a ser classificado, e 10 classes: cabeça, casca, côco, concha, marago, pau, pedra, perfeito, pergaminho e quebrado (apresentadas da Figura 1). Analogamente ao treinamento da MLP de cor, o da de forma foi feito capturando-se imagens contendo grãos ou impurezas de uma classe específica e então extraindo seus contornos para serem utilizados no cálculo das assinaturas, dos comprimentos e das larguras. Foi coletado um conjunto de cerca de 60 formas para cada uma das classes.

Além de indicar os padrões de forma presentes, também é gerado um histograma com a distribuição dos grãos por peneira, uma classificação utilizada pelas cooperativas de café. Em geral esta classificação é feita por um jogo de peneiras, que separa os grãos pela forma e pelo tamanho. As peneiras têm crivos com diversas medidas e dois formatos diferentes: podem ser oblongos, para separar os cafés mocas, ou circulares, para separar os cafés chatos. As medidas dos crivos das peneiras são dadas em frações de 1/64 de polegada e o número da peneira corresponde ao numerador da fração. Por exemplo: peneira 19 = 19/64 de polegada [9].

3. Resultados e Discussão

Com as MLPs devidamente treinadas foram utilizados conjuntos de grãos fornecidos pela Cooxupé em diversos testes com o intuito de se validar os métodos desenvolvidos.

3.1 Identificação dos objetos

Constatou-se que o processo de reconhecimento de formas na imagem apresenta algumas falhas, como ilustrado na Figura 9. A mais perceptível acontece na presença de objetos com coloração muito escura, visto que uma porção deles é rotulada como parte do fundo no processo de segmentação, resultando na extração de formas diferentes das reais. Outro problema ocorre na execução do algoritmo de *watershed,* que algumas vezes divide a forma de um único objeto em várias (o que ocorre com freqüência com os paus) e em certas ocasiões traça erroneamente a fronteira entre dois objetos que se encostam. Essas falhas se devem principalmente ao

sistema de iluminação da amostra, que não distribui a luz homogeneamente e ao fato de a superfície sobre a qual a amostra é depositada ter uma tonalidade de cor próxima a de certos objetos de interesse.

Apesar das falhas, de modo geral o algoritmo de reconhecimento de objetos se mostrou muito eficiente. Posicionando-se a amostra na faixa da bandeja onde a iluminação é mais homogênea e são gerados menos reflexos, são detectados praticamente todos os objetos (com exceção das formas muito escuras - defeito que pode ser corrigido alterando-se a cor do fundo usado).

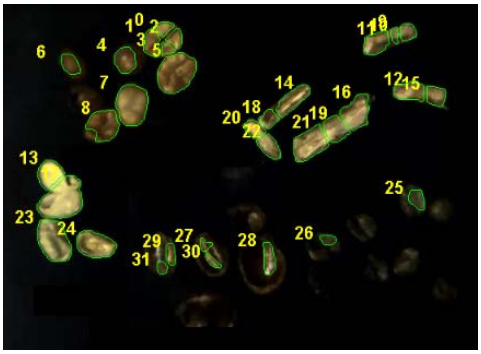


Figura 9. Deficiências na identificação de objetos.

3.2 Classificação por cor

Analisando-se visualmente as imagens capturadas das amostras, verificou-se que as cores apresentavam significante diferença em comparação àquelas observadas nos objetos reais, tornando mais difícil o reconhecimento de alguns dos padrões que definem algumas das classes das sementes. Somando-se a isso as limitações impostas pelo sistema de iluminação, que dificulta a extração satisfatória de contornos de objetos muito escuros, estão sendo alterados os processos de iluminação construídos no primeiro protótipo.

Para teste apenas, usando condição não-ideal de iluminação, foi classificada uma amostra contendo 15 grãos pergaminho, 15 perfeitos e 15 pedras e calculadas as taxas de acerto para cada uma das classes (Tabela 1).

Tabela 1. Taxa de acerto da classificação por cor.

Classe	Pergaminho	Pedra	Perfeito
Taxa de acerto	100%	80%	27%

Apesar da taxa de acerto para os grãos perfeitos ter sido baixa, do total de sementes perfeitas classificadas, 40% foram classificadas como "secador", grão visivelmente muito semelhante ao "perfeito", o que indica que houve certo êxito no reconhecimento do padrão, porém este não foi tão preciso quanto o necessário.

Sendo assim, o sistema de iluminação ainda está sendo modificado para se melhorar o processo de análise.

3.3 Classificação por forma

A extração de contornos de objetos escuros é prejudicada pela iluminação presente. Portanto, as classes de forma que apresentam uma coloração escura (cascas e côcos) não foram usados nos testes para validar o método de classificação por forma, além de serem menos presentes nas amostras. Foram utilizados 15 grãos quebrados, 15 pergaminhos e 15 maragos. As taxas de acerto resultantes estão na Tabela2.

Tabela 2. Taxa de acerto da classificação por forma.

Classe	Pergaminho	Quebrado	Marago
Taxa de acerto	46,67%	60,00%	40%

3.4 Distribuição de peneiras

Foram conduzidos testes para a obtenção da distribuição por peneira em três amostras de grãos, cada uma proveniente de um processo de separação por peneira com crivos diferentes. As peneiras em questão eram 15, 17 e 19. Nos três casos pôde-se perceber uma maior concentração em torno da furação utilizada na separação, a qual também apresentou em todos os casos o maior número de ocorrências. A figura 10 mostra o histograma obtido da amostra da peneira 19.

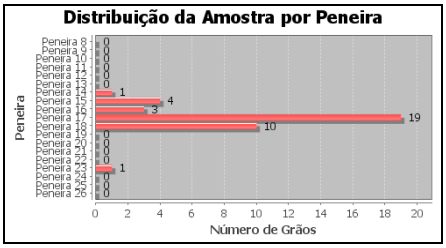


Figura 10. Distribuição por peneira de uma amostra de peneira 17.

3.5 Distinção entre grãos e impurezas

Comparando-se as classificações geradas pelo s*oftware* com as reais classes presentes nas amostras, observou-se que foi capaz de distinguir entre impurezas – paus e pedras – e grãos, sejam eles com defeitos ou não, com uma acurácia satisfatória. Dado que a presença de impurezas na amostra exerce uma depreciação no café muito maior que os grãos defeituosos, essa característica indica um importante resultado. Na figura 11 é mostrado um exemplo de classificação, na qual essa distinção é notada. Em verde estão os objetos corretamente classificados (como paus, pedras ou grãos) em pelo menos umas das duas análises, de cor e de forma, e em vermelho

aqueles cujas classificações não foram corretas em nenhuma delas.

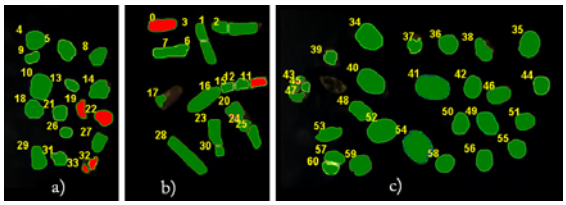


Figura 11. Acertos de classificação de grãos e impurezas a) pedras b) paus c) grãos.

A interface implementada no Qualicafé mostra a classificação do grão, tomado por cima e por baixo, como pode ser observado na Figura 12.

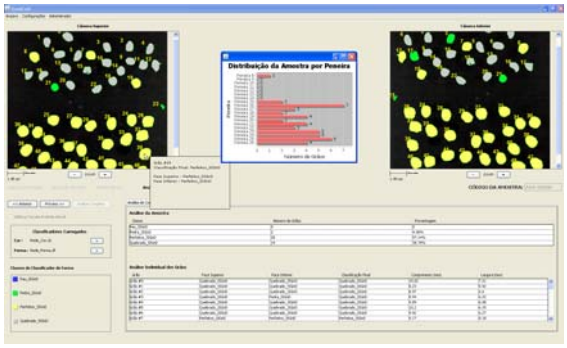


Figura 12: Tela de Resultados por análise

4. Conclusões e trabalhos futuros

Este trabalho buscou desenvolver um software que substituísse o atual processo manual de classificação de grãos de café por um método automático, mais eficiente e confiável. Os resultados obtidos mostraram que as atuais técnicas de processamento de imagem são capazes de extrair da imagem de um grão informações suficientes para definir sua forma e coloração, alguns dos atributos utilizados por especialistas humanos para classificar os grãos. Também se pode concluir que com as classificações geradas pelas redes neurais é possível fazer o reconhecimento de padrões de forma satisfatória.

Mesmo com as dificuldades geradas pela iluminação não-adequada, que acarreta grandes limitações em ambos os processo classificatórios, foi possível se obter bons resultados preliminares – distinguir grãos de impurezas e identificar certas classes com um acerto significativo – sugerindo que uma melhora no sistema de iluminação ampliará o substancialmente o potencial do método.

5. Agradecimentos

Agradecemos ao apoio financeiro da Cooxupé, Guaxupé, MG, aos alunos Gabriel Marcondes, Daniel Bueno e Alex Rodrigues pelas contribuições na fase inicial do projeto

6. Referências

[1] BEWLEY, J.D.; BLACK, M. Seeds: physiology of development and germination. 2º ed. New York: Plenun Press, p.199-267. 1994.

[2] KRZYZANOWSKI, C.F. Relato de testes de vigor disponíveis para as grandes culturas. Informativo ABRATES. V.1, n.2, p.59-61. 1994.

[3] MARCOS FILHO, J. Utilização de testes de vigor em programas de controle de qualidade de sementes. Informativo ABRATES. v.4, n.2, p.59-61. 1995.

[4] VIEIRA, D.R.; CARVALHO, N.M. Testes de vigor em sementes. Jaboticabal, FUNEP. 164 p. 1994.

[5] BRASIL. Ministério da Agricultura e Reforma Agrária. Regras para análise de sementes. Brasília, 1992. 365p.

[6] KHATCHATOURIAN, O.; PADILHA, F.R.R. Reconhecimento de variedades de soja por meio do processamento de imagens digitais usando redes neurais artificiais Eng. Agríc., Jaboticabal, v.28, n.4, p.759-769, out./dez. 2008.

[7] MACDOUGALL, D.B. Colour in food: Improving quality. Cambridge: Woodhead publishing, 2002. pp.115-142.

[8] HALL, M.; EIBE, F.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1. 2009.

[9] THOMAZIELLO, R.A. Classificação do café. Disponível em:<http://www.megaagro.com.br/cafe/art_class_cafe.asp >. Acesso em: 17 fev. 2011.

Análise óptica de sistemas eletro-ópticos por meio do cálculo da MTF

Elisa Signoreto Barbarini, Evandro Luís Linhari Rodrigues
Laboratório de Visão Computacional, Departamento de Engenharia Elétrica, Escola de Engenharia de São Carlos, Universidade de São Paulo.
Av. Trabalhador São Carlense, 400 – Cep:13566-590 – São Carlos, SP
elisa.barbarini@usp.br, evandro@sc.usp.br

Abstract

One of the widely used methods for performance analysis of an optic system is the determination of the Modulation Transfer Function (MTF). The MTF represents a quantitative and direct measure of image quality, and, besides being an objective test, can be used on concatenated optical system. This paper presents the application of software called SMTF (software modulation transfer function), built in C++ and Open CV platforms, for MTF calculation on electro-optical system. Through this method, the performance of a digital fundus camera and an infrared sensor was analyzed. The MTF information assists the analysis of the optical system alignment, and also defines its resolution limit by the MTF graphic. The result obtained from the implemented software is compared with the theoretical MTF curve from the analyzed systems.

Keywords: Modulation Transfer Function, Optical Alignment, Fourier Transform, Optical Performance.

1 Introdução

Equipamentos que integram óptica, eletrônica e mecânica são cada vez mais difundidos e utilizados em diferentes aplicações como na área médica, de sensoriamento e imageadores. Como exemplo destes equipamentos tem-se o retinógrafo digital, o microscópio cirúrgico oftalmológico e o sensor infravermelho.

Esses equipamentos devem possuir um bom desempenho óptico, permitindo que o sistema capture imagens com alta resolução e contraste.

A Função de Transferência Óptica (OTF) tornou-se a função mais utilizada para avaliar a qualidade da imagem no *design* de sistemas ópticos. A OTF fornece uma avaliação da estrutura da imagem indicando o contraste de um objeto senoidal para uma gama de frequências espaciais. Portanto, a informação veiculada pela OTF pode ser utilizada para construir de forma detalhada uma imagem [10].

A OTF é uma função complexa sendo possível calcular o módulo e a fase dessa função. O valor do módulo da OTF corresponde ao valor da MTF do sistema.

Como discutido por Scaduto [9] existem diferentes métodos para realizar o cálculo da MTF, dentre eles, os mais utilizados são os alvos com padrões senoidais, alvo do tipo fenda, ponto (*pinhole*), tipo borda (*knife-edge*), dentre outros. Outros testes ainda são baseados em métodos interferométricos [11].

Segundo Samei et al. [8] dentre as várias técnicas existentes para o cálculo da MTF, o alvo tipo borda é, na atualidade, o mais difundido.

Uma das justificativas para a utilização do alvo tipo borda é sua simplicidade de construção, sendo que não possui dimensões críticas, como o diâmetro reduzido do ponto necessário, no caso da função de espalhamento de ponto, e como a largura de um alvo do tipo fenda.

Além disso, quando comparado às outras técnicas apresenta um fluxo de energia muito superior, tornando a sua detecção mais conveniente [5].

Uma desvantagem de utilizar o degrau para cálculo da MTF é que durante o cálculo é realizada uma derivada espacial, sendo, portanto, acentuado o ruído existente na imagem [1].

O trabalho apresentado nesse artigo utiliza um alvo do tipo borda para realizar os cálculos da MTF dos sistemas eletro-ópticos analisados.

É utilizado primeiramente um sensor infravermelho para fazer a validação do método proposto e posteriormente o método é aplicado ao retinógrafo.

É utilizado um alvo padrão para todos os sistemas. O que muda entre um sistema e outro é a posição que o alvo é colocado devido à diferença da distância focal de cada equipamento.

2 Desempenho óptico

A imagem formada por um sistema óptico ideal deveria ser uma reprodução ponto a ponto exata de um objeto. Contudo, imperfeições estão sempre presentes num sistema óptico real e podem ser causados por efeitos de difração nas bordas das lentes, aberrações e erros na construção [13].

Uma das formas mais utilizadas para verificar o desempenho óptico de um sistema é utilizar o cálculo da MTF. Para realizar esse cálculo é necessário conhecimento sobre a Transformada de Fourier (FT) e a Função de Transferência Óptica.

2.1 Transformada de Fourier

A FT é uma ferramenta importante no processamento de imagens, a qual é utilizada para decompor uma imagem em componentes do tipo seno e cosseno. A FT é utilizada em uma vasta gama de aplicações, como análise de imagens, filtro para imagens, reconstrução de imagens e compressão de imagens.

Como apresentado por Scaduto [9], a saída da transformada de Fourier representa a imagem no domínio da frequência, enquanto que a imagem está no domínio do espaço. No domínio da frequência, cada ponto da imagem representa uma frequência particular.

Para cálculo da FT analiticamente pode ser utilizada a Equação (1).

$$\int f(x) e^{-i.2.\pi.x.s} dx \qquad (1)$$

Porém pode ser aproximada numericamente pela Equação (2).

$$\sum_{n=o}^{N-1} y_n e^{-ik\frac{2\pi}{N}n} \quad k \in [0, N-1] \qquad (2)$$

Nas Equações (1) e (2) f(x) e y_n representam a função a qual será aplicada a transformada, x representa a frequência que varia no intervalo de frequências onde a função é definida, e N representa o período onde a função está definida.

2.2 Função de transferência óptica

A OTF representa a capacidade de resolução de um sistema de imagem em função da frequência espacial dos objetos, demonstrando com exatidão para todas as frequências o grau de degradação que o sistema impõe no processo de produção da imagem [7], sendo a sua equação definida pela Equação (3), onde mtf(x,y) é a função de transferência de modulação do sistema, a ptf(x,y) é função de transferência de fase do sistema e x e y são as coordenadas da imagem.

$$otf(x,y) = mtf(x,y) * ptf(x,y) \quad (3)$$

Utilizando a FT e o teorema da convolução, tem-se que a função de transferência óptica, para um sinal senoidal, que pode ser calculada de acordo com a Equação (4), após ser aplicada a FT em todos os termos da Equação (3).

$$OTF(\varepsilon,\eta) = MTF(\varepsilon,\eta).PTF(\varepsilon,\eta) \qquad (4)$$

A partir da OTF pode-se definir a MTF e a Função de Transferência de Fase (PTF), conforme as Equações (5) e (6) respectivamente.

$$MTF(\varepsilon,\eta) = |OTF(\varepsilon,\eta)| \qquad (5)$$

$$PTF(\varepsilon,\eta) = e^{-i2.\pi.\lambda(\varepsilon,\eta)} \qquad (6)$$

Sendo $\varepsilon = 1/X$ e $\eta = 1/Y$, onde X e Y são os períodos espaciais nas direções x e y respectivamente.

Para as imagens utilizadas nesse artigo será analisada apenas a MTF dos sistemas, portanto $OTF(\varepsilon,\eta) = MTF(\varepsilon,\eta)$

2.3 Função de Transferência de Modulação

Embora o equacionamento da MTF tenha sido apresentado na Equação (5), segundo Holst [4] e Wolfe [12], a MTF pode ser dada pela relação adimensional expressa pela Equação (7).

$$MTF = \frac{Modulação\ do\ sinal\ da\ imagem}{Modulação\ do\ sinal\ do\ objeto} \qquad (7)$$

A Figura 1 apresenta a relação entre os sinais de entrada e saída de um sistema óptico, possibilitando realizar o cálculo da MTF.

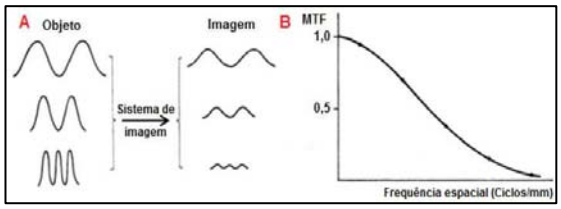


Figura 1 - (A) Relação entre a distribuição de intensidade de alvos com componentes senoidais de mesma amplitude e diferentes frequências espaciais e suas respectivas imagens; (B) A correspondente função de transferência de modulação (MTF) do sistema [7].

O gráfico proporcionado pela MTF pode oferecer uma análise sobre o limite de resolução do sistema. Esse limite é dado pelo valor da frequência espacial (ciclos/mm) onde a MTF é igual a zero [6]. A resolução de um sistema óptico pode ser definida como a separação, tanto em ângulo no espaço-objeto como em distância no plano-imagem, de dois pontos discretos que podem ser facilmente discernido [1].

2.4 Métodos para o cálculo da MTF

Existem diferentes métodos para o cálculo de MTF de um sistema óptico. O método que será abordado utiliza um padrão de alvo com frequência conhecida. Este tipo de alvo proporciona uma forma de medida que elimina a dependência da medida com a posição do alvo, permitindo a obtenção de uma MTF média que ocorre para uma cena natural. O teste consiste em capturar a imagem do alvo produzida pelo sistema em teste, usando um *charge-coupled device* (CCD), e analisar as frequências espaciais contidas na imagem detectada [3].

Esse método pode ser usado para testar apenas um arranjo de detectores como também sistemas ópticos completos [2].

3 Materiais e métodos

O trabalho aqui apresentado pode ser dividido em três etapas: a primeira apresenta a implementação do algoritmo necessário para os cálculos por meio do software desenvolvido; na segunda etapa o SMTF é testado utilizando uma imagem estática, com apenas um alvo para o sensor infravermelho e na terceira etapa a MTF é calculada utilizando cinco alvos, como o apresentado na Figura 2.

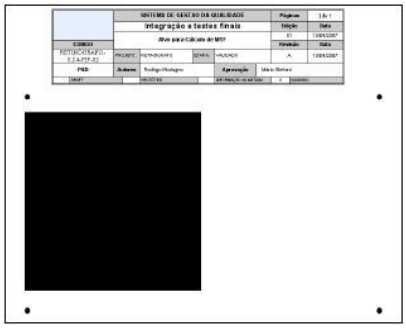


Figura 2 – Alvo do tipo borda para cálculo de MTF [5]

3.1 Algoritmo do software para cálculo da MTF

Abaixo é descrito o algoritmo utilizado para implementação do SMTF utilizando a linguagem de programação C++ e a biblioteca OpenCV.

1. Os vetores dos sinais de entrada, derivada dos sinais de entrada e módulo da FT, são criados;
2. A quantidade de entradas é representada pela quantidade alvos capturados pela imagem;
3. O tamanho do vetor dos sinais de entrada é definido pelo tamanho da linha do sinal sobre o alvo;
4. Os sinais de entrada são obtidos. Estes sinais são representados por linhas posicionadas sobre as áreas de transição de preto para branco nos alvos;
5. Verifica-se o valor mínimo do sinal de entrada de cada linha posicionada sobre os alvos;
6. A derivada do sinal é calculada, subtraindo o sinal deslocado de uma posição à frente do sinal original;
7. O máximo valor da derivada é calculado. Esse valor será utilizado para que apenas a transição do sinal seja considerada, zerando os pontos com baixa derivada;
8. Os valores com baixa derivada são zerados. Para isso é determinado um *threshold* zerando os pontos cuja derivada seja menor que 10% do máximo valor – esse valor foi determinado empiricamente;
9. Possuindo, dessa forma, a derivada do degrau de entrada. Calcula-se então o módulo da FT;
10. Calcula-se o valor máximo da FT;
11. É realizada a divisão do módulo da FT pelo valor calculado, normalizando, dessa forma, os valores da FT;
12. Por fim, é desenhado o gráfico do sinal de entrada e o módulo da FT, o qual caracteriza a MTF do sistema analisado.

3.2 Medidas utilizando o sistema óptico para o sensor infravermelho

Após a implementação do algoritmo apresentado acima, foi colocado um alvo em um anteparo, perpendicular ao eixo óptico do sensor infravermelho, na região central da imagem capturada.

Utilizando o sensor infravermelho foram adquiridas algumas imagens, como a apresentada na Figura 3, durante o alinhamento do sistema óptico do sensor.

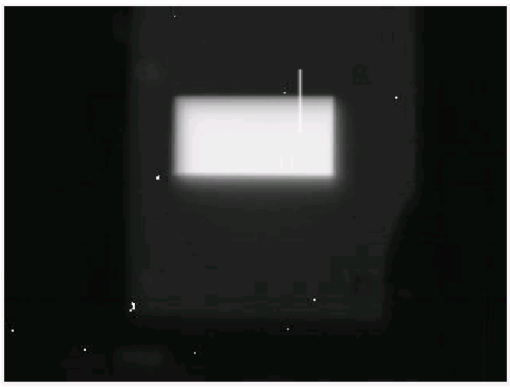


Figura 3 - Imagem captada pelo sensor infravermelho para cálculo da MTF

Após a imagem capturada foi necessário indicar o início e o final de uma linha, posicionada na transição da parte preta para a parte branca do alvo. Uma vez determinada a linha é possível fazer o cálculo da MTF e determinar o seu gráfico a partir do SMTF como pode ser visto na Figura 4.

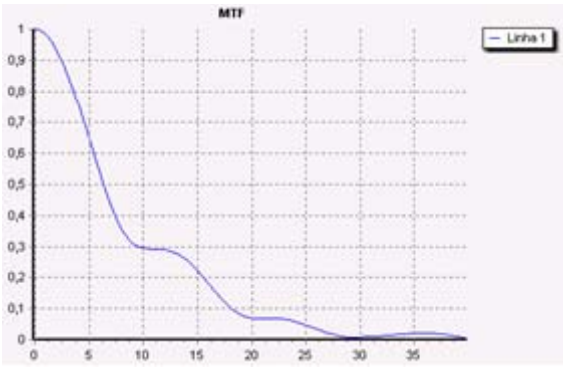


Figura 4 - Gráfico da MTF para um alvo utilizando o sensor infravermelho.

Após os cálculos, foi feito um comparativo entre a MTF calculada e a teórica.

A Figura 4 e a Figura 5 apresentam respectivamente a MTF calculada e a teórica do sensor infravermelho, após o alinhamento.

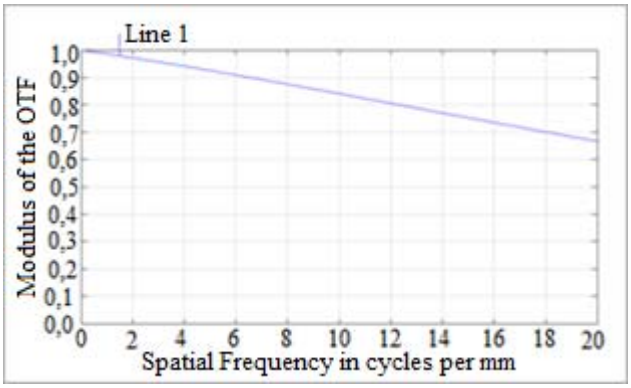


Figura 5 - Gráfico teórico da MTF para o sensor infravermelho.

Comparando os gráficos de MTF apresentadas pelo software com os teóricos verificou-se que a MTF do sensor ficou bem abaixo do esperado (teórico). Após essa verificação foram avaliadas a montagem e fabricação da óptica do sensor infravermelho e constatou-se que houve falha na fabricação e montagem da mesma.

3.3 Medidas utilizando o sistema óptico para o retinógrafo

Foram utilizados cinco alvos, sendo um central, perpendicular ao eixo óptico do retinógrafo, e os outros quatros alvos são colocados nas laterais de forma que fiquem equidistantes do alvo central.

Para o retinógrafo foi realizada uma modificação no software permitindo que ao invés da imagem ser capturada e após isso as linhas serem traçadas para o cálculo da MTF, as linhas são traçadas na imagem que está sendo adquirida e os alvos são colocados de forma que as linhas coincidam com a transição do preto para o branco.

Dessa forma, uma vez que os alvos são posicionados corretamente no anteparo perpendicular ao eixo óptico do retinógrafo, é possível visualizar continuamente a mudança na MTF quando se realiza qualquer mudança no alinhamento óptico do sistema. Com isso o processo de alinhamento é otimizado.

Por meio desse cálculo é possível observar que quando se tem um desalinhamento na horizontal o valor da MTF dos gráficos a direita são diferentes dos valores dos gráficos correspondentes ao lado esquerdo da imagem. Já quando se tem um desalinhamento vertical os valores da MTF dos gráficos localizados na parte de baixo estarão diferentes da parte de cima.

A imagem capturada pelo retinógrafo está apresentada na Figura 6.

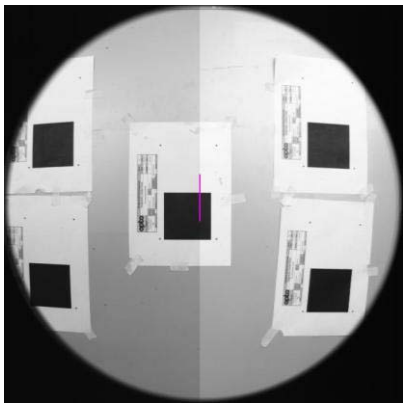


Figura 6 - Imagem capturada pelo retinógrafo para cálculo da MTF.

Na Figura 7 estão apresentados os gráficos resultantes do cálculo da MTF dos cinco alvos utilizados. Nos Gráficos de MTF do retinógrafo as linhas correspondem aos seguintes alvos: Linha 1 corresponde ao alvo superior esquerdo, a Linha 2 ao superior direito, a Linha 3 ao alvo central, a Linha 4 ao alvo inferior esquerdo e a Linha 5 ao alvo inferior direito.

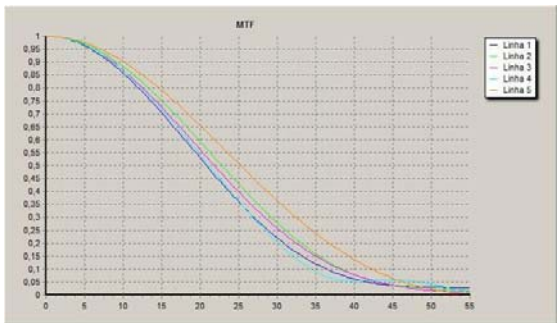


Figura 7 – Gráfico da MTF para cinco alvos utilizando o retinógrafo digital

A Figura 7 e a Figura 8 apresentam a MTF calculada e a teórica do retinógrafo, respectivamente, depois de realizado o alinhamento.

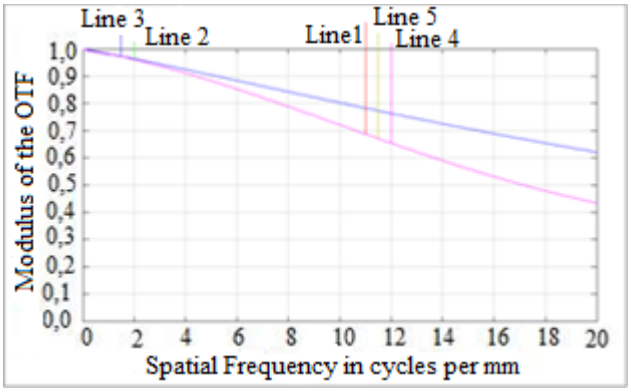


Figura 8 – Gráfico teórico da MTF do retinógrafo.

A partir do gráfico teórico do retinógrafo, é possível verificar que a MTF dos gráficos localizados nas laterais são iguais e a melhor MTF apresentada pelo sistema é do alvo central.

Considerando a MTF do retinógrafo é possível verificar que ocorreu uma degradação na MTF do sistema devido a um desalinhamento imposto pela mecânica do equipamento e por aberrações da óptica.

4 Conclusão

Durante a implementação do sistema proposto, para o cálculo de MTF, foi verificado que a iluminação do alvo influencia de forma direta o cálculo. Desta forma, para melhor aproveitamento do sistema os alvos devem ser iluminados de forma homogênea.

O software apresentou ótimo desempenho permitindo uma análise direta e objetiva do comportamento de sistemas ópticos. Com auxílio do software foi feito o alinhamento do retinógrafo digital e do sensor infravermelho.

Utilizando o SMTF para uma imagem estática conseguiu-se validar o funcionamento do mesmo, porém essa forma de análise demanda mais tempo do operador, pois a cada imagem capturada as linhas que determinam os sinais a serem analisados devem ser posicionadas de forma manual.

Com o software operando continuamente, uma vez que os alvos sejam posicionados de forma correta como descrito anteriormente, o cálculo da MTF é imediato. Qualquer modificação realizada no alinhamento do sistema óptico é analisada de forma rápida e direta, facilitando o alinhamento.

Esse trabalho permitiu observar, de forma prática, que nem sempre uma diferença entre os valores teóricos da MTF e os calculados pelo SMTF representa falta de alinhamento do sistema. As discrepâncias entre os valores da MTF puderam auxiliar no diagnóstico de outros problemas, como por exemplo, erro na fabricação das lentes presentes no sistema óptico, problemas na montagem do sistema, ou problemas com os filmes presentes em algumas lentes, entre outros.

Dando sequência na pesquisa, serão realizados ainda testes com um microscópio cirúrgico oftalmológico, após a implementação de um sistema óptico que acople um CCD ao mesmo, extraindo as curvas de MTF pelo software implementado e comparando com o teórico.

5 Referências Bibliográficas

[1] BOREMAN, G. D. *Modulation Transfer Function in Optical and Electro-Optical Systems.* Bllingham. SPIE Press, 2001. 110p..

[2] BOREMAN, G. D. *Modulation Transfer Function measurement using spatial noise targets.* Proc. SPIE, 1995, v. 2426, p. 386-390.

[3] DANIELS, A.; BOREMAN, G. D. DUCHARME, A. D. *Random targets for Modulation Transfer Function testing.* Proc. SPIE, 1993. v. 1969, p. 184-192.

[4] HOLST, G. C. *Infrared imaging system testing.* In: ACCETA, J. S.; SCHUMAKER, D. L. *The infrared and electro-optical systems handbook.* Ann Arbor: ERIM; Bellingham: SPIE, 1996. v. 5, cap. 4, p. 195-243.

[5] MODUGNO, R. G. *Uma contribuição ao projeto de retinógrafos Digitais*. São Carlo, 2009. 128 p..

[6] OLIVEIRA, P. D. JR. *Esquema Computacional para Avaliação Automática de Parâmetros de Qualidade de Equipamentos Mamográficos* São Carlos, 2009. 226 p..

[7] ROSSMANN, K. *Point spread function, line spread function and modulation transfer function.* Radiology, 1969. v. 12. p. 65-77.

[8] SAMEI, E.; BUHR, E.; GRANFORS, P.; VANDENBROUCKE, D.; WANG, X. *Comparison of edge analysis techniques for the determination of the MTF of digital radiographic system.* Physics in Medicine and Biology, v. 50. P.3613-3625, jul, 2005.

[9] SCADUTO, L. C. N. *Desenvolvimento e avaliação do desempenho de sistema óptico aplicado a sensoriamento remoto orbital*. São Carlos, 2008. 240p..

[10] SHANNON, R.R. *The Art and Science of Optical Design*. Cambridge University Press, 1997. 614p..

[11] WILLIAMS, T. L. *The OpticalTransfer Function of imaging systems*. Bristol: Institute of Physics Publishing, 199. 504p..

[12] WOLFE, W. L. *Introduction to infrared system design.* Bellingham: SPIE, 1996.

[13] YASOUKA, F. M. M. *Técnicas de desenho óptico e avaliação da qualidade de imagens em sistemas ópticos por medição de função de transferência óptica.* São Carlos, 1997. 260p..

Comparação de métodos de classificação da cobertura do solo urbano utilizando imagens IKONOS II com e sem o auxílio de dados LiDAR

Leonardo Rodrigues de Deus [1]
Rafael Santos [1]
Leila Maria G. Fonseca [1]
[1] Instituto Nacional de Pesquisas Espaciais
ldedeus@dpi.inpe.br; rafael.santos@lac.inpe.br; leila@dpi.inpe.br

Abstract

This study aims to use remote sensing data, comprising images of the IKONOS sensor II and LiDAR data, to classify urban land cover, and to evaluate whether LiDAR data can provide more precise classification results. We used three different classification methods implemented in the Weka software: IBK, J48, and MLP to process the data. The results showed that, for all classifiers, the use of IKONOS II images in association with LiDAR data enabled the urban land cover be classified more accurately, improving the accuracy by 5% for the study area.

1. Introdução

Os dados de sensoriamento remoto de alta resolução são uma fonte importante de informação a respeito das áreas urbanas, visto que as imagens produzidas possuem alta resolução espacial, o que permite uma melhor discriminação das feições presentes na cena.

No processo de extração de informações, as imagens de sensoriamento remoto, normalmente, são classificadas utilizando métodos automáticos para distinção entre as classes de cobertura do solo que são de interesse em cada aplicação.

Contudo, as imagens de alta resolução espacial possuem baixa resolução espectral, compreendendo apenas a faixa de comprimento de onda do visível e infravermelho próximo o que, segundo Pinho et al. (2007), dificulta a distinção de uma série de alvos urbanos que apresentam comportamento espectral semelhante para estes comprimentos de onda, como é o caso de ruas pavimentadas com asfalto e edificações com cobertura de amianto escuro.

Dessa forma, é necessário que novos métodos sejam desenvolvidos para auxiliar nos processos de classificação automática das imagens.

Nesse contexto, o uso combinado de imagens multiespectrais IKONOS II com dados LiDAR, pode melhorar a precisão da classificação da cobertura do solo, uma vez que os dados LiDAR fornecem informação de altura dos objetos. Esta informação pode favorecer a discriminação entre alvos que apresentam respostas espectrais semelhantes, mas possuem diferença de altura, como telhado e solo exposto.

LiDAR (Light Detection and Ranging) é um sistema topográfico que se baseia na emissão de milhares de pulsos laser de luz infravermelha por segundo, e no registro do retorno do sinal. O sistema realiza tarefas como: medir as distâncias até os alvos, a intensidade da energia refletida pelo objeto e os parâmetros de atitude do feixe (azimute e elevação), podendo assim determinar com precisão as elevações da superfície (JENSEN, 2009).

O objetivo deste estudo consiste em utilizar imagens do sensor IKONOS II e dados LiDAR para classificar a cobertura do solo urbano e avaliar se a informação de altura fornecida pelo dado LiDAR possibilita uma melhora dos resultados da classificação. Além disso, três métodos de classificação baseados em técnicas de mineração de dados são analisados.

2. Metodologia

A área de estudo deste trabalho consiste de uma porção central da cidade de Uberlândia-MG, a qual coincide com a região da cidade que apresenta a maior aglomeração de edificações verticalizadas (TÓMAS, 2010), onde foram selecionadas diversas amostras para testar os classificadores.

Foram utilizados dois tipos de dados para a realização do trabalho: imagens multiespectrais do sensor IKONOS II e dados LiDAR.

A imagem IKONOS II é uma imagem óptica, fusionada, com 1 metro de resolução espacial e composta por 4 bandas espectrais, correspondentes às faixas do azul (0,45 – 0,52 μ, Banda 1), verde (0,52 – 0,60 μ, Banda 2), vermelho (0,63 – 0,69 μ, Banda 3) e infravermelho próximo (0,76 – 0,90 μ, Banda 4). A Figura 1 mostra a imagem IKONOS II, composição colorida, da área de estudo.

Em relação aos dados LiDAR, foram utilizados dois arquivos raster, com tamanho de pixel de 1 metro. O primeiro, mostrado na Figura 2, corresponde ao Modelo

Digital de Altura (MDA) da área de estudo, o qual é resultado da subtração entre o Modelo Digital de Superfície (MDS), primeiro retorno do pulso laser, e o Modelo Digital de Terreno (MDT), último retorno do pulso laser, obtidos a partir da nuvem de pontos LiDAR. Nesse arquivo raster cada pixel da imagem apresenta um valor de altitude referente aos alvos que foram atingidos pelos pulsos emitidos pelo equipamento. O segundo arquivo, Figura 3, é uma imagem Intensidade, que corresponde à intensidade da energia refletida pelos objetos atingidos pela luz infravermelha emitida pelo equipamento.

Figura 1. Imagem IKONOS II da área de estudo.

Figura 2. Imagem MDA da área de estudo.

Dois experimentos, com diferentes grupos de dados foram realizados: (1) utilizando quatro bandas do sensor IKONOS II; (2) utilizando quatro bandas do sensor IKONOS II em conjunto com as imagens MDA e Intensidade, geradas a partir da nuvem de pontos do sistema LiDAR.

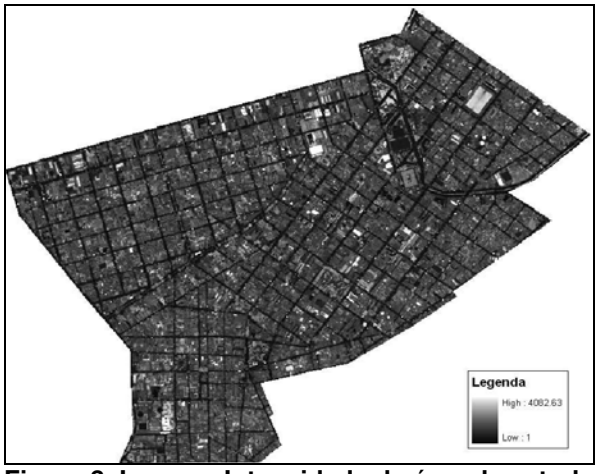


Figura 3. Imagem Intensidade da área de estudo.

Para realizar a classificação da cobertura do solo urbano foi utilizado o software Weka (Waikato Environment for Knowledge Analysis), o qual é uma coleção de algoritmos de aprendizagem de máquina para realizar tarefas de mineração de dados, que contem várias ferramentas para pré-processamento dos dados e visualização, além de algoritmos para regressão, classificação, agrupamentos, mineração de regras de associação e seleção de atributos (HALL et al., 2009).

Os classificadores utilizados neste estudo foram: IBK, que é um algoritmo de aprendizagem baseado em instâncias (Instance-Based Learning – IBL); J48, que é um classificador na forma de uma árvore de decisão, baseado no algoritmo C4.5; e o classificador *Multi-Layer Perceptron* (MLP), que é um tipo de rede neural artificial que usa um sistema de aprendizado supervisionado por correção de erros (retropropragação) para predizer a classificação das instâncias analisadas.

Para o classificador IBK, foram utilizados valores de K iguais a 1,3,5,8,11,15. O parâmetro K corresponde ao número de vizinhos mais próximos utilizados no processo. Em relação ao classificador MLP, o mesmo foi executado utilizando 1, 5, 10, 25 e 50 neurônios em uma camada escondida para avaliar sua eficiência na classificação.

Devido ao grande número de alvos urbanos, foram selecionadas amostras de 12 classes de cobertura do solo: asfalto, edifício, piscina, solo exposto, sombra, telhado azul, telhado branco, telhado escuro, telhado marrom, telhado marrom escuro, vegetação e vegetação rasteira. A Figura 4 mostra uma das amostras selecionadas para cada uma das 12 classes.

Com base no conjunto de amostras das classes de cobertura do solo uma tabela de dados foi gerada, onde cada pixel das amostras refere-se a uma instância e o valor associado a cada pixel, em cada uma das imagens, IKONOS II e LiDAR, refere-se a um atributo da tabela, bem como a classe de cobertura do solo determinada para as amostras selecionadas. A Figura 5 mostra parte da

tabela gerada, contendo as instâncias e os atributos utilizados para predizer a classificação da cobertura do solo.

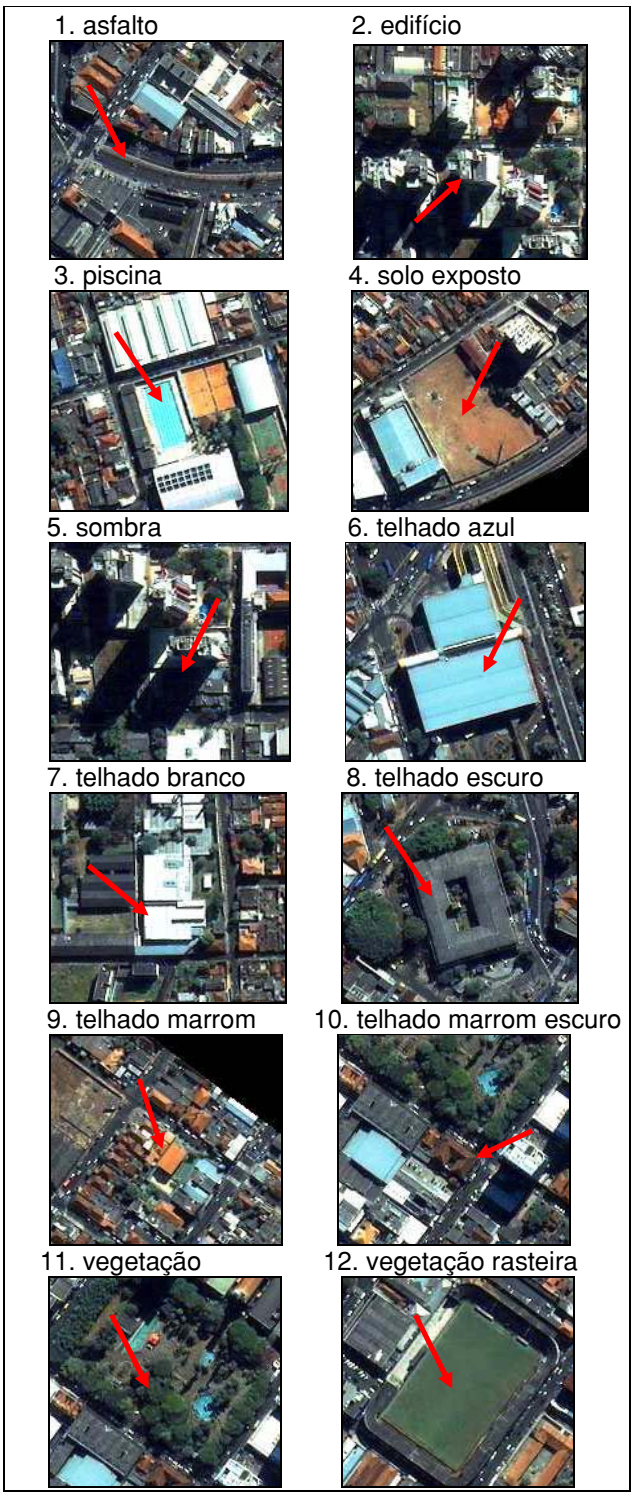


Figura 4. Amostras dos alvos considerados

No.	ikonos_b1 Numeric	ikonos_b2 Numeric	ikonos_b3 Numeric	ikonos_b4 Numeric	mda Numeric	intensidade Numeric	classe Nominal
2266	581.0	629.0	741.0	740.0	4.34082	37.596031	TELHADO_MARROM_ESCURO
2274	301.0	316.0	362.0	378.0	5.312073	37.596031	TELHADO_MARROM_ESCURO
462	670.0	750.0	750.0	634.0	0.050049	37.596367	SOLO_EXPOSTO
522	565.0	630.0	627.0	533.0	0.046997	37.596367	SOLO_EXPOSTO
582	640.0	713.0	708.0	585.0	0.051697	37.596367	SOLO_EXPOSTO
29371	557.0	564.0	434.0	343.0	40.494873	37.599899	EDIFICIO
29376	568.0	548.0	421.0	355.0	42.724548	37.599899	EDIFICIO
25778	524.0	533.0	571.0	608.0	3.283081	37.607563	TELHADO_MARROM_ESCURO
25798	395.0	398.0	416.0	468.0	3.447083	37.607563	TELHADO_MARROM_ESCURO
2545	359.0	232.0	153.0	145.0	1.717651	37.60828	SOMBRA
2563	367.0	234.0	155.0	151.0	2.863647	37.60828	SOMBRA
2581	364.0	230.0	154.0	155.0	4.410095	37.60828	SOMBRA
1115	647.0	722.0	645.0	586.0	-0.016235	37.616226	SOLO_EXPOSTO
31891	474.0	376.0	293.0	327.0	4.877686	37.61623	TELHADO_ESCURO
13651	1512.0	1597.0	1491.0	1332.0	3.368652	37.616688	TELHADO_BRANCO
13712	1490.0	1569.0	1458.0	1259.0	3.440918	37.616688	TELHADO_BRANCO
20700	681.0	832.0	835.0	713.0	-0.045959	37.618652	SOLO_EXPOSTO
20783	699.0	865.0	867.0	749.0	-0.010559	37.618652	SOLO_EXPOSTO
29563	712.0	789.0	633.0	461.0	70.72052	37.619015	EDIFICIO
29571	677.0	717.0	602.0	462.0	70.886719	37.619015	EDIFICIO
29579	689.0	705.0	607.0	474.0	70.995605	37.619015	EDIFICIO
31071	595.0	528.0	399.0	397.0	4.628723	37.62077	TELHADO_ESCURO
26410	331.0	231.0	158.0	588.0	2.87793	37.623207	VEGETACAO
26411	512.0	406.0	278.0	866.0	-0.110107	37.623207	VEGETACAO

Figura 5. Amostras da tabela de atributos.

3. Resultados e Discussões

3.1 Resultados para o Classificador IBK

A Tabela 1 sintetiza os resultados da precisão com que os dados amostrais foram classificados utilizando o classificador IBK com os diferentes valores de K.

Tabela 1: Resultados para o Classificador IBK

K	Precisão das Instâncias Classificadas Corretamente (%)	
	IKONOS II	IKONOS II + LiDAR
1	100	100
3	97,0635	99,6853
5	96,6326	99,5478
8	96,0918	99,3644
11	95,7801	99,2605
15	95,5448	99,1658

Analisando a Tabela 1 observa-se que, para os dois experimentos, a aplicação do classificador IBK com K igual a 1 foi perfeita, apresentando 100% de precisão para as instâncias classificadas.

A Tabela 1 mostra ainda que, à medida que o valor de K aumenta, a precisão das instâncias classificadas corretamente diminui, e que o uso do dado LiDAR melhora os resultados da classificação.

Esse resultado indica que o conjunto de amostras apresenta uma grande mistura de informações, visto que o classificador com K igual a 1 apresentou os melhores resultados, e independe de quanto se aumente este valor, os resultados são sempre piores, indicando que os dados amostrais não se caracterizam pela presença de *outliers*,

que podem ser ruído no espaço amostral do conjunto de dados.

As matrizes de confusão geradas na predição da classificação com diferentes valores de K, para o segundo experimento, o qual apresentou os melhores resultados, mostraram, como pode ser visto na Figura 6, que as classes piscina, sombra e vegetação rasteira tiveram 100% de acerto independente do número de vizinhos mais próximo utilizados, mostrando que os atributos utilizados foram excelentes para distinguir estes 3 tipos de alvos com a aplicação deste classificador.

Para leitura das figuras que mostram matrizes de confusão deve-se considerar:

. a = asfalto
. b = piscina
. c = solo exposto
. d = sombra
. e = edifício
. f = telhado azul
. g = telhado branco
. h = telhado escuro
. i = telhado marrom
. j = telhado marrom escuro
. k = vegetação
. l = vegetação rasteira

Valor de K	a	b	c	d	e	f	g	h	i	j	k	l	<-- Classificado como
1	0	652	0	0	0	0	0	0	0	0	0	0	b = piscina
3	0	652	0	0	0	0	0	0	0	0	0	0	
5	0	652	0	0	0	0	0	0	0	0	0	0	
8	0	652	0	0	0	0	0	0	0	0	0	0	
11	0	652	0	0	0	0	0	0	0	0	0	0	
15	0	652	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	1830	0	0	0	0	0	0	0	0	d = sombra
3	0	0	0	1830	0	0	0	0	0	0	0	0	
5	0	0	0	1830	0	0	0	0	0	0	0	0	
8	0	0	0	1830	0	0	0	0	0	0	0	0	
11	0	0	0	1830	0	0	0	0	0	0	0	0	
15	0	0	0	1830	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	5751	l = vegetacao rasteira
3	0	0	0	0	0	0	0	0	0	0	0	5751	
5	0	0	0	0	0	0	0	0	0	0	0	5751	
8	0	0	0	0	0	0	0	0	0	0	0	5751	
11	0	0	0	0	0	0	0	0	0	0	0	5751	
15	0	0	0	0	0	0	0	0	0	0	0	5751	

Figura 6. Matriz de confusão das classes piscina, sombra e vegetaçao rasteira - Classificador IBK

Outra classe que também foi muito bem discriminada foi a classe Telhado Azul, que só não teve 100% de acerto para todos os valores de K utilizados porque para K igual a 3, uma única instância foi classificada como Telhado Branco.

Para as demais classes de cobertura do solo, os resultados mostraram que à medida que o valor de K aumentou as confusões entre as classes também aumentaram. A maior parte das confusões observadas refere-se a alvos que possuem características espectrais semelhantes, como asfalto e telhado escuro ou solo exposto e telhado marrom escuro, por exemplo, como mostra a Figura 7.

Valor de K	a	b	c	d	e	f	g	h	i	j	k	l	<-- Classificado como
1	291	0	0	0	0	0	0	0	0	0	0	0	a = asfalto
3	284	0	0	0	0	0	0	7	0	0	0	0	
5	277	0	0	0	0	0	0	14	0	0	0	0	
8	276	0	0	0	0	0	0	15	0	0	0	0	
11	271	0	0	0	0	0	0	20	0	0	0	0	
15	265	0	0	0	0	0	0	26	0	0	0	0	
1	0	0	2566	0	0	0	0	0	0	0	0	0	c = solo exposto
3	0	0	2561	0	0	0	0	0	1	4	0	0	
5	0	0	2560	0	0	0	0	0	0	6	0	0	
8	0	0	2559	0	0	0	0	0	0	7	0	0	
11	0	0	2558	0	0	0	0	0	0	8	0	0	
15	0	0	2557	0	0	0	0	0	0	9	0	0	

Figura 7. Matriz de confusão das classes asfalto e solo exposto - Classificador IBK

Mas algumas confusões não eram esperadas, como é caso da confusão da classe asfalto e telhado escuro, mostrada na Figura 7, e vegetação e vegetação rasteira, Figura 8, visto que com a utilização Modelo Digital de Altura (MDA) os alvos com alturas diferentes deveriam ser bem discriminados. Isto indica que o MDA utilizado deve ser revisto no intuito de avaliar se não existem áreas com erros de altitudes.

Valor de K	a	b	c	d	e	f	g	h	i	j	k	l	<-- Classificado como
1	0	0	0	0	0	0	0	0	0	0	3000	0	k = vegetacao
3	0	0	0	0	0	0	0	1	0	0	2997	2	
5	0	0	0	1	0	0	0	1	0	0	2993	5	
8	0	0	0	4	0	0	0	1	0	0	2987	8	
11	0	0	0	4	0	0	0	0	0	0	2986	10	
15	0	0	0	7	0	0	0	0	0	0	2981	12	

Figura 8. Matriz de confusão da classe vegetação - Classificador IBK

Contudo, de maneira geral, os resultados obtidos com a utilização do classificador IBK mostraram que a utilização de imagem óptica em conjunto com dados LiDAR, embora tenha apresentado algumas confusões, inclusive entre alvos de alturas diferentes, favorece para que se obtenha uma classificação da cobertura do solo urbano com alto grau de precisão, visto que a classificação das amostras de teste sempre alcançou resultados acima de 99% de precisão, como mostrado na Tabela 1.

3.2 Resultados para o Classificador J48

A Tabela 2 apresenta os resultados gerais obtidos com o classificador J48.

Os resultados apresentados na Tabela 2 indicam que o uso das imagens IKONOS II juntamente com os dados LiDAR é a melhor opção para classificar a cobertura do

solo urbano, obtendo um ganho de aproximadamente 3% na precisão da classificação.

Tabela 2: Resultados para o Classificador J48

Atributos	Precisão das Instâncias Classificadas Corretamente (%)	Número de Folhas	Tamanho da Árvore
IKONOS II	96,984	453	905
IKONOS II + LiDAR	99,6639	230	459

Com mais atributos para classificar a cobertura do solo, o número de folhas da árvore de decisão diminui e conseqüentemente o tamanho da árvore também, indicando uma menor necessidade de recursos computacionais no processo.

A análise das matrizes de confusão, utilizando o classificar J48, mostrou também que quando se utiliza apenas as bandas do satélite IKONOS II para predição da classificação da cobertura do solo urbano, o grau de confusão entre as classes analisadas é maior, diminuindo com a utilização dos dados LiDAR no processo.

Usando dados LiDAR, o número de instâncias classificadas incorretamente é praticamente irrelevante frente ao número total de instâncias pertencentes a cada classe, como pode ser visto na Figura 9. Do total de 32.726 amostras, apenas 110 foram classificadas incorretamente, o que representa menos de 0,4% dos dados classificados.

a	b	c	d	e	f	g	h	i	j	k	l	<-- Classificado como
285	0	0	0	0	0	0	6	0	0	0	0	a = asfalto
0	652	0	0	0	0	0	0	0	0	0	0	b = piscina
0	0	2558	0	0	0	0	0	2	6	0	0	c = solo exposto
0	0	0	1829	0	0	0	0	0	0	1	0	d = sombra
0	0	0	0	1282	0	14	11	1	0	0	0	e = edifício
0	0	0	0	0	5637	1	0	0	0	0	0	f = telhado azul
0	0	0	0	2	3	4290	3	0	3	0	0	g = tellhado branco
0	0	3	0	1	4	8	4864	1	1	1	0	h = telhado escuro
0	0	4	0	0	0	1	0	875	2	0	0	i = telhado marrom
0	0	13	0	0	0	0	10	4	1597	0	0	j = telhado marrom escuro
2	0	0	0	1	0	0	0	0	1	2996	0	k = vegetação
0	0	0	0	0	0	0	0	0	0	0	5751	l = vegetação rasteira

Figura 9. Matriz de confusão segundo grupo de atributos - Classificador J48

A Figura 9 mostra que, dentre as confusões que mais se destacam, tem-se a classe Edifício que foi classificada como Telhado Branco e Telhado Escuro, o que provavelmente deve-se ao fato de as amostras da classe Edifício conterem prédios que tem telhados compostos por materiais diferentes, e a classe Telhado Marrom Escuro que teve algumas instâncias classificadas como Telhado Escuro e como Solo Exposto.

Mais uma vez observa-se que alvos com diferença de altitude, embora para um número bastante reduzido de instâncias, apresentaram confusões, ressaltando a necessidade de revisão do arquivo MDA utilizado.

A análise da árvore de decisão obtida com a utilização conjunta das imagens ópticas e dados LiDAR, apresentou no topo da árvore as bandas 1 e 4 do satélite IKONOS II, como mostra a Figura 10.

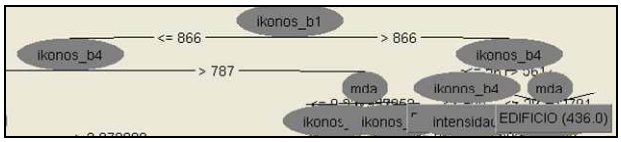


Figura 10. Topo da árvore de decisão gerada pelo Classificador J48 para o segundo grupo de atributos.

Conforme exposto por Sousa e Kux (2005), os alvos urbanos apresentam respostas espectrais diferentes para as faixas do azul e do infravermelho do espectro eletromagnético, justificando então o fato destas duas bandas estarem no topo da árvore de decisão.

Observou-se ainda, pela árvore gerada que o atributo MDA também aparece nas camadas mais altas da árvore distinguindo algumas instâncias da classe Edifício dos demais alvos.

Comparando os resultados obtidos pelo classificador J48 frente aos resultados obtidos pelo classificador IBK, observa-se que a diferença de precisão entre as instâncias classificadas corretamente considerando apenas o uso do segundo grupo de atributos pode ser considerada irrelevante, uma vez todos os resultados estiveram acima de 99% de precisão. Contudo, o classificador J48 leva uma grande vantagem na questão do tempo de processamento, apresentando o resultado em pouco mais de 5 segundos, enquanto o IBK levou aproximadamente 5 minutos para predizer a classificação para o mesmo grupo de atributos.

3.3 Resultados Classificador MLP

A Tabela 3 mostra os resultados das instâncias classificadas corretamente a partir da utilização do classificador MLP.

Analisando a Tabela 3, observa-se que à medida que o número de camadas escondidas aumenta a porcentagem de instâncias classificadas corretamente também aumenta. E da mesma forma que para os outros dois classificadores testados, o uso de dados IKONOS e LiDAR em conjunto melhora a precisão da classificação.

Tabela 3: Resultados Classificador MLP

Número de Camadas Escondidas	Precisão das Instâncias Classificadas Corretamente (%)	
	IKONOS II	IKONOS II + LiDAR
1	56,5576	59,2281
5	91,4838	97,6135
10	92,8008	99,0008
25	95,8504	99,3919
50	95,7679	99,4653

Em relação ao segundo experimento (IKONOS + LiDAR) observou-se que, ao se utilizar apenas uma camada escondida, o classificador não consegue distinguir as classes consideradas, classificando todas as instâncias em apenas 5 classes diferentes, Telha Azul, Telhado Branco, Telhado Escuro, Vegetação e Vegetação Rasteira, como mostra a Figura 11, obtendo 100% de erro para as instâncias pertencentes às demais classes.

À medida que o número de camadas escondidas aumenta a confusão entres as classes diminui gradativamente.

Usando 50 camadas escondidas, a confusão entre as classes é mínima, e está relacionada com os alvos que possuem respostas espectrais semelhantes.

a	b	c	d	e	f	g	h	i	j	k	l	<-- Classificado como
285	0	0	0	1	0	0	5	0	0	0	0	a = asfalto
0	652	0	0	0	0	0	0	0	0	0	0	b = piscina
0	0	2556	0	0	0	0	0	1	9	0	0	c = solo exposto
0	0	0	1830	0	0	0	0	0	0	0	0	d = sombra
0	0	0	0	1246	0	51	11	0	0	0	0	e = edifício
0	0	0	0	0	5637	1	0	0	0	0	0	f = telhado azul
0	0	0	0	2	3	4273	15	2	6	0	0	g = tellhado branco
4	0	0	1	1	0	3	4870	1	3	0	0	h = telhado escuro
0	0	5	0	0	0	0	0	871	6	0	0	i = telhado marrom
0	0	16	0	0	0	0	22	2	1584	0	0	j = telhado marrom escuro
0	0	0	0	0	0	0	1	0	0	2997	2	k = vegetação
0	0	0	0	0	0	0	0	0	0	1	5750	l = vegetação rasteira

Figura 11. Matriz de confusão Classificado MLP com 50 camadas escondidas.

Para analisar se a precisão do classificador MLP atingiria 100%%, o classificador foi testado com números maiores de camadas escondidas, 100, 200 e 500.

Os resultados obtidos mostraram que de 10 até 200 camadas escondidas a precisão de instâncias classificadas corretamente é praticamente a mesma, com 99% de precisão, variando o valor apenas nas casas decimais. Com mais de 200 camadas escondidas o classificador perdeu a precisão, estando com 500 camadas escondidas com 97% de precisão.

O classificador MLP, em relação os dois classificadores empregados, apresentou uma desvantagem significativa relacionada ao custo computacional para realizar a predição da classificação. O tempo de processamento ao utilizar este classificador é bem mais alto do que ao se utilizar o classificar IBK ou o J48.

4. Considerações Finais

Os resultados obtidos usando os três classificadores mostraram que as bandas do sensor IKONOS II são capazes de distinguir os alvos que compõem a cobertura do solo urbano. Porém, a utilização dos dados LiDAR propiciou uma melhora na precisão das instâncias classificadas corretamente da ordem de 5%.

Então, havendo a disponibilidade de dados LiDAR para realização da classificação da cobertura do solo urbano, estes devem ser empregados em conjunto com os dados de imagens ópticas para melhorar os resultados da classificação de áreas urbanas.

Em relação aos classificadores, o classificador J48 apresentou a melhor relação custo benefício, pois o tempo de processamento computacional foi bem inferior aos demais, e a precisão foi tão alta quanto qualquer outro. No entanto vale a pena ressaltar que o conjunto de regras gerado foi muito grande para ser interpretado.

5. Referências Bibliográficas

Pinho, C. M. D.; Almeida, C. M.; Kux, H. J. H.; Rennó, C. D.; Fonseca, L. M. G. Classificação de cobertura do solo de ambientes intra-urbanos utilizando imagens de alta resolução espacial e classificação orientada a objetos. In: Almeida, C. M.; Câmara, G.; Monteiro, A. M. V. (Ed.). **Geoinformação em urbanismo: cidade real x cidade virtual**. São Paulo: Oficina de Textos, 2007. cap. 8, p. 171-192.

Jensen, J. R. **Sensoriamento Remoto do Ambiente: uma perspectiva em recursos terrestres**. 2ª ed. São José dos Campos: Parêntese, 2009, 598 p.

Tomás, Lívia Rodrigues. **Inferência populacional urbana baseada no volume de edificações residenciais usando imagens IKONOS-II e dados LiDAR**. 2010. 130p. (INPE-16712-TDI/1651). Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos. 2010.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v. 11, n. 1, p. 10-18, 2009.

Sousa, R. C. A.; Kux, H. J. H. Comportamento espectral e alvos urbanos: simulação com as bandas espectrais do satélite CBERS. SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12, 2005, Goiânia. **Anais...** São José dos Campos: INPE, 2005. p. 1-8. Disponível em: <http://www.obt.inpe.br/cbers/cbers_XIISBSR/351_comp_esp_urbano.pdf>. Acesso em: 05 Out 2010.

ALGORITMO BASEADO NO OPERADOR DE ENERGIA DE TEAGER (TEO) PARA SEGMENTAÇÃO DE LINHAS E PALAVRAS EM TEXTOS MANUSCRITOS E IMPRESSOS

Josimeire Tavares[1,2], Igor S. Peretta[1], Gerson F. M. Lima[1], Keiji Yamanaka[1], e Mônica S. Pais[1]

[1]*Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, C.P.593, Uberlândia/MG*
[2]*Instituto Luterano de Ensino Superior-ILES/ULBRA, Av. Beira Rio,1001,Itumbiara/GO*

Abstract

This paper presents a method for text line and word segmentation in freestyle handwritten, the SLP-TEO method. Results with this method applied to machine printed documents are also presented. This method uses the Teager Energy Operator (TEO) which is able to extract the envelope function of a signal consisting of two or more frequency components. In the proposed method (SLP-TEO), the operator TEO is applied to the projection histogram of a binarized text image and the returned values are used to perform the lines and words segmentation. Therefore, the SLP-TEO method is able to perform the text line and word segmentation in handwritten and printed documents using the same algorithm. The simplicity of the method does not prevent the high percentage of hits in the tests in handwritten documents: 98.5% in line segmentation and 89.8% in word segmentation. Comparison with a method based on the state-of-art Hough transform is also presented.

1. Introdução

Segmentação de linhas de texto e segmentação de palavras são alguns dos principais e mais críticos componentes de um sistema de reconhecimento de documentos impressos ou manuscritos. É primordial que se obtenha resultados satisfatórios nessas etapas para uma maior eficiência dos sistemas de reconhecimento óptico de caracteres (*Optical Character Recognition* - OCR).

Nos últimos anos, vários trabalhos têm sido divulgados sobre a segmentação de linhas de textos e segmentação de palavras em documentos manuscritos e impressos. Os autores têm focalizado na solução de segmentação de linhas e palavras em textos manuscritos por ser considerada ainda um problema crítico [9], [10].

Os documentos manuscritos apresentam vários problemas como: inclinação de linhas, diferença entre ângulos de linhas, vários tipos e tamanhos de letras, além de linhas sobrepostas ou fragmentadas [10]. Quanto à etapa de segmentação de palavras, esta ainda é uma tarefa difícil devido à falta de padronização dos espaços entre palavras, pontuação ao longo das linhas de texto, inclinação de palavras em uma mesma linha comprometendo a linha superior e inferior, além de letras e alguns tipos de acentuação presentes em algumas línguas como o português, espanhol, francês, turco entre outras [9].

Quanto ao processo em textos impressos, a segmentação de linhas e palavras se torna mais simples, pois este tipo de documento apresenta um padrão de espaçamento entre linhas, entre palavras e entre caracteres. Alguns métodos se tornaram padrão para esse tipo de texto, como os métodos baseados em projeção. Além disso, alguns trabalhos utilizam uma adaptação dos métodos baseados em projeção com a finalidade de segmentação de linhas de textos manuscritos como no trabalho [3].

Este trabalho apresenta um método para segmentação de linhas e segmentação de palavras de texto impresso ou manuscrito, utilizando o mesmo algoritmo para os dois tipos de texto. O método proposto é denominado "Segmentação de Linhas ou Palavras baseado no Operador de Energia de Teager" (SLP-TEO). As principais contribuições deste trabalho são: (a) Utilizar o mesmo algoritmo independente do tipo de texto, sem necessidade de ajustes para os textos manuscritos; (b) Utilização do Operador de Energia de Teager (*Teager Energy Operator* - TEO) para segmentação de linhas e palavras em textos; (c) Utilização do mesmo algoritmo para a segmentação de linhas e palavras. Neste último item, a única diferença está na definição do valor de referência — a variável altura da menor letra para segmentação de linhas e a largura da menor palavra para segmentação de palavras.

O método SLP-TEO é o mesmo tanto para segmentação de linhas quanto para segmentação de palavras, e aplicável tanto para textos impressos quanto para textos manuscritos. Neste trabalho, apresentamos resultados semelhantes aos resultados apresentados em [3], [9] tanto para segmentação de linhas como para segmentação de palavras. Quanto à segmentação de linhas em textos manuscritos, o percentual de acertos na base de dados utilizada foi de 98,5%. Para segmentação de palavras, obtivemos 89,8% de acerto para a mesma base. Este resultado é bastante expressivo quando comparado aos resultados de outros algoritmos, considerando a simplicidade do algoritmo do SLP-TEO e a praticidade no uso de um único algoritmo para textos manuscritos e impressos.

2. Trabalhos Relacionados

Muitos autores têm se empenhado em desenvolver métodos que atendam a necessidade de reconhecimento de textos, principalmente textos manuscritos onde há uma maior dificuldade nas etapas de segmentação.

Dentre as metodologias empregadas para a segmentação de linhas em documentos de texto manuscritos, podemos citar algumas categorias mais utilizadas: métodos baseados em projeção; métodos que utilizam manchas (*smearing*); e aplicações da transformada de Hough. A transformada de Hough é considerada como o estado da arte para segmentação linhas e palavras de textos manuscritos [10], [14].

No trabalho de Bar-Yosef, Hagbi e Kedem [2], são utilizados documentos degradados. A segmentação de linhas destes documentos é feita através do método de projeção. Basicamente o autor executa a soma dos valores dos pixels ao longo da direção horizontal para encontrar espaços entre linhas de texto através dos valores da projeção. Em seguida, são aplicadas técnicas estatísticas para se encontrar o mínimo local, e algoritmos de passagem por zero para detectar as intensidades dos picos. No trabalho de Louloudis, Gatos e Halatsis [10], a primeira etapa de processamento da imagem do texto inclui a binarização e o realce da imagem, extração de componentes conectados e estimação de altura do caractere. Na segunda etapa, uma transformada de Hough baseada em blocos é aplicada para detecção de possíveis linhas de texto. A terceira etapa visa corrigir possíveis fragmentações para detectar linhas de texto que não haviam sido detectadas previamente. Finalmente, caracteres conectados verticalmente são separados e designados às suas linhas de texto de origem. Em outro exemplo, no trabalho de Saha et al. [14], temos uma primeira etapa de detecção de bordas através do algoritmo *Edge Sobel*. Aplicando-se a transformada de *Hough* com os parâmetros ajustados à imagem binarizada, temos as linhas de texto identificadas como componentes conectados. Essas linhas são então segmentadas por um algoritmo do tipo *Connected Component Labeling* (CCL). Cada linha de texto segmentada é então submetida a essas mesmas etapas, utilizando-se parâmetros específicos para a segmentação de palavras.

Quanto à segmentação de palavras, muitos métodos têm sido utilizados, como os métodos baseados em medidas estatísticas associados aos métodos de projeção [1]. Normalmente, os métodos baseados em projeção são associados a diferentes medidas de distâncias — *convex Hull-based*, distância Euclidiana, dentre outras. Liwicki, Scherz e Bunke [8] apresenta um método onde as palavras são segmentadas utilizando as distâncias dos componentes adjacentes de uma linha. As distâncias utilizadas nesse trabalho são a distância Euclidiana e a *convex Hull-based*.

3. Base de Dados

Foi adotada a base de dados *IAM - Database* [11], uma conceituada base de imagens para reconhecimento de documentos de textos manuscritos. A base de dados contém formulários de textos com uma parte impressa e a transcrição em escrita manual do mesmo texto por vários autores. Cada texto desta base foi digitalizado com resolução de 300 dpi e armazenado no formato de imagem PNG com 256 níveis de cinza.

A Figura 1 apresenta como exemplo um dos textos dessa base de dados utilizada neste trabalho.

Figura 1. Formulário extraído IAM-Database.

4. Metodologia

4.1. Operador de Energia de Teager (TEO)

Teager e Teager, em um trabalho sobre modelamento não-linear da voz [15], apresentaram um operador de energia que mais tarde foi denominado como operador de energia de Teager (*Teager Energy Operator* - TEO), ou operador Teager-Kaiser. Este operador, de acordo com Kaiser [5], ao ser aplicado a um sinal composto por uma única frequência variante no tempo, é capaz de extrair a medida de energia do processo mecânico que gerou esse sinal.

O operador de energia de Teager é definido no domínio discreto [6] como:

$$\Psi\left[x\left(n\right)\right] = x_n^2 - x_{n+1} \cdot x_{n-1}, \tag{1}$$

onde Ψ é o operador de Teager e x_n é o valor da n-ésima amostra do sinal abstraído a partir da projeção linear da imagem analisada.

Uma importante característica do TEO analisada por Kaiser [5] é que, quando aplicado a sinais compostos por duas ou mais componentes de frequência, "seria como se o algoritmo [TEO] fosse capaz de extrair a função envelope do sinal" [tradução nossa]. Como exemplo, a dissertação de Peretta [12] utiliza esta característica do TEO, a de extrair a função envelope de um sinal, para detectar as fronteiras de um comando de voz inserido em um sinal de áudio. Outro trabalho deste grupo utiliza essa mesma premissa [13].

4.2. TEO aplicado ao sinal do histograma

Ao considerarmos os valores de projeção (histograma) de uma imagem de texto binarizada como sendo um sinal amostrado, pela teoria de séries de Fourier [4] , este sinal (função arbitrária dentro de dado intervalo) pode ser expresso por uma série de senos e cossenos. Ou seja, teremos como resultado dessa abstração um sinal composto por diversas componentes de frequência. Se aplicarmos o TEO a este sinal, seremos capazes de extrair a função envelope deste sinal abstrato. Esta é a base teórica do método SLP-TEO.

O método SLP-TEO é aplicado aos valores de projeção horizontal da imagem do texto para segmentação de linhas de texto. Para segmentação de palavras, o método é aplicado aos valores de projeção vertical da imagem de cada linha de texto segmentada.

5. Projeto Desenvolvido

5.1. Pré-processamento

Para esta pesquisa, de cada imagem de texto da base de dados utilizada, foram extraídos em separado os textos impressos e manuscritos. Dois conjuntos de testes foram preparados. O primeiro conjunto foi constituído apenas por imagens de texto impresso e o segundo, apenas os manuscritos.

Em seguida, estas imagens foram binarizadas utilizando-se o método de Otsu [4] para determinação do limiar global de cada imagem.

5.2. Segmentação de Linhas

O algoritmo de segmentação de linhas recebe como entrada uma imagem binarizada de texto impresso ou manuscrito. A projeção (histograma) horizontal desta imagem é determinada. Em seguida, seus valores são normalizados dentro dos limites [0,1]. O operador TEO é aplicado a esses valores como se fossem parte de um sinal amostrado, gerando um vetor com os valores resultantes desta operação. A seguir, o método SLP-TEO converte os valores deste vetor em valores absolutos. Este vetor é designado como vetor ABS-TEO.

Em seguida, é necessário encontrar as fronteiras das linhas de texto. Essas fronteiras são constituídas pelos limites superior e inferior de cada linha de texto contida na imagem binarizada. A altura mínima (Hmin) para uma linha de texto válida foi definida como 6 pixels, considerando que seria a altura mínima para representar as letras do alfabeto. Também foi definido empiricamente o valor do limiar necessário para permitir que o método SLP-TEO detecte linhas de texto ($1,25 \cdot 10^{-3}$). O fluxograma na Figura 2 detalha o método SLP-TEO para detecção de linhas.

Cada linha da imagem de texto possui um correspondente no vetor ABS-TEO. O método SLP-TEO varre o vetor ABS-TEO à procura de fronteiras. Se o valor correspondente da linha no vetor é maior ou igual ao limiar definido e se a variável de controle é falsa, o método marca aquela linha da imagem como sendo a fronteira **superior** de uma linha de texto. Se o valor correspondente da linha no vetor é menor do que limiar definido e se a variável de controle é verdadeira, o método marca aquela linha da imagem como sendo a fronteira **inferior** de uma linha de texto. Ao encontrar uma fronteira inferior, o método checa ainda se a altura em pixels encontrada é maior do que a altura mínima pré-determinada. Se sim, as fronteiras superior e inferior são validadas. Se não, as fronteiras encontradas são removidas.

Em uma segunda etapa do método SLP-TEO é feito o ajuste das fronteiras encontradas. Esse ajuste é importante, pois é uma forma de se adequar a fronteira encontrada com a altura das linhas da imagem do texto. É verificado através dos pontos definidos pela fronteira a possibilidade de "inflar" essa fronteira, sem contudo, exceder os limites da imagem e respeitando os limites das fronteiras detectadas acima ou abaixo da fronteira analisada no momento.

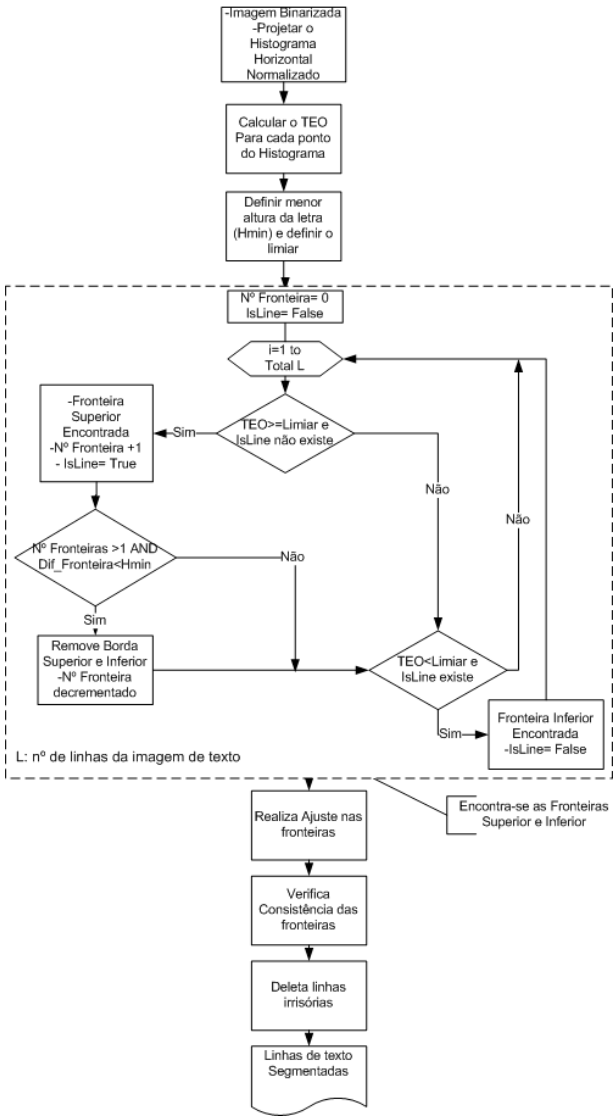


Figura 2. Fluxograma do Algoritmo de Segmentação de Linhas.

Na terceira etapa, é verificada a consistência das linhas encontradas. Neste processo, é calculado a altura média das linhas de texto encontradas. Se entre uma fronteira superior e a fronteira inferior exatamente anterior não houver ao menos 1 pixels entre elas e ao menos uma das alturas das linhas de texto analisadas for menor ou igual à altura média calculada, então essas fronteiras são agrupadas.

5.3. Segmentação de Palavras

O algoritmo de segmentação de palavras recebe como entrada uma imagem binarizada de cada linha de texto previamente segmentada. A projeção (histograma) vertical desta imagem é determinada. Em seguida, seus valores são normalizados dentro dos limites [0,1]. O operador TEO é aplicado a esses valores como se fossem parte de um sinal amostrado composto por várias componentes de frequência, gerando um vetor com os valores resultantes desta operação. A seguir, o método SLP-TEO converte os valores deste vetor em valores absolutos. O método é o mesmo utilizado para segmentação de linhas de texto (incluindo seu fluxograma), apenas com as seguintes modificações:

- Ao invés de uma altura mínima (Hmin), é definido uma largura mínima (Lmin) através da Equação 2;
- O método localiza fronteiras direita e esquerda ao invés de superior e inferior.

$$L_{min} = \max(10, \lceil 0.15 \cdot H \rceil), \qquad (2)$$

onde L_{min} é a largura mínima (em pixels) necessária ao método e H é a altura (em pixels) da imagem da linha de texto segmentada.

Os mesmos ajustes realizados para a detecção de fronteiras das linhas de texto na subseção 5.2 são aplicados para a detecção de fronteiras das palavras a serem segmentadas.

6. EXPERIMENTOS E RESULTADOS

Dois tipos de experimentos foram conduzidos. No primeiro experimento, foram utilizadas 150 imagens binarizadas de textos manuscritos retiradas da base de dados *IAM-Database* [11]. As imagens utilizadas para este experimento contém 1.054 linhas de texto. No segundo experimento, foram utilizadas 150 imagens binarizadas de textos impressos, retiradas da mesma base de dados.

Na primeira etapa do primeiro experimento, dois métodos de segmentação de linhas de texto foram utilizados, com o objetivo de comparação: o método baseado na transformada de Hough [14] (BTH) e o método SLP-TEO. Para o método BTH, obtivemos um total de 86,2% de linhas segmentadas corretamente, resultado este bastante coerente com os divulgados na literatura recente. Para o método SLP-TEO, os resultados encontrados mostram uma maior eficiência, pois 98,5% do total de linhas foram segmentadas corretamente. Note que, para textos manuscritos, este percentual normalmente não é atingido.

Na segunda etapa, os mesmos dois métodos de segmentação foram utilizados para segmentação de palavras. Pequenos ajustes em ambos os métodos foram necessários. Também foi respeitada a premissa de que a medida da correta segmentação de palavras precisaria ser efetuada dentro de imagens de linhas de texto corretamente segmentadas.

Note que ocorreram apenas 6 casos dentro do universo das 150 imagens iniciais em que o método BTH atingiu 100% de correta segmentação de linhas sem que o método SLP-TEO também tenha atingido essa porcentagem. Do outro lado, foram 104 casos onde o método SLP-TEO obteve 100% de sucesso na segmentação de linhas sem que o método BTH tenha atingido essa porcentagem.

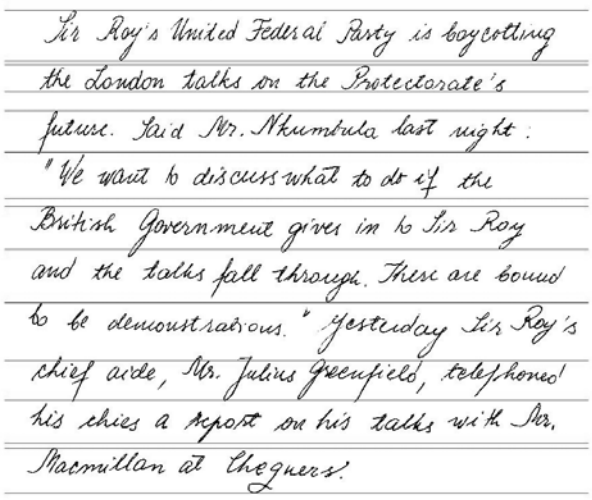
Para esta segunda etapa, portanto, foram utilizadas apenas 26 imagens binarizadas de textos manuscritos das 150 originais. Estas imagens de texto foram escolhidas porque tanto o método BTH quanto o método SLP-TEO atingiram 100% das linhas de texto corretamente segmentadas. Com isso, trabalhamos com um total de 1.733 palavras a serem segmentadas. Para o método BTH, obtivemos um total de 81,2% de correta segmentação dessas palavras, enquanto que, para o método SLP-TEO, obtivemos um total de 89,8% de correta segmentação.

Como complemento, aplicamos o método SLP-TEO para segmentação de palavras às 130 imagens binarizadas em que houve 100% de correta segmentação de linhas por este mesmo método. Como resultado, de um total de 8526 palavras, 87,5% foram corretamente segmentadas.

A Figura 3(a) mostra uma imagem de texto manuscrito cujas linhas foram segmentadas pelo método SLP-TEO. Neste caso, mesmo que as linhas de texto apresentem diferentes extensões verticais, o método SLP-TEO consegue delimitá-las corretamente. A Figura 3(b) apresenta a segmentação de palavras pelo método SLP-TEO.

No segundo experimento, aplicamos o método SLP-TEO em 150 imagens binarizadas de textos impressos retiradas da base de dados *IAM-Database* [11][1]. Não houve modificações no algoritmo para a adaptação aos textos impressos, ou seja, os mesmos parâmetros que foram utilizados para textos manuscritos foram utilizados para os textos impressos. Como resultados deste experimento, foi obtida uma correta segmentação de 99,5% no total de 781 linhas, e 99,9% no total de 9831 palavras. A segmentação de palavras foi limitada apenas às 147 imagens de texto que tiveram 100% de suas linhas de texto corretamente segmentadas pelo método SLP-TEO. A Figura 6 apresenta um exemplo de imagem texto impresso cujas linhas e palavras foram segmentadas pelo método SLP-TEO.

O sumário com o comparativo dos resultados obtidos está apresentado na Tabela 1.



(a) Linhas Segmentadas pelo método SLP-TEO.

(b) Palavras Segmentadas pelo método SLP-TEO.

Figura 3. Segmentação de Texto Manuscrito Utilizando-se o SLP-TEO.

Texto	Segmenta	#Textos	BTH	SLP-TEO
Manuscrito	Linhas	150	86,2%	98,5%
Manuscrito	Palavras	26	81,2%	89,8%
Manuscrito	Palavras	130	∅	87,5%
Impresso	Linhas	150	∅	99,5%
Impresso	Palavras	147	∅	99,9%

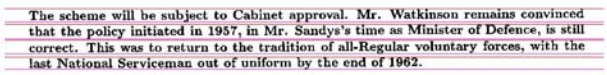
Tabela 1. Resultados dos métodos BTH e SLP-TEO para segmentação de linhas e palavras.

7. Conclusão

Baseado na capacidade do operador de energia de Teager (TEO) de extrair a função envelope do sinal com múltiplas componentes de frequência e aplicando este operador ao "sinal" gerado com os valores das projeções lineares de uma imagem de texto, o método SLP-TEO foi desenvolvido e utilizado para a segmentação de linhas e palavras de textos manuscritos ou impressos.

O método SLP-TEO apresentou resultados bastante interessantes quanto a segmentação de linhas tanto para textos manuscritos como para textos impressos. Um dos diferenciais do método está no fato de que um mesmo algoritmo é utilizado para ambos os tipos de texto. Além

1 O método BTH não é indicado na literatura para segmentação de textos impressos.



(a) Linhas Segmentadas.

(b) Palavras segmentadas.

Figura 4. Segmentação de texto impresso com o método SLP-TEO.

disso, o método SLP-TEO demonstrou sua eficiência na segmentação de palavras para textos manuscritos ou impressos, independente de qualquer mensuração de distâncias entre palavras e/ou caracteres.

Prático, eficiente e relativamente simples, o método SLP-TEO teve um bom desempenho em relação à segmentação de palavras em textos manuscritos, desempenho este superior aos métodos mais recentemente utilizados. Note que esta é uma etapa que ainda hoje é considerada problemática em segmentação de imagens de texto.

Segundo Li [7], não é fácil estender um algoritmo de segmentação utilizado em textos impressos adequando-o para segmentação de textos manuscritos. Neste trabalho, mostramos que um mesmo algoritmo pode ser utilizado tanto para segmentação de linhas quanto para palavras, independentemente da natureza do texto, seja ele impresso ou manuscrito.

Alguns ajustes ao método ainda são necessários. O método SLP-TEO realiza uma separação linear das linhas, "cortando" componentes conectados em um ponto aproximadamente equidistante das concentrações de pixels de duas linhas adjacentes. Ainda é objeto de pesquisa o aprimoramento do método para a correta separação de componentes conectados entre linhas de um determinado texto. Esta funcionalidade provavelmente será implementada em uma etapa de pós-processamento. Além disso, serão necessários testes com textos que possuam uma inclinação muito acentuada, ou curvilínea — a base de dados utilizada não apresenta exemplos com esse tipo de problema.

7.1. References

Referências

[1] E. Ataer and P. Duygulu. Retrieval of Ottoman documents. In *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, MIR '06, pages 155–162, Santa Barbara, California, USA, Oct 2006. ACM.

[2] I. Bar-Yosef, N. Hagbi, K. Kedem, and I. Dinstein. Line Segmentation for Degraded Handwritten Historical Documents. In *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*, ICDAR '09, pages 1161–1165, Washington, DC, USA, 2009. IEEE Computer Society.

[3] R. dos Santos, G. Clemente, T. I. Ren, and G. Cavalcanti. Text Line Segmentation Based on Morphology and Histogram Projection. In *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*, pages 651 –655, july 2009.

[4] R. C. Gonzalez and R. E. Woods. *Processamento de Imagens Digitais*. Editora Blucher, São Paulo, 2000.

[5] J. Kaiser. On a simple algorithm to calculate the 'energy' of a signal. In *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, pages 381–384, Abril 1990.

[6] J. Kaiser. Some useful properties of Teager's energy operators. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 3, pages 149 –152, apr 1993.

[7] Y. Li, Y. Zheng, D. Doermann, and S. Jaeger. A New Algorithm for Detecting Text Line in Handwritten Documents. pages 35–40, La Baule, France, 2006.

[8] M. Liwicki, M. Scherz, and H. Bunke. Word Extraction from On-Line Handwritten Text Lines. 2008.

[9] G. Louloudis, B. Gatos, and C. Halatsis. Text Line Detecting in Unconstrained Handwritten Documents Using a Block-Based Hough Transform Approach. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 2, pages 599 –603, 2007.

[10] G. Louloudis, B. Gatos, I. Pratikakis, and C. Halatsis. Text line and word segmentation of handwritten documents. *Pattern Recognition*, 42(12):3169 – 3183, 2009. New Frontiers in Handwriting Recognition.

[11] U. V. Marti and H. Bunke. The IAM-database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5:39–46, 2002. 10.1007/s100320200071.

[12] I. S. Peretta. A Novel Word Boundary Detector based on the Teager Energy Operator for Automatic Speech Recognition. Master's thesis, Universidade Federal de Uberlândia, Dezembro 2010.

[13] I. S. Peretta, G. F. M. Lima, J. A. Tavares, and K. Yamanaka. A Spoken Word Boundaries Detection Strategy for Voice Command Recognition. *Learning and Nonlinear Models*, 8(3):148–156, 2010.

[14] S. Saha, S. Basu, M. Nasipuri, and D. K. Basu. A Hough Transform based Technique for Text Segmentation. *Journal of Computing*, abs/1002.4048, Feb 2010.

[15] H. M. Teager and S. M. Teager. Evidence for Nonlinear Production Mechanisms in the Vocal Tract. In *NATO Advanced Study Institute on Speech Production and Speech Modeling*, Julho 1989.

Diferenciação da *Greening* do citros de outras doenças foliares a partir de descritores de cor e forma

Patricia Pedroso Estevam Ribeiro[1]; Lúcio André de Castro Jorge[2]; Maria Stela Veludo de Paiva[3].

1- Mestrando, Escola de Engenharia de São Carlos - Universidade de São Paulo - USP, Depto. de Engenharia Elétrica, São Carlos-SP, patriciapedrosoestevam@hotmail.com
2- Pesquisador, Embrapa Instrumentação Agropecuária, São Carlos – SP, lucio@cnpdia.embrapa.br
3- Professora, Escola de Engenharia de São Carlos - Universidade de São Paulo - USP, Depto. de Engenharia Elétrica, São Carlos-SP, mstela@sc.usp.br

Abstract

The citrus culture since the beginning of the twentieth century has been subjected to several attacks of pests and diseases, among these diseases, Greening or Huanglongbing (HLB), is currently considered the most serious disease of citrus worldwide. Until now there isn't cure or treatment, and disease control is currently carried out by visual inspection method. In order to help the visual inspection method, this work test image processing techniques to help in differentiating the Greening from other diseases, by color image segmentation of leaves and classification based on shape descriptors. The results show that only the yellow stain is not a strong differentiator of this disease.

1. Introdução

A citricultura paulista, desde o início do século XX, tem sido exposta a vários ataques de pragas e doenças, e recentemente de uma forma mais intensa, apesar do crescimento da produção de laranja, (cerca de 318,6 milhões de caixas em 2010, segundo os dados do Instituto de Economia Agrícola), ocasionando uma perda de 10% da produção média nos últimos anos [1], [9].

Dentre essas doenças, a *Greening,* também conhecida como *Huanglongbing* (HLB), é considerada atualmente a mais grave doença dos citros no mundo [2]. Causada pela bactéria *Candidatus* Liberibacter *spp.,* é transmitida pelo inseto psilídeo *Diaphorina citri,* que adquire e transmite a bactéria às demais plantas ao se alimentar de uma planta já contaminada [3].

Por não possuir cura ou tratamento e por não existir variedade comercial de copa ou porta-enxerto resistente à doença, o controle da *Greening* só pode ser feito com inspeção constante, eliminação imediata de plantas com sintomas e o controle do inseto transmissor. Atualmente, o método de inspeção visual e a análise do PCR (*Polymerase Chain Reaction*) [4], são os mais utilizados para diagnosticar a doença. O método PCR é utilizado para diagnosticar o patógeno de plantas suspeitas, mas o custo elevado e o longo tempo para a análise, o tornam proibitivo de ser aplicado em escala necessária para o controle. As inspeções visuais são realizadas por inspetores caminhando a pé ou em plataformas movidas por tratores ao lado das plantas cítricas, como apresentado na Figura 1. Apesar de ser o método mais aplicado atualmente, sua eficácia depende de vários fatores, tais como, o conhecimento e prática na detecção de plantas sintomáticas, época do ano, genótipo e altura das plantas, incidência de raios solares nas plantas e no rosto do inspetor, apresentando em média 47,61% de precisão na detecção de plantas sintomáticas [23].

(a) (b)

Figura1. Inspeções visuais realizadas em pomares de citros, em a) realizada a pé, b) realizada por plataformas movidas por tratores [3].

O sintoma característico da *Greening* aparece inicialmente, em alguns ramos apresentando folhas mosqueadas (manchas de formas irregulares, mescladas com o verde amareladas no fundo verde). Estas manchas são facilmente confundidas com outras doenças e deficiências nutricionais que se assemelham às características visuais da *Greening* [3] como por exemplo, as doenças CVC (Clorose Variegado dos Citros), Rubelose e as deficiências de Magnésio, Manganês e Zinco, confundindo os inspetores na identificação destas doenças pelo método visual. Nos experimentos apresentados por [11], [12], [13] é destacado a variação da mensuração entre os inspetores, evidenciando a necessidade de técnicas complementares.

Com os avanços na área de processamento digital de imagem, é possível fazer uso de métodos computacionais, que auxiliem na diferenciação destas doenças. Basset et al., [5], utilizam técnicas de visão computacional para inspeção da qualidade de produtos. Yohekawa et al., [7], verificam que os fatores de forma são uteis para a identificação de plantas por meio de suas folhas.

Este trabalho tem como objetivo desenvolver uma ferramenta baseada em descritores de cor e forma, para auxiliar na diferenciação da *Greening* e outras doenças em citros, possibilitando um diagnóstico mais rápido e preciso.

O presente artigo está organizado nas seguintes seções: a seção 2 aborda a teoria relacionada ao trabalho, na seção 3 os materiais e métodos, na seção 4 resultados, discussão e finalmente na seção 5 a conclusão.

2. Extração de atributos

2.1. Segmentação por cor

A segmentação de uma imagem consiste em subdividir uma imagem em seus componentes básicos, com as características mais relevantes, sendo que estas características dependem do objeto de interesse.

Para a segmentação por cor, encontram-se, na literatura, inúmeras aplicações com o uso de redes neurais artificiais (RNA), com bons resultados [16], [17]. Em Cavani et al., [19] e Simões et al., [18] é utilizada segmentação por cor em imagens de frutas, utilizando de uma RNA do tipo *Perceptron* multicamada (PMC) [22].

2.2. Descritores de forma

Em processamento de imagens, a aplicação de descritores de forma, permite analisar e extrair características e parâmetros dos objetos da imagem.

2.2.1. Razão de aspecto

Razão de aspecto é a razão entre o eixo maior e eixo menor do objeto selecionado [20], como por exemplo, 4:3 ou 16:9, podendo ser observado na Figura 2, e implementado pela Equação 1 [8].

Figura 2. Ilustração da região representada pela razão de aspecto.

$$\text{Razão de Aspecto} = \frac{\text{Diâmetro Máximo}}{\text{Diâmetro Mínino}} \qquad (1)$$

2.2.2. Diâmetro médio

São as medidas do comprimento do diâmetro medido a cada 2 graus de intervalo, passando pelo centróide do objeto, podendo ser observado na Figura 3 [20].

Figura 3. Ilustração da região representada pelo diâmetro médio.

2.2.3. Razão do raio (*Radius ratio*)

E a razão entre o raio máximo e o raio mínimo, em relação aos pontos centrais do objeto [20]. Podendo ser observado na Figura 4.

Figura 4. Ilustração da região representada pela razão do raio.

2.2.4. Aspecto arredondado (*Roundness*)

Redondeza mede o a espessura média do objeto selecionado [20]. Podendo ser observado na Figura 5, Equação 2.

Figura 5. Ilustração da região representada redondeza.

$$\text{Aspecto Arredondado} = \frac{\text{Perímetro}^2}{4.\pi.\text{Área}} \qquad (2)$$

2.2.5. *Feret* médio

O diâmetro *feret* é à medida que caracteriza o tamanho do objeto selecionado [8]. O *feret* médio é a média dos *ferets* em várias direções.

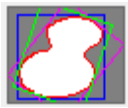


Figura 6. Ilustração da região representada pelo diâmetro *feret*.

3. Materiais e métodos

A Figura 7 apresenta as etapas de processamentos aplicadas nas imagens foliares, sendo estas: aquisição, segmentação por cor, extração dos descritores de forma e classificação. A classificação visava a obtenção de um

classificador para diferenciar espacialmente as manchas amarelas das amostras das doenças e deficiências cujos sintomas mais se assemelham ao *Greening*.

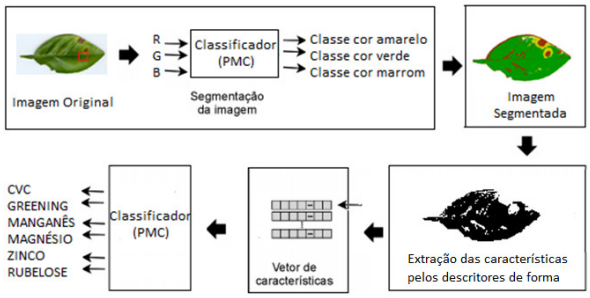


Figura 7. Etapas de processamento: aquisição, segmentação por cor, extração de atributos e classificação das doenças.

3.1. Aquisição dos dados

Foram fornecidas pela empresa FISHER seis tipos de folhas de citros com sintomas de doenças e deficiência nutricional, conforme a Figura 8. Estas amostras foram selecionadas por um técnico agrônomo denominado pragueiro, que identificou os sintomas, apenas por meio da inspeção visual, segundo instruções do manual técnico do *Greening*, fornecido pela FUNDECITRUS [3].

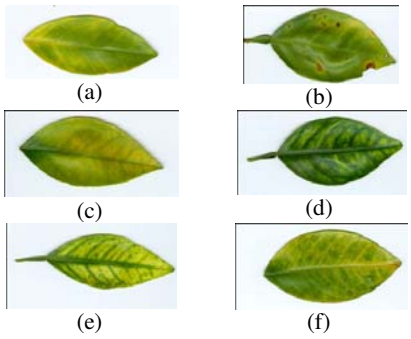


Figura 8. Amostras foliares com a) *Greening*, b) CVC, (c) Deficiência de Manganês, (d) Deficiência de Magnésio, (e) Deficiência de zinco, (f) Rubelose.

Conforme apresentado na Figura 8, os sintomas são:

- *Greening:* apresenta folhas mosqueadas ou clorose assimétrica.
- Deficiência de Magnésio: amarelecimento em forma de "V" invertido.
- Deficiência de Manganês: surge clorose entre as nervuras.
- Deficiência de Zinco: apresenta clorose acentuada do limbo entre as nervuras.
- CVC (Clorose Variegado dos Citros): apresenta pequenas manchas amareladas e irregulares, espalhada na frente, e lesões de cor palha nas costas da folha.
- Rubelose: apresenta lesão nas forquilhas dos ramos principais; as folhas da copa tornam-se amareladas.

Foram selecionadas 60 amostras de folhas, sendo 10 amostras para cada tipo de doença/deficiência.

Estas amostras obtidas foram digitalizadas por um scanner fotográfico de mesa do modelo da HP *Scanjet* G4050, com resolução de 100 DPI, dimensão de 400 x 200 pixels, utilizando somente a parte frontal da folha.

3.2. Segmentação

A segmentação por cor foi utilizada para separar as manchas amarelas do fundo verde das folhas e de áreas com alguma necrose, caracterizada pela cor marrom.

Foi aplicada a RNA com o algoritmo *backpropagation*, com uma camada de entrada, três camadas escondidas e três camadas saídas. Para o treinamento da rede foram utilizadas 46 amostras de um total de 60, sendo o restante utilizado para teste. Os parâmetros que melhor se ajustaram no treinamento foram: 0,3 para a taxa de aprendizado, 0,2 para o momento e 500 para a quantidade de épocas. Foi utilizado o esquema de validação cruzada com 10 *folds* e função de ativação sigmóidal.

Os parâmetros de entrada da RNA foram às componentes de cor RGB de cada pixel das amostras selecionadas sobre a imagem, representando as cloroses, padrões de verde e necroses. Como saída, cada pixel analisado foi rotulado como sendo da classe amarelo, verde e marrom, correspondentes às manchas de clorose, área sadia e necrose, respectivamente. Um exemplo de imagem segmentada pode ser observada na Figura 9 (a).

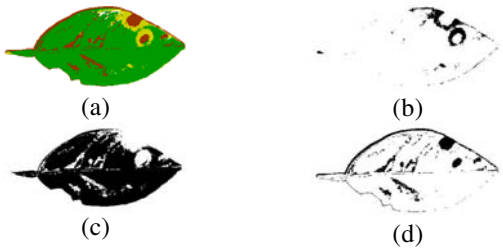


Figura 9. Apresenta as imagens das folhas segmentada por cor, em 3 classes de cores, a) imagem segmentada, b) somente amarelo, c) somente verde, d) somente marrom.

Após a imagem ser segmentada por cores (amarelo, verde e marrom), foram utilizadas cada imagem binária, geradas das porções amarela, verde e marrom, da imagem como apresentado nas Figuras 9(b), 9(c), 9(d).

3.3. Extração dos descritores

Por meio das imagens binarizadas foram extraídos os valores estatísticos médio, máximo e mínimo das formas das manchas para os descritores de forma citados na seção 2.2. Os valores médio, máximo e mínimo são demonstrados pelas Equações 3, 4 e 5, como segue:

$$Média = \frac{1}{n} . \sum_{i=0}^{n} f(x_i) \qquad (3)$$

Na Equação 3, *n* corresponde ao número de objetos ou manchas da imagem binarizada *f*(*x*).

$$Máximo = \text{Max } f(x_i) \qquad (4)$$

Na Equação 4, o valor Máximo representa o maior valor dentre os objetos ou manchas encontrados na imagem binarizada *f*(*x*).

$$Mínimo = \text{Min } f(x_i) \qquad (5)$$

Na Equação 5, o valor Mínimo representa o menor valor dentre os objetos ou manchas encontrados na imagem binarizada *f*(*x*).

Em seguida foram criados oito vetores de características descritos na Tabela 1.

Tabela 1. Tabela com a descrição de cada vetor.

Vetor	Cor	Atributos	Instâncias
1°	Amarelo	15	60
2°	Verde	15	60
3°	Marrom	15	60
4°	Amarelo+verde	30	60
5°	Amarelo+verde+marrom	45	60
6°	Amarelo+verde+marrom	45	50
7°	Amarelo+verde+marrom	45	40
8°	Amarelo+verde+marrom	45	30

Do primeiro ao quinto vetor, foram utilizadas seis classes de doenças: CVC, Magnésio, Manganês, Zinco, *Greening*, Rubelose.

Para o sexto vetor foram utilizadas cinco classes de doenças: CVC, Magnésio, Manganês, Zinco, *Greening*.

Para o sétimo vetor foram utilizadas quatro classes: CVC, Manganês, Zinco, *Greening*.

Para o oitavo vetor foram utilizadas três classes de doenças: Manganês, Zinco, *Greening*.

3.4. Classificação

Para a classificação de padrões foi utilizada a API (*Application Programming Interface*) do *toolbox* WEKA [14], para construção do classificador foi utilizado uma RNA do tipo *Perceptron* multicamada (PMC) com o algoritmo *backpropagation* [15].

4. Resultados e discussão

O conjunto de classificação utilizou a combinação de atributos das classes de doenças, buscando encontrar a influência da distribuição das cores e forma na classificação entre as classes de doenças.

Durante o processo de classificação foram utilizados 80% das instâncias para o treinamento da RNA e 20% das instâncias para os testes.

4.1. Influência das Cores

Para verificar a influência da distribuição espacial das cores nas folhas, como por exemplo, a distribuição da *Greening* que apresenta as manchas amarelas assimétricas nas folhas, foi aplicada uma RNA e os parâmetros da mesma foram ajustados com base nos testes realizados previamente. Dessa forma, para a RNA foi utilizado uma camada de entrada, três camadas neurais escondidas e uma camada de saída. Para a taxa de aprendizado foi utilizado o valor 0,3, para o momento 0,2 e o número de épocas foi limitado a 500. Em todos os testes foi utilizado o esquema de validação cruzada com 10 *folds*, e a função de ativação foi a sigmóide.

A Tabela 2 apresenta os resultados da classificação geral. Foram utilizadas 60 instâncias para cada vetor, desse total, o classificador conseguiu classificar corretamente trinta e duas instâncias para o 1º vetor, representando um percentual de 53,33%, vinte e seis para o 2° vetor, representando 43,33%, dezesseis para o 3º vetor, representando 26,67%, trinta e três para o 4º vetor, representando 55% e trinta e oito para o 5º vetor, representando 63,33%. Dessa forma é possível observar que apenas a distribuição de uma única cor na folha não é o suficiente para diferenciar a *Greening* de outras doenças. O quinto vetor, por exemplo, é composto pelas cores amarelo, verde e marrom, e apresenta o melhor resultado (63,33%) na classificação geral das instâncias se comparado aos demais vetores. Os valores obtidos para o erro quadrático médio (EQM) e o valor do erro absoluto médio (EAM) para este vetor também foram inferiores aos demais valores obtidos com os outros vetores.

Tabela 2. Resultados EQM e EAM obtidos na classificação da RNA para cada classe de cor do 1º ao 5° vetor.

Vetor	Cor	EQM	EAM	Classificação Correta (%)
1º	Amarelo	32,70%	15,43%	53,33%
2º	Verde	37,87%	20,46%	43,33%
3º	Marrom	42,47%	24,70%	26,67%
4º	Amarelo+verde	33,74%	15,36%	55%
5º	Amarelo+verde+marrom	31,59%	14,84%	63,33%

Na Tabela 3 para o quarto vetor o classificador obteve para a CVC a precisão de acertos de 71,40%, para o Magnésio 42,90%, para o Manganês 60%, para o Zinco 58,30%, para a *Greening* 57,10% e para a Rubelose 50%. Para o quinto vetor o classificador obteve para a CVC 63,60%, para o Magnésio 57,10%, para o Manganês 63,60%, para o Zinco 69,20%, para a *Greening* 50% e para a Rubelose 60%.

Dessa forma observa-se que os melhores resultados obtidos com o classificador só foram obtidos mediante a combinação das cores distribuídas nas folhas.

Tabela 3. Valores de Precisão obtidos na classificação por RNA para cada classe de cor em relação às doenças do 1° ao 5° vetor.

		Precisão					
Vetor	Cor	CVC	Magnésio	Manganês	Zinco	Greenin	Rubelose
1°	Amarelo	50,00%	70,00%	58,30%	50,00%	41,70%	50,00%
2°	Verde	25,00%	44,40%	41,70%	60,00%	36,40%	50,00%
3°	Marrom	37,50%	0	30,00%	30,00%	31,30%	28,60%
4°	Amarelo+verde	71,40%	42,90%	60,00%	58,30%	57,10%	50,00%
5°	Amarelo+verde+marrom	63,60%	57,10%	63,60%	69,20%	50,00%	70,00%

4.2. Influência das doenças

Conforme mencionado na seção 3.3 que descreve os vetores de características utilizados neste trabalho, a similaridade presente nas doenças leva a maiores erros no processo de classificação. A Tabela 4 apresenta os resultados obtidos com o classificador para os vetores 6º, 7º e 8º. O oitavo vetor apresenta o melhor resultado para a classificação, pois, o mesmo utiliza apenas as classes Magnésio, Zinco e Greening. Das trinta instâncias analisadas do 8º vetor, vinte e cinco delas foram classificadas corretamente, representando um percentual de 83,33% de acerto. O valor do EQM e do EAM foram respectivamente 31,75% e 15,39%.

Tabela 4. Resultados EQM e EAM obtidos na classificação da RNA para cada classe de cor do 6º ao 8° vetor.

Vetor	Cor	EQM	EAM	Classificação Correta (%)
6°	Amarelo+verde+marrom	32,15%	15,23%	62%
7°	Amarelo+verde+marrom	35,21%	17,57%	70%
8°	Amarelo+verde+marrom	31,75%	15,39%	83,33%

A Tabela 5 apresenta os resultados obtidos com a precisão dada pelos classificadores para cada classe de doença. Para a CVC a precisão dada no sexto vetor foi de 66,7%, para o sétimo vetor foi de 75%. Para o Magnésio a precisão para o sexto vetor foi de 60% e para o oitavo vetor 100%. Para o Manganês a precisão do sexto vetor foi de 63,60% e para o sétimo vetor foi de 58,30%. Para o Zinco a precisão do sexto vetor foi de 57,10%, para o sétimo vetor 87,50% e para o oitavo vetor foi de 80%. Para a *Greening* a precisão do sexto vetor foi de 63,60%, para o sétimo vetor 66,70% e para o oitavo vetor 72,70%.

Dessa forma, é possível observar que o menor número de doenças melhora os resultados obtidos com o classificador.

Tabela 5. Valores de Precisão obtidos na classificação por RNA para cada classe de cor em relação às doenças do 6° ao 8° vetor.

		Precisão					
Vetor	Cor	CVC	Magnésio	Manganês	Zinco	Greenin	Rubelose
6°	Amarelo+verde+marrom	66,70%	60,00%	63,60%	57,10%	63,60%	Não Usada
7°	Amarelo+verde+marrom	75,00%	Não Usada	58,30%	87,50%	66,70%	Não Usada
8°	Amarelo+verde+marrom	Não Usada	100,00%	Não Usada	80,00%	72,70%	Não Usada

5. Conclusão

A aplicação da RNA apresenta um melhor resultado na classificação das classes quando utilizada para as classes que mais se assemelham ao *Greening*, neste caso as classes Manganês e Zinco. A acurácia obtida para essas instâncias foi de 83,33%.

Este trabalho mostrou que apenas a cor amarela e sua distribuição espacial, indicada pelos descritores de forma não consegue diferenciar totalmente a *Greening* de outras doenças. Para uma melhor diferenciação destas doenças, faz-se necessário a combinação da distribuição das cores amarelo, verde e marrom.

Para trabalhos futuros, pretende-se aumentar o numero de amostras foliares e o número de atributos extraídos, verificando assim o descritor de forma que melhor auxilia na diferenciação das doenças.

6. Agradecimentos

Agradecemos o apoio financeiro do CNPq (processo 578627/2008-6) e o fornecimento das amostras de foliares pela empresa FISHER.

7. Referência bibliográfica

[1] PORTAL REDE DE INOVAÇÃO E PROSPECÇÃO TECNOLÓGICA PARA O AGRONEGÓCIO. Disponível em: <http://www.ripa.com.br/index.php?id=1823> Acesso em: 04 Jan. 2011.

[2] BOVÉ, J. M. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. **Journal of Plant Pathology**, Pisa, v. 88, p.7-37, 2006.

[3] FUNDECITRUS. Manual Técnico do Greening 2009. Disponível em:

<http://www.fundecitrus.com.br/ImageBank/PageFlip/pageflip.aspx?idPage=143>. Acesso em: 29 abr. 2010.

[4] INNIS, M. A.; GELFAND, D. H.; SNINSKY, J. J.; WHITE, T. J. (eds.), PCR Protocols: A guide to methods and applications, Academic Press: San Diego, 1990.

[5] BASSET, O.; BUQUET, B.; ABOUELKARAM, S.; DELACHARTRE, P.; CULIOLI, J. Aplication of texture image analysis for the classification of bovine meat**. Food Chemistry**, v. 69, n. 4, p. 437 - 445, 2000.

[6] GONZALEZ, R.; WOODS, R. **Processamento Digital de Imagens**, Pearson Education do Brasil, 2010.

[7] YONEKAWA, S.; SAKAI, N.; KITANI, O. Identification of idealized leaf type using simple dimensionless shape factors by image analysis. **Transactions of the ASAE**. v. 39, n.4, p. 1525-1533, 1996.

[8] RUSS, J. C. **The Image Processing Handbook**. 3ed. CRC Press LLC, 1998.

[9] INSTITUTO DE ECONOMIA AGRÍCOLA. Divulgação da safra paulista de laranja 2009/10. Disponível em: <http://www.iea.sp.gov.br/out/LerTexto.php?codTexto=12002>. Acesso em: 28 Dez 2010.

[10] AGRISOFT (2009). A citricultura em riso – o pesadelo da Greening. Disponível em: <http://www.agrosoft.org.br/agropag/103543.htm>. Acesso em: 11 Jan. 2011.

[11] NUTTER, JR., F.W.; SCHULTZ, P.M. Improving the accuracy and precision of disease assessments: selection of methods and useof computer-aided training programs. **Canadian Journal of Plant Pathology**, Ottawa, v.17, n.1, p.174-184, 1995.

[12] MARTINS, M.C.; GUERZONI, R.A.; CÂMARA, G.M.S., MATTIAZZI, P., LOURENÇO, S.A. & AMORIM, L. Escala diagramática para a quantificação o complexo de doenças foliares de final de ciclo em soja. **Fitopatologia Brasileira** vol.29, p.179-184, 2004.

[13] KOWATA, L. S.; MAY-DE-MIO, L. L.; DALLA-PRIA, M.; SANTOS H. A. A. Escala Diagramática para avaliar severidade de míldio na soja. **Scientia Agraria,** Curitiba, v.9, n.1, p.105-110, 2008.

[14] WEKA. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/ >. Acesso em: 29 abr. 2010.

[15] WITTEN, I. H.; FRANK, E. Data mining: practical machine learning tools and techniques. 2 ed. San Francisco, 2005.

[16] SIMÕES, A. S. Segmentação de imagens por classificação de cores: uma abordagem neural. DISSERTAÇÃO DE MESTRADO. ESCOLA POLITÉCNICA DA USP. São Paulo. 2000.

[17] SIMÕES, A. S.; REALI COSTA, A. H. Using Neural Color Classification in Robotic Soccer Domain. NTERNATIONAL JOINT CONFERENCE IBERAMIA'2000 and SBIA'2000, Workshop Proceedings, Meeting on Multi-Agent Collaborative and Adversarial Perception, Planning, Execution, and Learning. L. N. Barros, R. M. Cesar Jr., F. G. Cozman, A. H. Reali Costa (eds.). São Paulo, ISBN 85-901664-1-4, November 2.000, Atibaia, SP, Brazil, p.208-213.

[18] SIMÕES, A. S.; REALI COSTA, A. H.; ANDRADE M. T. C. Utilizando um classificador fuzzy para seleção visual de laranjas. **Workshop de Computação - WORKCOMP´2001**. Carlos H. C. Ribeiro and Milton T. S. Sakude. p. 113-117. ITA, 2001.

[19] CAVANI, F. A.; TRONCO, M. L.; SOUSA, R. V. de; PORTO, A. J. V. Segmentação e Classificação de Imagens de laranjeiras utilizando JSEG e Perceptron Multicamadas. **Revista Minerva**, v. 3, p. 189-197, 2006.

[20] IMAGE PRO PLUS. Disponível em: <http://www.mediacy.com/>.Acessado em: 18 Fev. 2011.

[21] THEODORIDIS, S**. Pattern Recognition**. Academic Press, USA, 1st edition, 1999.

[22] SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. Redes Neurais Artificiais para engenharia o e ciências aplicadas curso pratico. Ed. Artliber. 2010.

[23] A INSTRUÇÃO NORMATIVA N°53 E A NECESSIDADE DE UM CONTROLE EFETIVO DO GREENING NO BRASIL, 2009, Disponivel em <http://www.fundecitrus.com.br/ImageBank/FCKEditor/file/pdf/artigo_controle_greening.pdf>. Acessado em : 18 Fev. 2011.

Uma abordagem para verificação facial em dispositivos móveis

Tiago de Freitas Pereira
tpereira@cpqd.com.br

Marcus de Assis Angeloni
massis@cpqd.com.br

CPqD - Centro de Pesquisa e Desenvolvimento em Telecomunicações
Rodovia Campinas - Mogi-Mirim (SP-340) - km 118,5
Campinas, SP

Abstract

In a multiservice world where mobile adoption increases in a fast pace, a new way to provide easy, secure and fast access is a challenge. With the expansion of the number of mobile users, along with the increasing number of different services accessed via mobile devices, there are also increasing obstacles related to the use of traditional passwords for authentication. These are usually related with the choice of a strong password with no change among different services (once the password is compromised, all the services will be subject to unauthorized access) or with the choice of a weak, easy to remember password (subject to brute-force or password guessing attacks). In this paper we present a texture based approach for face verification in mobile devices.

1. Introdução

Nos últimos anos, observou-se um expressivo crescimento nas vendas de dispositivos móveis, e o consequente aumento no emprego de tais aparelhos para acesso a transações bancárias, correio eletrônico, redes sociais e documentos corporativos. Um dos dispositivos móveis que tem alavancado as vendas é o *smartphone*, que é um telefone celular dotado de funcionalidades e acessórios extras, que possui um hardware superior aos telefones celulares convencionais, maior número de conexões disponíveis e um sistema operacional sobre o qual é possível criar novas aplicações. Segundo o estudo "*IDC Latin America Predictions 2011*" apresentado no início de 2011, a venda de aparelhos *smartphones* deve superar a venda de computadores pessoais no Brasil em 2011, com um crescimento de mercado, segundo a IDC, de 67,1% [13]. Um outro estudo, realizado pela consultoria Gartner, aponta que as vendas de smartphones no mundo em 2010 cresceram 72,1% em relação a 2009 [5].

Acompanhando este aumento no número de aparelhos, cresceram também o número de aplicações e serviços destinados a esta plataforma móvel. Com a popularização deste dispositivo, ele começa a ser mais visado por criminosos e fraudadores e, por apresentar dimensões e peso reduzido, o risco de perda e furto do aparelho é grande, podendo resultar em prejuízo financeiro e violação de dados pessoais e corporativos. Diante desse cenário, um problema já conhecido de aplicações para computadores começa a se aplicar também aos *smartphones*, que é o emprego de senhas e sua vulnerabilidade intrínseca, associada ao fato de lidar com dois princípios antagônicos: elas devem ser fáceis de lembrar e difíceis de adivinhar. É comum o usuário ter uma senha forte e utilizá-la em diversos serviços (neste caso, uma vez que a senha está comprometida, todos os serviços estarão sujeitos a acessos não autorizados) ou ter diversas senhas fracas (possibilitando a memorização), que são mais fáceis de serem quebradas. Outro problema é que por se tratar de um celular, é comum um descuido maior por parte de seus usuários do que manipulando um computador, por vezes acessando serviços bancários ou informações pessoais na presença de outras pessoas, que podem acompanhar seus movimentos e copiar sua senha.

Uma alternativa promissora para acesso seguro e que poupa o usuário de ter de memorizar várias e complexas senhas é através do emprego de características biométricas na autenticação do usuário. A biometria vem sendo usada no campo forense já faz algumas décadas, e nos últimos anos tem ganho espaço também em aplicações civis e comerciais, como no controle de acesso, ensino a distância, controle de ponto, presença em aulas para obtenção da carteira de habilitação, obtenção de passaporte e controle de fronteira [6]. No entanto, pouco foi desenvolvido para dispositivos móveis, em parte pelo hardware limitado que até pouco tempo atrás era um grande impeditivo para adaptação dessa e de outras aplicações. Com o recente aumento da capacidade computacional dos *smartphones* e dos acessórios acoplados a ele, como é o caso das câmeras digitais, torna-se conveniente a implementação

de métodos biométricos para estes dispositivos. A face do usuário é uma característica interessante de ser utilizada neste tipo de aparelho, visto que todos os *smartphones* são dotados de câmera, e os mais modernos inclusive possuem câmera frontal para uso em chamadas com vídeo.

2. Biometria

Qualquer característica humana física ou comportamental (por exemplo, face, íris, impressão digital, voz e geometria da mão) pode ser usada como característica biométrica desde que satisfaça alguns requisitos, como universalidade (toda pessoa deve possuir a característica), unicidade (deve permitir distinguir as pessoas), permanência (não deve se alterar demasiadamente ao longo do tempo) e coletabilidade (deve poder ser medida quantitativamente) [6].

Um sistema biométrico é um sistema de reconhecimento de padrões que coleta os dados biométricos do usuário, extrai um conjunto de características discriminativas destes dados e efetua a comparação com um modelo previamente gravado no banco de dados. Ele pode operar de duas formas: como identificação, onde as características extraídas são comparadas com todos os modelos cadastrados no banco de dados, para identificar quem é o usuário (comparação 1:N, onde N é o número de usuários cadastrados); e como verificação, onde as características são comparadas com o modelo referente a identidade declarada pelo usuário, verificando se ele é quem diz ser (comparação 1:1).

Atualmente os sistemas biométricos vêm sendo empregados na área forense, civil e comercial, como no controle de acesso, *login* na rede de computadores, ensino a distância, controle de ponto, presença em aulas para obtenção da carteira de habilitação, obtenção de passaporte, controle de fronteira e investigação criminal [6].

2.1. Reconhecimento de face

O reconhecimento de faces é uma tarefa que as pessoas executam rotineiramente e sem esforço em suas relações diárias. O uso da face tem várias vantagens sobre outras tecnologias biométricas: ela é natural, não intrusiva, não exige equipamentos sofisticados para seu funcionamento, e é de fácil uso [7].

Embora a pesquisa em reconhecimento automático de face tenha iniciado em 1960, ela ainda apresenta muitos problemas sem solução [7]. O correto rastreamento e detecção da face, a exposição da face a diferentes tipos de iluminação na imagem, as poses e rotações da cabeça são alguns dos desafios a serem vencidos.

No entanto, nos últimos anos observou-se avanços significativos na área, novas abordagens propostas, e uma grande atenção ao assunto, sobretudo, devido as demandas de segurança, como em câmeras de vigilância em locais de grande aglomeração de pessoas [16]. Parte dessa evolução se deve ao amadurecimento das técnicas de análise e processamento de imagens e da modelagem da face, e também aos avanços na tecnologia das câmeras digitais e dispositivos móveis.

As abordagens de verificação de faces podem ser divididas em duas categorias: as abordagens discriminativas e generativas [11]. Nas técnicas discriminativas, onde se enquadram as abordagens baseadas em SVM (*Support Vector Machines*) e *boosting*, o resultado do algoritmo é uma decisão binária, enquanto que nas técnicas generativas como as baseadas em GMM (*Gaussian Mixture Model*) e HMM (*Hidden Markov Model*), um modelo estatístico é fornecido como saída e a decisão binária é obtida através de um limiar de aceitação gerado pelo sistema.

Um sistema de reconhecimento de face envolve algumas etapas, tais como: a captura da imagem ou de quadros de um vídeo, a detecção da face e a separação da mesma do resto da imagem, um pré-processamento visando reduzir o ruído introduzido pela iluminação e outros fatores do ambiente, a extração de características da imagem da face tratada, e por fim, a criação do modelo do usuário para gravação no banco de dados (durante o cadastro), ou a comparação das características extraídas com os modelos previamente gravados no banco de dados (durante a verificação ou identificação).

3. Sistema proposto

Este trabalho pertence a um projeto maior envolvendo pesquisa e desenvolvimento em biometria, e como resultado está sendo construída uma API (*Application Programming Interface*) para uso em *smartphones*. Uma das frentes a serem trabalhadas é o desenvolvimento da tecnologia de reconhecimento facial.

O modelo de API que estamos desenvolvendo é inspirado na especificação da BioAPI 2.0 [1]. A BioAPI é um conjunto de padrões internacionais para aplicações que usam tecnologias e dispositivos biométricos, estabelecido pelo BioAPI Consortium. Este consórcio foi criado em 1998, e atualmente reúne mais de 120 empresas ao redor do mundo, além dos organismos padronizadores dos EUA e a ISO (*International Organization for Standardization*).

A especificação da BioAPI é bastante flexível quanto as etapas executadas no cliente e as executadas no servidor remoto. Dessa forma, as diferentes funções especificadas na BioAPI (captura, processamento, verificação, criação de *template*) possuem diferentes graus de liberdade na implementação, ou seja, parte da implementação pode estar alocada no *smartphone* e parte em um servidor.

A abordagem proposta foi implementada em *smart-*

phones da plataforma Android da Google[1]. O sistema de verificação facial proposto está dividido em três módulos: detecção de face, extração de características e criação de *template*/verificação.

3.1. Detecção da face

O processo de detecção da face consiste em segmentar a região da mesma e destacá-la do fundo. O método de detecção facial utilizado neste trabalho é baseado no algoritmo Viola-Jones [14].

Por ser um algoritmo com reduzidas taxas de falsos positivos e com um tempo de resposta baixo [14], o algoritmo Viola-Jones tem sido muito utilizado recentemente para a tarefa de detecção de faces.

Para esta tarefa é explorado um conjunto de características de Haar e aplicada uma técnica de boosting para classificação de face e não face. As características de Haar exploram o contraste presente naturalmente na face e são representadas por formas geométricas como se pode observar na Figura 1. Posicionado sob uma região da imagem, o valor da característica é calculado somando-se os pixels da parte branca da forma geométrica e subtraindo pelo somatório dos pixels da parte preta da forma geométrica.

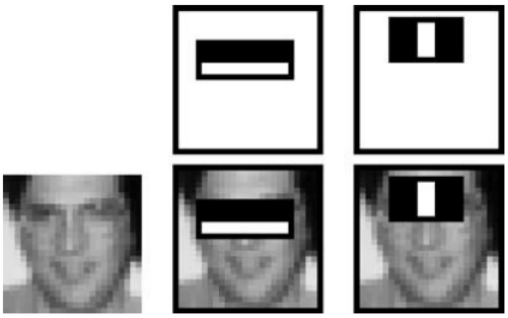


Figura 1. Representação geométrica das características de Haar. Adaptado de [14].

Cada uma das características tem a capacidade de distinguir uma face de uma não face, agindo como um classificador, com uma taxa de erro alta (basta ser abaixo de 50%). Na literatura estes classificadores são chamados de classificadores fracos. Este conjunto de classificadores são combinados na forma de uma árvore degenerada onde cada nó dessa árvore é capaz de rejeitar ou descer mais um nível na árvore (para o próximo classificador), como se pode observar na Figura 2. A técnica de boosting utilizada pelo algoritmo Viola-Jones é o AdaBoost [4].

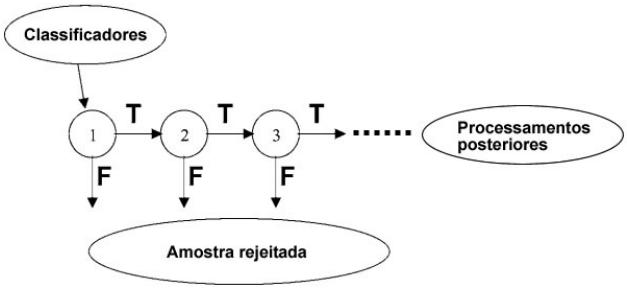


Figura 2. Árvore degenerada do AdaBoost. Adaptado de [14].

3.2. Extração de características

As características utilizadas neste trabalho, para descrição da face, são baseadas no descritor LBP (*Local Binary Pattern*). O LBP é um descritor de textura que foi adaptado para representação de imagens da face, e desde então vem sendo estudado e modificado para este tipo de tarefa, obtendo resultados promissores [2].

O descritor LBP consiste em um histograma de operadores LBP, calculados a partir de uma vizinhança determinada ao redor de um pixel seguindo a Equação 1.

$$LBP(x_c, y_c) = \sum_{n=0}^{N-1} f(i_n - i_c)2^n \qquad (1)$$

onde i_c corresponde ao pixel sendo visitado (x_c, y_c), N é o número de amostras coletadas de sua vizinhança, i_n é a intensidade de um pixel na vizinhança de i_c e $f(x)$ é definido segundo a Equação 2.

$$f = \begin{cases} 0 & se \quad x < 0 \\ 1 & se \quad x \geq 1 \end{cases} \qquad (2)$$

Seguindo a notação de Ojala et al.[9] o número de pontos amostrados e o raio da vizinhança podem ser de formas e tamanhos variados. Resultados satisfatórios tem sido obtidos extraindo 8 pontos na vizinhança de um pixel, utilizando raio 2 e considerando a vizinhança de forma circular como mostrado na Figura 3c. Os pontos amostrados que não coincidem com o centro de um pixel são bilinearmente interpolados [9].

1 http://developer.android.com/sdk/index.html

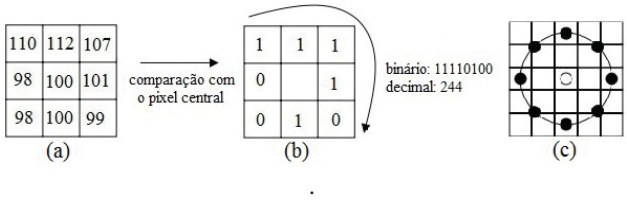


Figura 3. Operador LBP. (a) e (b) ilustram o operador LBP básico, onde os pixels da vizinhança são comparados com cada pixel e apresenta como resultado um código binário. (c) Um exemplo de vizinhança circular (com 8 pontos de amostra e raio 2). Baseado em [2].

Após o cálculo do operador LBP em cada pixel, o histograma é computado. Este histograma de operadores LBP é a descrição da face no espaço de parâmetros.

Para preservação da informação espacial da descrição da face através dos operadores LBP, a imagem facial é dividida em regiões de igual tamanho, e para cada uma delas, um histograma de operadores LBP é computado. Estes histogramas são concatenados a fim de descrever toda a face. Desta maneira a face é descrita em três níveis: o nível de pixel com o cálculo de cada operador, o nível regional com o cálculo de histogramas de cada região e a descrição global com a concatenação dos histogramas [2] como podemos observar na Figura 4.

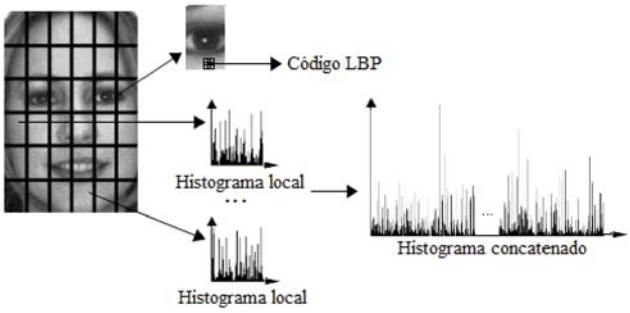


Figura 4. Descrição da face nos três diferentes níveis: nível de pixel (código LBP), nível regional (histograma local) e nível global (histograma concatenado). Imagem de face utilizada foi extraída da base de dados AR Face[2].

Utilizando 8 pontos de amostragem teremos 256 classes de padrões LBP. Uma maneira eficiente de se trabalhar com o LBP é considerando apenas os padrões que possuem até duas trasições de bit 0-1 e vice-versa. Este sub-conjunto de padrões LBP é chamado de padrões uniformes [9]. Além de melhor representar os pontos, arestas e regiões planas, a construção de histogramas usando padrões uniformes representa uma significativa redução da dimensionalidade, já que com estes padrões teremos apenas 58 classes de padrões para um conjunto de amostragem de 8 pontos.

Alguns trabalhos, que utilizam o LBP com os padrões uniformes, para não descartarem os padrões não uniformes, costumam os alocar em uma nova classe de padrões. Experimentos efetuados por [15] apontaram que esta 59ª classe cresce demasiadamente em relação as outras 58 em imagens muito ruidosas ou com má condição de iluminação. Para diminuir este ruído, [15] propôs uma aproximação de cada padrão uniforme pela distância de Hamming. A ideia é reclassificar os padrões não uniformes, os classificando nos padrões uniformes que minimizem a distância de Hamming em relação ao seu valor. Este tipo de refinamento se faz necessário já que estamos tratando deste problema em ambientes móveis.

3.3. Criação do template e verificação

O método generativo implementado neste trabalho é baseado no modelo de misturas gaussianas (GMM), que foi proposto por [10] para verificação de locutor e vem sendo amplamente utilizado para verificação facial por [3, 11, 12].

Nesta estratégia o template do usuário é representado por uma soma de gaussianas ponderadas. Para cada usuário estas gaussianas são inicializadas a partir de um conjunto de dados do usuário alvo e suas médias são posicionadas usando o algoritmo LBG [8]. A partir dessa inicialização, os parâmetros do GMM (pesos, vetores de média e vetores de variâncias) são iterativamente refinados aplicando o algoritmo EM (*Expectation Maximization*)[10].

Uma vez gerado o GMM de um usuário, ele é usado para gerar o *score*, que representa a probabilidade de uma amostra de entrada ter sido gerada por este GMM. Espera-se que este *template* gere notas altas para amostras genuínas e notas baixas para amostras que não são do usuário cadastrado.

Para construção de um sistema de verificação de faces uma grande base de dados é fundamental. Ela é usada para treinamento de um modelo genérico, chamado modelo de mundo, e também é utilizada para calcular a curva de falsa aceitação, empregada para posicionar o limiar de aceitação de um usuário. Amostras usadas para treinamento do modelo de mundo não podem ser usadas para compor o con-

2 http://www2.ece.ohio-state.edu/ aleix/ARdatabase.html

junto de usuários impostores (usados para compor a curva de falsa aceitação).

A função do modelo de mundo neste trabalho é a normalização dos *scores* de verificação gerados pelo *template* do usuário. Para esta tarefa, ele deve ser representativo para o universo de faces. O processo de criação do modelo de mundo é análogo ao processo de criação do *template* de um usuário, com a diferença que o modelo de mundo utiliza amostras LBP de várias faces.

O critério utilizado para aceitação de uma amostra de face para o GMM de um usuário foi inspirado na BioAPI [1]. Nesta especificação o limiar de aceitação é uma percentagem na curva de falsa aceitação, ou seja, se for especificado 0% de falsa aceitação, para uma amostra de face ser aceita seu *score* deve ser superior a todas as amostras da curva de falsa aceitação costruída para tal GMM. É importante ressaltar aqui a importância de uma base de dados volumosa. Se o conjunto de impostores não for representativo, selecionar 0% de falsa aceitação não terá representatividade em termos de taxa de erro do sistema em campo.

4. Resultados

A base de dados de faces utilizada para os testes preliminares, avaliando desempenho e taxa de acerto, foi a da Universidade de Essex (*University of Essex*)[3]. O conjunto *face94*, utilizado para os testes, possui 20 frames de 132 indivíduos. Destes, 113 são homens (compondo o grupo denominado *male*) e 19 são mulheres (compondo o grupo denominado *female*). Os frames possuem dimensões 180x200 pixels e compreendem basicamente toda a região da cabeça e parte da região dos ombros, como podemos observar na Figura 5.

Figura 5. Imagens da base de dados da University of Essex.

Foram realizados três experimentos com esta base de dados. O primeiro computou os *templates* masculinos, com a construção do modelo de mundo e cálculo da curva de falsa aceitação, utilizando apenas amostras masculinas. O segundo computou os *templates* femininos, com a construção do modelo de mundo e cálculo da curva de falsa aceitação, utilizando apenas amostras femininas. E o terceiro computou o modelo de mundo e as curvas de falsa aceitação utilizando amostras de ambos os sexos.

Experimento	EER Médio
Base de dados masculina	0,16%
Base de dados feminina	0,00%
Base de dados mista	0,15%

Tabela 1. Experimentos com a base de dados da Universidade de Essex.

Para todos os experimentos, as imagens das faces detectadas foram redimensionadas para 120x120 pixels, divididas em regiões de 10x10 pixels para cálculo dos histogramas locais, e o ponto de operação do LBP foi: 8 pontos de amostragem na vizinhança e considerando a varredura circular de raio 2.

Para a construção dos *templates* de cada usuário foram utilizadas 15 amostras das 20 disponíveis por indivíduo, e as 5 demais foram usadas para teste de falsa rejeição. Com este conjunto de amostras, para criação do *template*, um GMM com 7 misturas gaussianas foi utilizado. Este número de misturas foi escolhido empiricamente e levando em consideração o número reduzido de amostras, que inviabiliza (matematicamente) o emprego de um número de misturas maior.

O experimento utilizando somente o subconjunto *male* foi feito da seguinte forma. Para constituir o modelo de mundo foram utilizadas amostras dos indivíduos 001 até o 057. Para a construção dos *templates* foram utilizadas amostras dos indivíduos 058 até o 113. Para construção das curvas de falsa aceitação de um indivíduo foram utilizados todos os indivíduos do conjunto de construção dos *templates* (descartando as amostras do usuário em questão). Para a construção do modelo de mundo utilizando o subconjunto *female*, foram utilizadas as amostras 001 até a 009, e para a construção dos *templates* e das curvas de falsa aceitação foram utilizadas as amostras 010 até a 019. Para o experimento descrito, obtivemos o resultado apresentado na Tabela 1.

Para verificar o desempenho computacional deste algoritmo embarcado em um *smartphone*, esta base foi processada pelo aparelho Galaxy S da Samsung, com o sistema operacional Android 2.1. É importante notar que, para este experimento, foi escolhido o pior cenário na arquitetura da BioAPI, ou seja, todo o processamento foi efetuado no celular. Para a construção de cada *template*, o tempo médio foi de 1 minuto e 2 segundos, que é um tempo razoável já que esta tarefa é executada poucas vezes. Para a verificação,

3 http://cswww.essex.ac.uk/mv/allfaces/index.html

o tempo médio do processamento da imagem, geração de *score* e normalização com o modelo de mundo de 5 imagens (as 5 imagens do usuário) foi de 2,230 segundos, o que é razoável também. Com este tempo de verificação pode-se considerar a utilização desta abordagem em uma aplicação em tempo real de verificação de faces embarcada em *smartphones*.

5. Conclusão

Este trabalho explorou uma alternativa para autenticação biométrica através da face, utilizando descritores de textura e embarcado em *smartphones* da plataforma Android. O desenvolvimento desta abordagem foi todo baseado na BioAPI, que é um padrão ISO para construção de bibliotecas de biometria. Os resultados preliminares assinalaram um desempenho computacional promissor na plataforma móvel. Experimentos com a base de dados de face da Universidade de Essex mostraram baixas taxas de erro e um desempenho promissor embarcado em um *smartphone*, mas ainda é necessário construir um experimento mais fiel as condições que serão encontradas em um sistema em campo. Para isso uma base de dados com capturas utilizando *smartphones* explorando diferentes condições de iluminação e de pose é fundamental para calcular o real potencial desta abordagem.

Agradecimentos

A elaboração deste artigo foi feita dentro do Projeto BIOMODAL, realizado com apoio da FINEP e recursos do FUNTTEL.

Referências

[1] *Biometric application programming interface, Part 1: BioAPI specification*. ISO/IEC 19784-1, 2006.

[2] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary pattern. In *Eighth European Conference Computer Vision*, pages 469–481, 2004.

[3] F. Cardinaux, C. Sanderson, and S. Bengio. Face verification using adapted generative models. In *Int. Conf. Automatic Face and Gesture Recognition (AFGR)*, pages 825–830, 2004.

[4] Y. Freund and R. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.

[5] IG Tecnologia. Estudo do Gartner confirma explosão de vendas de smartphones em 2010. Disponível em: <http://tecnologia.ig.com.br/noticia/2011/02/10/estudo+do+gartner+confirma+explosao+de+vendas+de+smartphones+em+2010+10364013.html>er. Acessado em: 18 fev. 2011.

[6] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):4–20, 2004.

[7] A. K. Jain and S. Z. Li. *Handbook of Face Recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[8] Y. Linde, A. Buzo, and R. Gray. An algorithm for vector quantizer design. *IEEE Transactions on communications*, 28(1):84–95, 1980.

[9] T. Ojala, M. Pietikinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[10] D. Reynolds, T. Quatieri, and R. Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 10(1-3):19–41, 2000.

[11] Y. Rodriguez and S. Marcel. Face authentication using adapted local binary pattern histograms. In *In European Conference on Computer Vision (ECCV)*, pages 321–332, 2006.

[12] C. Sanderson and K. K. Paliwal. Fast features for face authentication under illumination direction changes. *Pattern Recogn. Lett.*, pages 2409–2419, October 2003.

[13] UOL Tecnologia. Venda de smartphones deve superar a de computadores no Brasil em 2011, diz IDC. Disponível em: <http://tecnologia.uol.com.br/ultimas-noticias/redacao/2011/02/17/venda-de-smartphones-deve-superar-a-de-computadores-no-brasil-em-2011-diz-idc.jhtm>. Acessado em: 18 fev. 2011.

[14] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, pages 137–154, May 2004.

[15] H. Yang and Y. Wang. A lbp-based face recognition method with hamming distance constraint. In *Proceedings of the Fourth International Conference on Image and Graphics*, ICIG '07, pages 645–649, Washington, DC, USA, 2007. IEEE Computer Society.

[16] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):433–440, 2003.

Autenticação Automática de Assinaturas *Online*: Investigando Alternativas de Representação

Cassia I. G. da Silva[1], Guilherme L. A. Mota[1], Aruquia B. M. Peixoto[2], Alexandre Sztajnberg[1]

[1]Programa de Pós Graduação em Ciências Computacionais
Universidade do Estado do Rio de Janeiro,

[2]Programa de Pós Graduação em Engenharia Mecânica
Pontifícia Universidade Católica do Rio de Janeiro

cassia@ime.uerj.br, guimota@ime.uerj.br, aruquia@gmail.com, alexszt@ime.uerj.br

Abstract

This work proposes a new representation to be used on online signature authenticating systems. Our approach is based on features used in graphoscopy which takes the form of a signature to classify it as genuine or forgery. For each point of the captured signature it calculates the normal vector to compare with the variation of the curvature in the signature's trajectory. The proposed approach is also invariant to scale, rotation and translation. The obtained vector is aligned using Dynamic Time Warping in the classification process. Experiments were performed using the linear classifier in conjunction with the Principal Component Analysis (PCA) to compare the results with those obtained in [3].

1. Introdução

Em nossa sociedade, a assinatura é uma forma legalmente aceita para conferir validade a documentos. A assinatura é um tipo de padrão biométrico, mais especificamente, um padrão comportamental e que vem sendo usado em aplicações forenses, possuindo tanta legitimidade quanto, por exemplo, uma impressão digital.

Como exemplo, em cheques bancários a assinatura é a principal forma de assegurar sua veracidade e garantir sua compensação. O recebimento de cheques com assinaturas falsificadas é uma fonte de prejuízo para bancos e comércios. Segundo [13], em dois anos, uma rede de lojas recebeu 25 mil cheques falsos. Se todos esses cheques tivessem sido aceitos o prejuízo chegaria a R$ 1,8 milhão.

O reconhecimento da autenticidade de assinaturas por computador tem sido um tema de grande interesse, dado seu potencial na redução de prejuízos. Além disto, possibilita maior rapidez e praticidade no processo, quando comparado à autenticação realizada por peritos.

Do ponto de vista de sua automação trata-se de um problema bastante complexo. Dentre as causas, podemos citar que uma pessoa não permanece com o mesmo tipo de assinatura ao longo de sua vida, passando por diversas mudanças como tamanho, características estilísticas ou até mesmo a abreviatura de parte do nome. Esta também sofre influência do meio ou do estado psicológico. Apesar dessas mudanças, alguns detalhes pessoais são preservados, o que torna possível a autenticação, mesmo passado muito tempo.

Por esses motivos, uma assinatura não pode ser considerada falsa apenas por alguns detalhes, nem podemos classificar uma assinatura como verdadeira por apresentar extrema exatidão com todas as características de uma original, podendo esta ter sido feita por uma especialista em falsificações. Características como essas justificam a complexidade do problema conforme mencionado.

Dois métodos são utilizados para a obtenção de dados: o *online* e o *offline*. No método *offline* a assinatura é feita sobre um papel e então digitalizada e transferida para o computador por meio de algum hardware (scanner, máquina fotográfica). Já no método *online*, a assinatura é obtida quando uma pessoa assina diretamente sobre um dispositivo de *hardware*, normalmente um *tablet.* Há diversas vantagens nesta forma de captura, dentre as quais é possível destacar a clareza dos traços, a sequência dos pontos da assinatura, a velocidade com que a pessoa assina (informação que não é possível obter pelo método *offline*) e a não introdução de erros na hora de sua captura.

Trabalhos avaliados mostram que os sistemas atuais não possuem um bom desempenho, onde a média de erros total é de 10%. Em contrapartida, o melhor sistema automático *online* disponível obteve uma taxa de erro igual (EER – *Equal Error Rate*) de 3%. Somando-se estes valores chegamos ao valor de 6%, enquanto o desempenho de um perito é de 0.5% de erro para falsas aceitações (aceitação de assinatura falsificada) e 7% para falsas rejeições (rejeição de assinatura verdadeira).

No presente artigo, investigamos e avaliamos o uso de representações de assinaturas combinadas com algumas

técnicas matemáticas e de visão computacional para a autenticação de assinaturas *online*. Observa-se que o emprego de técnicas diferentes ou a variação de atributos pode alterar o resultado de uma classificação positiva ou negativamente. O objetivo do presente trabalho é aprimorar a qualidade e o desempenho (velocidade, precisão e etc.) das classificações.

Como estratégia de avaliação, foi selecionado o conjunto de técnicas apresentados em [3], vencedor do *First International Signature Verification Competition* (1st SVC). A partir desta base, foram realizadas modificações nas características extraídas das assinaturas e empregadas as técnicas de análise vetorial, analisando o quanto a escolha dessas características pode influenciar no resultado final da classificação. Os resultados obtidos, com a abordagem aqui proposta, foram comparados com aqueles obtidos por [3], indicando que existe uma margem tangível para aprimorar os resultados das classificações utilizando-se de elementos de representação que permitam um melhor tratamento das variações intraclasse na assinatura de um indivíduo.

O restante do artigo está estruturado da seguinte forma: na Seção 2, são analisados alguns dos principais problemas surgidos durante a autenticação de assinaturas. Na Seção 3, é apresentada a metodologia que foi utilizada como base para a presente proposta, enunciando os passos seguidos. E nas seções que seguem, é apresentada a proposta, assim como os resultados obtidos e a comparação com os obtidos em [3].

2. Autenticação de assinaturas

O processo de autenticação de assinaturas se torna um trabalho minucioso, pois muitas considerações devem ser feitas antes da classificação. Justino et al [7] mostra que quando uma pessoa assina em espaços diferentes, varia a escala de sua assinatura de forma não linear. A Figura 1 mostra três assinaturas de um mesmo autor escritas em diferentes escalas, exemplificando a variação da assinatura dependendo do espaço fornecido para a mesma.

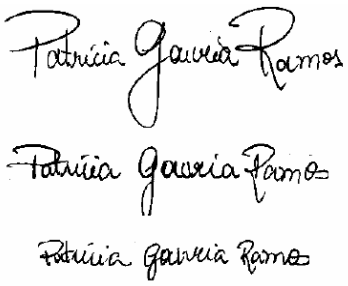


Figura 1: Variações da escala [7]

Cavalcanti et al. [8] observa que uma pessoa muda diversas vezes o tipo de sua assinatura ao longo de sua vida e que uma assinatura sofre alterações dependendo do estado físico e psicológico de seu dono, entre outras. Santos et al [6] ressalta que pode haver uma similaridade interpessoal que se reflete em uma semelhança entre assinaturas de autores distintos (observe Figura 2). Assinar não é um processo perfeitamente preciso com características idênticas, dependendo do estado psicológico e do espaço disponível pode-se perceber variações. E coloca em evidência que a única certeza é que quando duas assinaturas são idênticas, uma delas é provavelmente uma cópia.

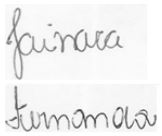


Figura 2: Similaridade interpessoal

Outro detalhe importante é como representar a assinatura de modo que esta seja determinante para a comparação. Segundo [5] há três principais categorias em que os sistemas de verificação de assinaturas podem centrar suas propostas: comparação ponto a ponto, comparação dos valores de atributos e comparação da forma da assinatura.

Comparação ponto a ponto: confere a assinatura teste com a assinatura de referência por diferentes partes, separadamente e combinando essas comparações para conseguir uma medida de similaridade global. Alguns trabalhos sugerem que esta abordagem pode fornecer bons resultados, mas pode falhar devido a variações entre assinaturas de um mesmo indivíduo. Segundo [5], os sistemas que apresentam os melhores resultados, até o momento, têm taxas de erro total de pelo menos 15%.

Comparação dos valores de atributos: neste contexto, deve-se levar em consideração o número e tipos de atributos que serão necessários para a verificação e a escolha apropriada de um meio para fazer a comparação desses atributos. Uma das técnicas, local, é realizar a comparação exaustiva envolvendo cada informação capturada. A outra técnica, global, utiliza alguns poucos atributos que descrevem a assinatura como um todo. Sistemas que utilizam esta técnica são normalmente mais rápidos, no entanto possuem baixo desempenho.

Comparação entre a forma da assinatura: Estes sistemas são baseados na captura da forma da assinatura. Esta técnica é bem diferente da comparação ponto a ponto. Eles testam partes da assinatura comparando-a com as respectivas partes de outra. Estes sofrem de um pobre desempenho quando submetidos à autenticação de falsificações habilidosas (que reproduzem a forma de uma assinatura genuína), mas possuem ótimo desempenho quando identificam as assinaturas como sendo de um usuário ou não.

3. Trabalho de referência

Dentre os trabalhos estudados, foi tomado como base aquele apresentado em [2] e refinado em [3]. O sistema proposto por [2] recebeu o prêmio de primeiro lugar em 1st SVC. Este sistema está dividido em quatro principais etapas: extração de atributos, criação de um perfil, treinamento e verificação.

3.1. Extração de atributos

Segundo [2], para ter um bom desempenho é preciso encontrar características que tenham pequenas variações dentro do conjunto de assinaturas genuínas, mas que sejam determinantes para a distinção de uma assinatura falsificada. Assim, foram utilizados atributos locais dos pontos da trajetória da assinatura: (i) diferença entre as coordenadas x e y de dois pontos consecutivos (que são chamados ao longo do texto de (Δ_x,Δ_y)), (ii) diferença entre o primeiro ponto da assinatura e os demais pontos e (iii) diferença de curvatura entre dois pontos consecutivos.

O atributo da diferença entre o ponto inicial de uma assinatura com os demais pontos é uma característica invariante à rotação, ou seja, se uma assinatura for capturada sem uma linha de apoio, pode haver alterações em relação ao posicionamento da assinatura com a base da folha, causando discrepâncias entre duas assinaturas. Tendo-se como base o primeiro ponto da assinatura, essas discrepâncias tornam-se menores já que a diferença entre um ponto qualquer e o primeiro ponto pode permanecer constante.

Observe que quando é feita a diferença entre coordenadas de dois pontos consecutivos não está sendo feito entre os caracteres da assinatura, visto que os pontos capturados pelo *tablet* em duas assinaturas genuínas não são os mesmos. Outro ponto interessante é que a diferença de curvatura pode ser considerada uma característica invariante à escala. Dentre os testes feitos pelo autor, os melhores resultados foram obtidos utilizando (Δ_x,Δ_y).

Todos os atributos mencionados são calculados para cada ponto extraído pelo *tablet*. Para cada característica utilizada, é obtido um vetor que será usado para efetuar as comparações entre elas. Logo após, é feito o alinhamento entre duas assinaturas para ser efetuada a autenticação.

Observa-se que, devido à diferença existente entre assinaturas genuínas de um autor, cada assinatura possui mais ou menos pontos de amostras (comparada uma a outra) implicando, assim, em vetores de atributos que possuem comprimentos distintos. Assim, utilizar a distância Euclidiana para esta comparação torna-se muito difícil. Por esse motivo, o alinhamento é feito com a técnica *Dynamic Time Warping* (DTW) – Seção 4.2.

3.2. Criação de um perfil

Consiste em tomar algumas assinaturas genuínas, chamadas de conjunto de referência, para que seja extraída alguma similaridade do indivíduo. Neste trabalho e em [2], foram fornecidas oito assinaturas para o conjunto de referência. Para este conjunto são calculados os vetores de atributos e, então, as assinaturas são alinhadas em pares (uma a uma) utilizando o DTW para que sejam encontradas as distâncias entre cada par. Dentre todas essas distâncias três foram consideradas para a classificação:

- Média das distâncias do conjunto de referência aos seus vizinhos mais próximos $\overline{D_{\min}}$;
- Média das distâncias do conjunto de referência aos seus vizinhos mais distantes $\overline{D_{\max}}$;
- Média das distâncias do conjunto de referência a um modelo (*template*) $\overline{D_{Template}}$.

Os valores médios de distância descrevem a variação (perfil) do usuário. Os valores das médias máximas, mínimas e a média de distâncias ao modelo para o seu conjunto de referência $(\overline{D_{\min}}, \overline{D_{\max}}, \overline{D_{Template}})$, foram selecionados de forma a serem utilizados na normalização de uma assinatura.

3.3. Treinamento

Para o treinamento em [2] foram utilizadas assinaturas genuínas e falsificadas (diferentes das assinaturas utilizadas na criação do perfil). Estas são inseridas no sistema para que seja determinado um limiar que faça a separação das assinaturas autenticas das falsificadas.

Cada uma é comparada com as assinaturas do conjunto de referência, obtendo assim três valores de distância $\left(D_{\min}, D_{\max}, D_{Template}\right)$. Esses valores são normalizados por suas médias correspondentes do conjunto de referência $(\overline{D_{\min}}, \overline{D_{\max}}, \overline{D_{Template}})$ resultando em um vetor tridimensional com as características utilizadas para a classificação - Equação (1).

$$Fy = \left[\frac{D_{\min}}{\overline{D_{\min}}} \quad \frac{D_{\max}}{\overline{D_{\max}}} \quad \frac{D_{Template}}{\overline{D_{Template}}} \right] \tag{1}$$

Finalmente, os classificadores são treinados para separar as assinaturas genuínas das falsificadas.

3.4. Verificação e resultados

Cada assinatura teste é comparada com todas as assinaturas do conjunto de referência, sendo obtido um

vetor tridimensional (contendo as informações do vizinho mais próximo, vizinho mais distante e distância ao *template*), normalizado pelas suas médias correspondentes obtidas na criação do perfil. Então, um classificador é usado para determinar se é uma assinatura genuína ou não.

Dois classificadores foram usados em [2] e [3]: Bayes e o classificador linear. Os melhores resultados foram obtidos empregando o classificador linear, após a redução da dimensionalidade de três dimensões para uma, com a técnica PCA, *Principal Components Analysis*. Os resultados observados foram adquiridos utilizando o atributo da diferença entre dois pontos consecutivos (Δ_x, Δ_y). Tais resultados são apresentados na Seção 5.

4. Representação alternativa da assinatura

A ideia central da representação da assinatura em na presente proposta é a de considerar a forma da assinatura capturada e tratá-la de maneira semelhante ao do sistema de reconhecimento de assinaturas por um perito. A forma da assinatura é uma propriedade que, para a grafoscopia, é indispensável para a autenticação de uma assinatura, dado que cada pessoa possui uma forma única e pessoal de assinar.

A característica, utilizada nesta investigação, é a variação da curvatura de uma assinatura, onde a ordem dos pontos ao longo da mesma é fundamental. Observa-se que a captura de uma assinatura com a técnica online preserva esta informação, o que não acontece com uma assinatura digitalizada representada por matriz de pontos.

É importante ressaltar que a variação da curvatura, entre assinaturas de um mesmo autor, não é grande. Não estão sendo considerados fatores externos que possam influenciar a escrita de um indivíduo, citados anteriormente.

Os processos de autenticação de assinaturas que fazem comparações entre as curvaturas têm um bom desempenho desde que as falsificações testadas sejam falsificações simples ou aleatórias, já que as falsificações habilidosas copiam de forma idêntica o formato da assinatura em questão. Para sanar esse problema, outro atributo pode ser levado em consideração. Um dos mais sugeridos é o tempo de duração da assinatura, devido ao fato de que essa informação dificilmente pode ser copiada.

4.1. Extração de atributos

O atributo escolhido para a avaliação deste trabalho é a variação da curvatura, considerando cada ponto adquirido pelo *tablet*, medido através dos vetores normais. É importante ressaltar que os vetores normais indicam a variação da curvatura ao longo de uma curva.

Essa variação é obtida com seguinte processo: dados três pontos consecutivos na trajetória da assinatura, capturados pelo *tablet*, são calculados os vetores, que darão o sentido de orientação da escrita. Isto é, no ponto P_i é calculado o vetor $\overrightarrow{P_{i-1}P_{i+1}}$, como mostra a Figura 3.

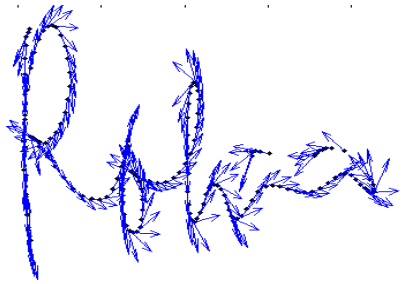


Figura 3: Vetores sobre os pontos da assinatura.

Somente os pontos capturados pelo *tablet* são mostrados nas figuras, e não a linha que os liga, formando a assinatura em si. Logo após, são calculados os vetores normais (em vermelho) a cada um desses vetores.

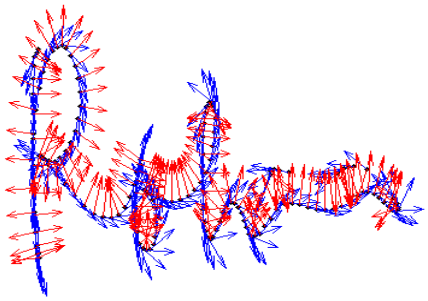


Figura 4: Vetores normais (em vermelho).

O resultado obtido é uma matriz de ordem *n*x2, onde n é numero de pontos capturados pelo *tablet* e as colunas correspondem às componentes (x, y) dos vetores encontrados. Esse atributo é invariante à escala, rotação e translação, já que a curvatura de uma assinatura pode não variar tanto tendo um ponto de vista global.

4.2 *Dynamic Time Warping* (DTW)

Também utilizado em [2] e [3] para alinhar duas assinaturas, o DTW é um algoritmo que mede a similaridade entre duas sequências (não necessariamente de mesmo comprimento) que podem variar de acordo com o tempo ou velocidade.

O DTW permite encontrar o melhor alinhamento não-linear entre vetores tal que a distância total entre os elementos do vetor correspondente seja minimizada. Esses elementos são 'deformados' de forma não-linear na dimensão do tempo para determinar uma medida de sua semelhança independente de certas variações. Aqui, em

[2] e [3], a distância total entre duas assinaturas S_1 e S_2 é calculada por:

$$C[i,j] = \min \begin{cases} C[i-1,j] + \gamma, \\ C[i,j-1] + \gamma, \\ C[i-1,j-1] + Dist(S_1[i], S_2[j]), \end{cases} \quad (2)$$

onde $S_k[i]$ representa o i-ésimo ponto da k-ésima assinatura, e

$$Dist(S_1, S_2) = \begin{cases} 0, & \text{se} \quad \|S_1 - S_2\| < Thr, \\ \|S_1 - S_2\| - Thr, & \text{caso contrário.} \end{cases} \quad (3)$$

C é a matriz a ser preenchida pelo algoritmo e γ é a constante de penalidade em uma das assinaturas. A função *Dist* é projetada para permitir a variação insignificante em assinaturas de referência usando o limite constante *Thr*.

Deste modo, após serem extraídas as características da assinatura, duas assinaturas são alinhadas utilizando o algoritmo DTW. O resultado da aplicação do algoritmo DTW nos fornece a distância entre as duas assinaturas que são introduzidas no classificador, como a proposta usada como base (Seção 3).

Todos os experimentos apresentados neste artigo foram feitos com a base de assinaturas adquirida em [4]. Esta base é composta por 94 usuários que cederam 20 exemplares de suas assinaturas, e os mesmos foram convidados a falsificar as assinaturas de outros usuários, assim obtendo 10 falsificações para cada usuário, dentre estas, 5 classificadas como habilidosas e 5 como altamente habilidosas. Os testes foram implementados com o software MATLAB® 7.8.0, versão 2009.

O experimento consistiu em ler a base de assinaturas, extrair e alinhar os atributos, realizar o treinamento do sistema e em seguida a classificação. Duas baterias foram realizadas: uma utilizando-se apenas as técnicas originais de [2] e outra utilizando a abordagem proposta. Como não possuímos as informações sobre os valores dos parâmetros utilizados por [2] para as equações (2) e (3), empregamos valores empíricos. Os resultados obtidos não foram idênticos aos originais, mesmo para o primeiro caso, mas bem aproximados. Com os resultados de cada bateria foi gerada a curva ROC, que indica falsos positivos e falsos negativos.

5. Avaliação e resultados

Nesta proposta, foram testados dois classificadores: o classificador linear em conjunto com o PCA e o classificador Discriminante de Fisher. No gráfico da Figura 5 estão representadas as curvas ROC para o classificador linear. Os melhores resultados foram obtidos variando-se os parâmetros do DTW para cada par de assinaturas a partir da média da Equação (3). Como podem ser observados no gráfico, estes foram de 4% de EER contra 3,75% obtidos pela abordagem em [3].

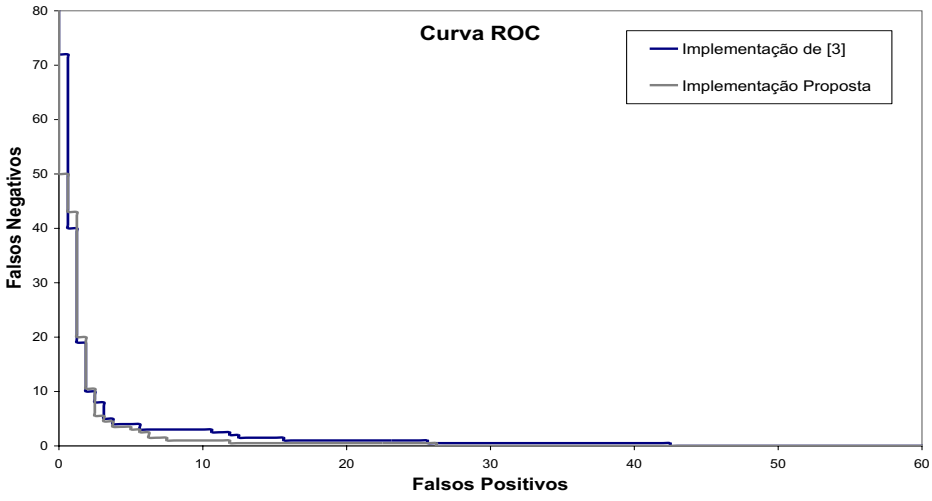


Figura 5: Resultados obtidos usando apenas as técnicas em [3] e com a mudança de atributos proposta

Uma outra característica citada pelo autor (diferença do primeiro ponto para os demais pontos da assinatura) foi testada obtendo uma média de erro de 5.5%.

Fixando o valor dos parâmetros para todos os usuários, foi observado que as distâncias obtidas com o alinhamento do DTW podem alterar significativamente o

resultado da classificação, assim como a taxa de erro. Assim, foi obtida uma média de EER de 8.8% para [3] e 9,6% para a representação sugerida.

6. Conclusão

Neste trabalho, avaliamos e comparamos os resultados obtidos com o uso de uma nova representação da forma de uma assinatura, com os resultados obtidos em [2] e [3], fazendo a mudança nas características utilizadas para a representação de assinaturas *online*. Para tal, escolhemos as características empregadas pela grafoscopia, onde se compara a variação da curvatura de uma assinatura, neste caso, por meio dos vetores normais.

A maior contribuição desse trabalho é propor uma representação da assinatura que capture as variações intraclasse e que, ainda, é invariante à escala, rotação e translação.

A representação sugerida para este trabalho mostra-se promissora, pois com poucas adaptações esta foi introduzida no sistema e mostrou um bom desempenho quando comparado com outras propostas desta área. Estes resultados são animadores, pois as falsificações utilizadas durante a classificação são tidas como habilidosas e altamente habilidosas, diferente de outros sistemas que só obtêm bons resultados com falsificações simples e aleatórias. Além disso, o segundo lugar do SVC obteve um erro (EER) de 4,4%.

Como trabalhos futuros, serão testados outros classificadores e as rotinas de treinamento continuarão sendo aprimoradas com o objetivo de melhorar os resultados.

Agradecimentos. Os autores gostariam de agradecer o apoio parcial da CAPES e Faperj.

7. Referências

[1] A. S. Britto Jr, C. O. A. Freitas, E. Justino, D. L. Borges, J. Facon, F. Bortolozzi and R. Sabourin,: "Técnicas em Processamento e Análise de Documentos Manuscritos". *Revista de Informática Teórica e Aplicada*. Porto Alegre - UFRGS, Vol.8, N°. 2, pp. 47 - 68, 2001.

[2] A. Kholmatov, and B. Yanikoglu, "Biometric authentication using online signatures". In: *Proc. International Symposium on Computer and Information Sciences (ISCIS)*, Springer LNCS-3280 373-380, 2004.

[3] A. Kholmatov, and B. Yanikoglu, "SUSIG: An On-Line Signature Database, Associated Protocols And Benchmark Results". *Pattern Analysis and Applications*, 25 Abril de 2008.

[4] A. Kholmatov, and B. Yanikoglu,, "Identity authentication using improved online signature verification method". *Pattern Recognition Letters,* vol. 26, pp. 2400-2408, 2005. Disponível em: http://biometrics.sabanciuniv.edu/eng/SUSig.html, 2010

[5] A. McCabe, J. Trevathan, "Implementation and analysis of a handwritten signature verification technique". ISBN 978-989-8111-12-8. *Proceedings of the Second International Conference on Security and Cryptography In: Secrypt 2007 International Conference on Security and Cryptography*, Barcelona, Espanha, Julho, 2007.

[6] C. R. Santos, "Análise de assinaturas manuscritas baseada nos princípios da grafoscopia". Dissertação de Mestrado em Informática Aplicada. Publicada em 2004, Puc-PR, 2004.

[7] E. J. R. Justino, "O Grafismo e os Modelos Escondidos de Markov na Verificação Automática de Assinaturas". Tese de Doutorado, Pontifícia Universidade Católica do Paraná, Brasil, 2001.

[8] G. D. C. Cavalcanti, R. C. Dória, P. J. L. Adeodato and E. C. B. Carvalho Filho: "Verificação Off-line de Assinaturas de Tamanhos Diferentes". *VII Brazilian Symposium on Neural Network,* 2002, Porto de Galinhas, Recife, PE. (SBRN 2002), p. 30-35, 2002.

[9] M. Hanmandlu,, M. H. M. Yusof, and V. K. Madasu, "Off-line Signature Verification and forgery detention using fuzzy modeling". *Pattern Recoginition* 38[3] 341-356, 2005.

[10] M. Heinen, F. Osório, "Aplicação de Redes Neurais Artificiais no Reconhecimento On-line de Assinaturas". *IV Congresso Brasileiro de Computação CBComp,* 2004, Itajaí, SC, Brazil. Anais do Concurso de Monografias, Dissertações e Teses em Computação e áreas afins. Itajaí, SC, Brazil: 2004. Vol. 1. pp. 1-92, 2004.

[11] M. Heinen, F. Osório, "Autenticação de Assinaturas Utilizando Análise de Componentes Principais e Redes Neurais Artificiais". *1st Workshop on Computational Intelligence (WCI 2006), 2006*, Ribeirão Preto, SP, Brazil. Proceedings of the International Joint Conference IBERAMIA/SBIA/SBRN 2006. Ribeirão Preto, SP, Brazil: ICMC-USP. p. 1-6, 2006.

[12] R. C. Dória, E. C. B. C. Filho, and E. F. A. Silva, "How distortions in different size signatures influence moment based techniques". *Conference of Pattern Recognition and Image Processing (PRIP)*, 1:219–226, 2001.

Reconstrução Digital de Documentos Mutilados usando Programação Dinâmica e Algoritmo de Prim Modificado

Andre M. Pimenta, Edson J. R. Justino, Luiz E. S. Oliveira
Pontifícia Universidade Católica do Paraná, Curitiba, PR, Brasil
Universidade Federal do Paraná, Curitiba, PR, Brasil

andrepimenta@yahoo.com
edson.justino@pucpr.br
lesoliveira@inf.ufpr.br

Abstract

In this work we propose a methodology for document reconstruction based on dynamic programming and a modified version of the Prim's algorithm. Firstly, we use polygonal approximation to reduce the complexity of the boundaries and extract features from them. Thereafter, these features are used to feed the LCS dynamic programming algorithm. The scores yielded by the LCS algorithm are then used into a modified Prim's algorithm to find the best match among all pieces. Comprehensive protocol and a database composed of 100 shredded documents support the efficiency of the proposed methodology. When compared to global search algorithms, this approach brings an improvement of 18% in the number of fragments reconstructed.

1. Introdução

Reconstrução de documento mutilado é uma tarefa laboriosa executada por especialista de diferentes áreas de atuação, tais como forense, restauração de acervos bibliográficos, entre outros. O tempo despendido nessa tarefa depende do tamanho e número de fragmentos do documento e pode ser mensurado em dias, semanas ou mesmo meses, dependendo dos cuidados que devam ser tomados no trato do mesmo. Documentos antigos ou raros, por exemplo, devem ser manuseados com extremo cuidado para evitar outros danos decorrentes da fragilidade do material que o compõe. A fim de reduzir ao mínimo possível a manipulação dos fragmentos e atenuar o esforço no manuseio, tem sido propostos métodos computacionais para diminuir a complexidade da remontagem, utilizando imagens digitais. Os métodos propostos, em sua maioria, foram desenvolvidos para solucionar problemas como os quebra-cabeças [1]. Em geral, eles são baseados em formas específicas e características como a cor, bem como as relações que podem existir entre várias peças do mesmo.

Um algoritmo eficiente para a resolução da montagem de quebra-cabeças foi proposto por Wolfson em [2]. Nesse trabalho, o autor apresenta um algoritmo para o casamento de duas curvas onde os limites são representados por sequências de atributos de forma, que são obtidos por aproximação poligonal. No entanto, o algoritmo falha quando o número de peças do quebra-cabeça se torna elevado. Outra estratégia interessante é proposta por Kong e Kimia [3]. Eles redefinem as bordas dos fragmentos através de uma técnica de aproximação poligonal para reduzir a complexidade dos contornos dos fragmentos. A programação dinâmica é usada na sequência para alinhar os fragmentos.

Estas técnicas foram usadas em outros domínios de aplicação, tais como arqueologia e restauração de arte. Nestes casos, o objetivo é reconstruir objetos bidimensionais ou documentos que foram danificados ou quebrados em um grande número de fragmentos irregulares. Willis e Cooper associam [4] o problema à reconstrução de artefatos em abordagem 2D e 3D. Leitão e Stolfi [5] propõem um algoritmo baseado na aproximação por curvas, na restauração de painéis cerâmicos. Resultados interessantes também foram relatados por Papaodysseus et al [6], onde o foco é a reconstrução de pinturas em afrescos em sítios arqueológicos.

Com relação à reconstrução de documentos para fins forenses, algumas obras podem ser encontradas na literatura. Num trabalho mais recente, De Smet [7] apresenta uma análise formal do problema de reconstrução de documentos rasgados, quando os fragmentos podem ser recuperados como uma pilha ordenada de pedaços. Justino et al [8] propõe uma reconstrução local dos documentos mutilados, baseada na aproximação poligonal e comparação de características, Chan et al [11] apresta uma proposta para reconstrução de documentos recortados por máquina de picotar papel, cujos fragmentos apresentam formas regulares.

O presente trabalho apresenta um método para reconstrução digital de documentos mutilados com fragmentos em formato irregulares, baseado na programação dinâmica e uma versão modificada do algoritmo de Prim [9].

Inicialmente aplica-se a aproximação poligonal para reduzir a complexidade dos contornos dos fragmentos e superar os problemas específicos enfrentados na reconstrução do documento. Um dos exemplos desses problemas encontra-se nas bordas dos fragmentos dos documentos rasgados a mão. Esses fragmentos apresentam bordas duplas decorrentes da permanência de parte do material num fragmento e em outro, veja Figura 1. Cabe lembrar que artigos relacionadas com remontagem de quebra-cabeças exploram o fato de as peças do mesmo apresentarem bordas ou contornos lisos e curvas bem definidas. No entanto, pedaços de papel picado a mão não seguem esse padrão.

(a) (b)

Figura 1. (a) Exemplo de fragmento com bordas duplas. (b) Exemplo de fragmento com borda simples.

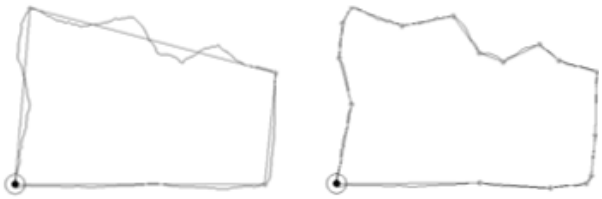
Depois da aproximação poligonal, é possível extrair características relevantes de cada vértice, que definem o contorno do documento e, em seguida, realizar a comparação com o algoritmo LCS (*(Longest Common Subsequence)* da programação dinâmica. Os escores gerados pelo algoritmo LCS são utilizados em um algoritmo de Prim modificado. O objetivo é encontrar a melhor correspondência entre todos os fragmentos. Foi demonstrado, em experimentos abrangentes, que o procedimento proposto produz resultados promissores para o problema da reconstrução do documento. Em relação ao trabalho anteriormente apresentado [8], o presente trabalho aborda alguns pontos-chave que permitem melhorar a taxa de reconstrução em cerca de 18%.

2. Método Proposto

O método proposto é formado por quatro etapas principais: pré-processamento, extração de características, comparação de características e reconstrução. Inicialmente, cada imagem do fragmento do documento é pré-processado através do uso da aproximação poligonal com o objetivo de reduzir a complexidade das bordas dos mesmos. A partir desse processo, podem-se extrair características relevantes de cada vértice do polígono e, em seguida, realizar a comparação com o algoritmo LCS da programação dinâmica. Os escores gerados pelo algoritmo LCS são utilizados em um algoritmo de Prim modificado a fim de encontrar a melhor correspondência entre todos os fragmentos. Nas seções seguintes são apresentadas, em detalhes, as partes que compõe o método proposto.

2.1. Pré-processamento

Algoritmos tradicionais para montagem de quebra-cabeças geralmente levam em conta as bordas lisas e cantos bem definidos das peças. No entanto, no tocante aos documentos mutilados o processo se torna mais complexo. O ato de rasgar um pedaço de papel à mão, muitas vezes produz algumas irregularidades nas bordas dos fragmentos, o que dificulta a obtençao de um encaixe perfeito dos fragmentos, veja Figura 1. Para superar este tipo de problema, foram experimentados algoritmos para a redução da complexidade do contorno, tais como os de aproximação poligonal, e neste caso, aquele que apresentou os melhores resultados foi o bem conhecido algoritmo de Douglas-Peucker (DP) [9]. Esse algoritmo é usado extensivamente em computação gráfica e sistemas de informação geográfica. A Figura 2 mostra um exemplo deste processo, em diferentes níveis de aproximação. Para mais detalhes sobre o pré-processamento, veja [8].



(a) (b)

Figura 2. Aproximação poligonal em diferentes níveis de detalhamento.

2.2. Extração de Características

Após a redução da complexidade através da aproximação poligonal, o próximo passo consiste na extração de características que devem ser utilizadas durante o processo de correspondência entre fragmentos. A extração de características pode ser visto também como um processo de redução da complexidade, já que converte o polígono em uma sequência de vértices com seus atributos. Neste trabalho, é proposto um conjunto de recursos simples que podem ser utilizados para efetuar a correspondência local.

A primeira característica é o ângulo de cada vértice em relação aos seus vizinhos. Considere, por exemplo, o vértice A e B no polígono mostrado na Figura 3. O ângulo é dado por:

$$\cos(\alpha) = \frac{uv}{|u||v|}. \quad (1)$$

Além do valor angular, também é verificado se tal ângulo é convexo ou côncavo. Por exemplo, na Figura 3, o vértice B tem um ângulo convexo enquanto o vértice C tem um ângulo côncavo. Para completar o conjunto de recursos, são calculadas as distâncias, ou arestas *u* e *v*, entre o vértice em análise e seus vizinhos (anterior e posterior em sentido horário). Essas distâncias são obtidas por meio do cálculo da distância Euclidiana. A Tabela 1 descreve os vetores de características extraídos do polígono mostrado na Figura 3. As duas últimas características são as coordenadas do vértice na imagem, serão utilizados posteriormente para o processo de translação e rotação do fragmento, durante a remontagem.

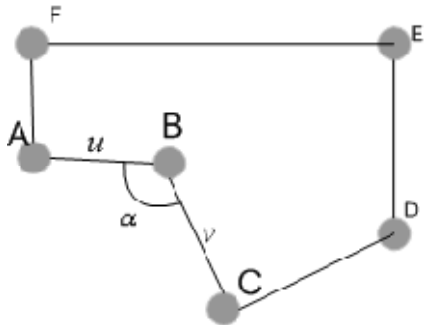


Figura 3. Extração da característica angular do polígono.

Tabela 1. Exemplo do vetor de características.

Vértice	Ângulo	Distâncias		X	Y
		Anterior	Posterior		
A	270	40.0	45.0	10	70
B	120	45.0	43.5	55	67
C	200	43.6	115.7	67	25
D	245	115.7	11.0	180	0
E	270	110.0	170.0	180	110
F	270	170.0	40.0	10	110

Esta tabela pode ser lida como segue: O ângulo do vértice B, que é menor do que 180 graus e portanto côncavo, é calculado usando os vértices A e C, possui o valor de 120 graus. As distâncias Euclidianas entre B e os seus vizinhos A e C são 45,0 e 43,6, respectivamente. As coordenadas do vértice B, na imagem, são 55 e 67.

2.3. Combinação

Nessa etapa, o principal objetivo é calcular o grau de semelhança entre as bordas de cada fragmento do documento. Para executar esta tarefa, usou-se o algoritmo LCS (a mais longa subsequência comum), que é um algoritmo de programação dinâmica destinado a encontrar a maior subsequência comum a todas as sequências em um conjunto. Neste caso, as sequências são formadas pelos nodos ou vertices com seus grupos de características.

O LCS começa com uma matriz E de tamanho (*M* +1) (*N*+1), onde cada célula representa o cunjunto de características do vertice, onde *M* e *N* representam os comprimentos das duas sequências (*X* e *Y*) que estão em análise. A primeira linha e coluna são preenchidas com zeros. Os demais valores são estabelecidos segundo a definição seguinte:

$$E_{ij} = \max \begin{cases} E_{i-1,j-1} + S \\ \max(E_{i,j-1} + P, 0) \\ \max\left(E_{i-1,j} + P, 0\right). \end{cases} \quad (2)$$

Sendo $S = 1$ se $X_i = Y_j$ e 0 caso contrário. *P* é o valor de penalidade. Considere, por exemplo, duas sequências $X = \{0223110133\}$ e $Y = \{0223101133\}$. A correspondente matriz *E* é ilustrada na Figura 4. Apesar do fato de ambas as sequências não coincidirem completamente, é possível recuperar-se a matriz LCS parcial resultante. A Figura 4 mostra a diagonal principal dos resultados parciais. A pontuação é dada pela última célula da combinação. Nesse caso, a maior combinação possui a pontuação 5.

	X	0	2	2	3	1	1	0	1	3	3
Y	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0
2	0	0	2	1	0	0	0	0	0	0	0
2	0	0	1	3	0	0	0	0	0	0	0
3	0	0	0	0	4	0	0	0	0	1	1
1	0	0	0	0	0	5	0	1	1	0	0
0	0	0	0	0	0	1	0	1	2	0	0
1	0	1	0	0	0	0	2	0	0	0	0
1	0	0	0	0	0	1	0	3	1	0	0
3	0	0	0	0	1	0	0	0	0	2	1
3	0	0	0	0	1	0	0	0	0	1	3

Figura 4. Matriz E para as sequências *X* e *Y*.

Nos experimentos realizados, um valor de penalidade elevada (*P* = -100) foi considerado pois a não combinação de qualquer característica de uma célula, implica em invalidar toda a sequência. Por exemplo, considerando que as duas sequências $Z = \{100, 90, 100\}$ e $W = \{100, 150, 100\}$, representando três características usadas para a reconstrução. Se, Z_2 e W_2 representam o ângulo, o

complementar de 90 seria 270, portanto o par (90, 150) não é compatível e não serve à reconstrução. Isto exemplifica por que punir qualquer incompatibilidade. No final do processo de comparação, existindo uma lista de todas as combinações possíveis com as suas respectivas pontuações então, a mesma deve ser usada para reconstruir a imagem do documento.

Para realizar a reconstrução, de uma forma mais estruturada, adotou-se uma representação gráfica, que mostra claramente a relação entre todos os fragmentos. A Figura 5a mostra a aproximação poligonal de um documento composto por sete fragmentos enquanto a Figura 5b apresenta todas as relações encontradas pela LCS em um gráfico, isto é, a lista com as combinações e pesos. No grafo da Figura 5b é possível observar o peso dado a cada combinação entre fragmentos. Quanto maior o valor do peso, mais fortemente relacionados estão os fragmentos envolvidos.

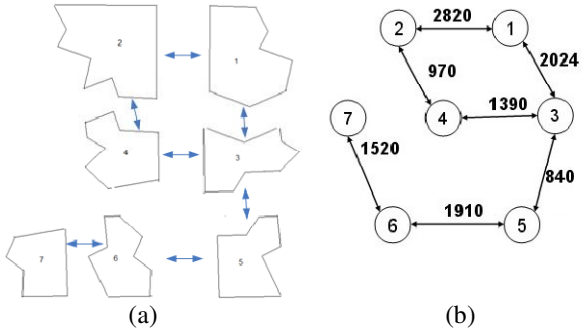


Figura 5. (a) Resultados da aproximação poligonal. (b) A representação gráfica das relações entre os fragmentos encontrados pelo algoritmo LCS.

2.4. Reconstrução

É possível observar na Figura 5, que o gráfico original pode possuir caminhos cíclicos, o que torna o problema de reconstrução mais complexo. Para atenuar tal complexidade, o gráfico gerado foi transformado em uma árvore geradora de custo mínimo, utilizando-se o algoritmo de Prim [10]. As seguintes alterações foram feitas no algoritmo para torná-lo adequado para a reconstrução do documento: i) Em vez de escolher aleatoriamente um vértice para iniciar o algoritmo, seleciona-se o vértice com maior peso, ii) Os melhores resultados obtidos tornam-se as conexões. Na verdade, poder-se-ia dizer que se está procurando o peso máximo de árvore estendida, pois quanto maior a pontuação melhor é a correspondência. iii) Para cada nó adicionado à árvore, deve-se alinhar o fragmento aos correspondentes, transladando e rotacionando o respectivo fragmento em relação a seu vizinho. Rotação e translação são guardadas para a reconstrução da imagem. Muitas vezes, pode-se ter uma correspondência de várias características, mas uma parte do fragmento pode obstruir o seu vizinho. Se isso vier a acontecer (falsos positivos), o nó é removido da árvore. Esse problema é mostrado na Figura 6, onde ocorre a oclusão entre um vértice e a área de encaixe.

Figura 6. (a) Combinação de dois fragmentos. (b) Oclusão produzida após o alinhamento.

É importante observar que a última modificação pode resultar na formação de várias árvores, tendo em vista a eliminação de um nó, isto ocorre pois existe a quebra da sequência entre os nós, tornando impossível encontrar um caminho, partindo do primeiro nó em direção ao último. Quando isso acontece, o documento será reconstruído apenas parcialmente, pois se considerou apenas os nós da maior árvore resultante. Depois da formação da árvore de Prim, a reconstrução do documento torna-se simples.

Parte-se de um nó e visitam-se todos os outros nós com as informações sobre rotação e translação encontradas anteriormente. A Figura 7 mostra a árvore de Prim para o gráfico apresentado na Figura 5 e respectivo documento reconstruído. É possível notar que algumas correlações podem ser removidas, como a 2-4 na figura 5 e 7, mas isso não produz impacto na reconstrução final.

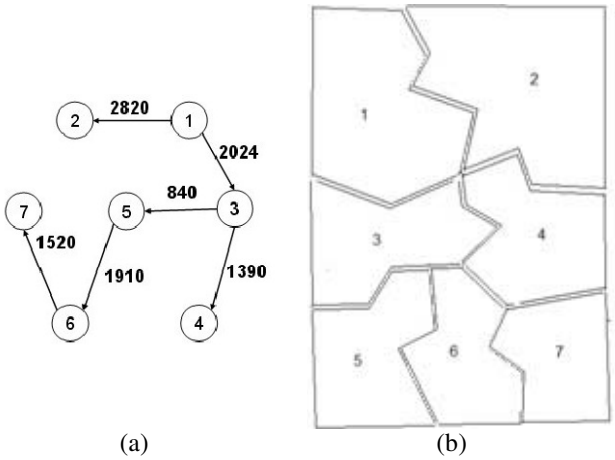


Figura 7. (a) A árvore de Prim. (b) o documento reconstruído.

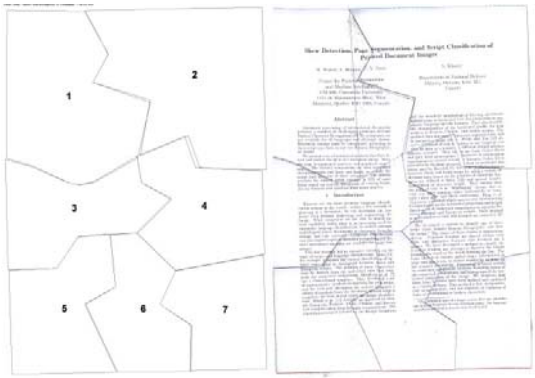
3. Experimentos

Para validar o método proposto, foi utilizada uma base de dados composta de 100 documentos. Estes documentos

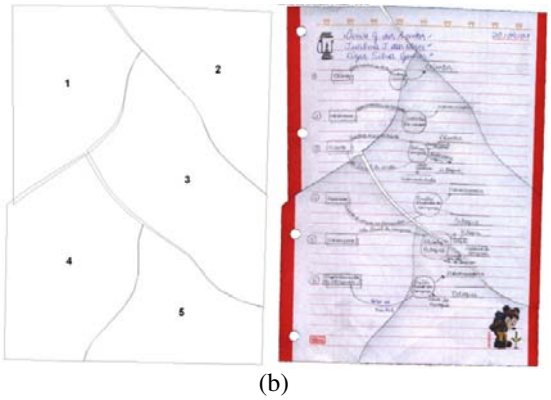
foram rasgados em 3 à 16 fragmentos e os tamanhos dos fragmentos ficaram entre 1x1 cm à 5x5 cm em média. É importante notar que os documentos foram aleatoriamente fragmentados, em outras palavras, não se usou critério para execução dessa tarefa. O mesmo banco de dados foi utilizado em [8].

A abordagem proposta foi capaz de reconstruir, em média, 75% dos fragmentos de cada documento. Em 61% dos documentos, a reconstrução foi completa ou apenas um fragmento se perdeu. Mesmo naqueles casos onde a reconstrução não é total, o método proposto se mostrou útil na redução do tempo de reconstrução, tendo em vista que é possível executar parcialmente o trabalho, reduzindo assim, a quantidade de fragmentos a recompor manualmente.

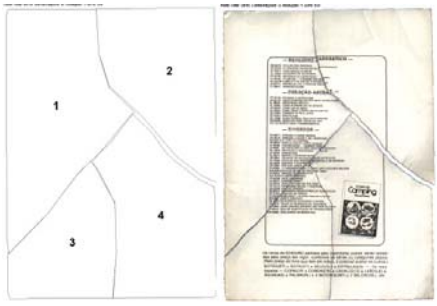
Na Figura 8 apresenta alguns exemplos de documentos totalmente e parcialmente reconstruídos. Pode-se notar que nos documentos reconstruídos por completo, os encaixes dos fragmentos, em alguns casos, não ocorreu perfeitamente. Isso se deve ao fato de a aproximação poligonal incorporar pequenas distorções nos processos de rotação e translação dos fragmentos.



(a)

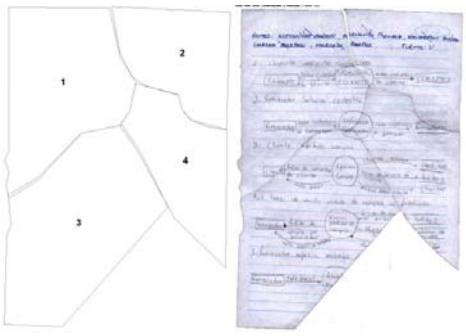


(b)

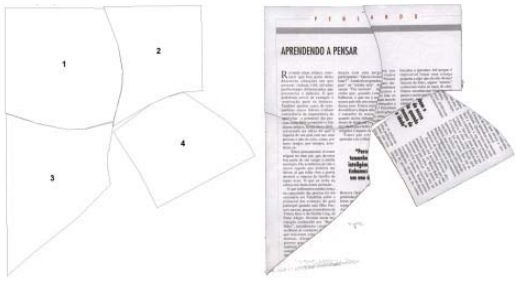


(c)

(d)



(e)



(f)

Figura 8. (a), (b), (c) e (d) Exemplos de documentos completamente reconstruídos. (c) Exemplo de documentos parcialmente reconstruídos. (d) Exemplo de documentos com fragmento incorretamente reconstruído.

A Tabela 2 compara os resultados obtidos nesse trabalho, com os relatados em [8], que utilizou a mesma base de dados. Como se pode notar, o método proposto traz um aumento considerável no desempenho do

algoritmo. É importante enfatizar que 75% de reconstrução correta não significam 75% do documento totalmente reconstruído, mas sim que, em média, o algoritmo é capaz de reconstruir 75% de cada documento. O mesmo vale para a taxa de falso positivo e de erro, apresentado na Tabela 2. Em média, o algoritmo falhou na correspondência entre fragmentos em 14%. Isto é, foi incapaz de encontrar as correspondências corretas, um exemplo pode ser visto na Figura 8d. O mesmo ocorreu com o erro, o algoritmo excluiu em média 11% de fragmentos por não conseguir uma associação com os demais. Um exemplo pode ser visto na Figura 8c.

A coluna de falso positivo representa os fragmentos que têm boas características para a correspondência, mas que foram escolhidas de forma claramente equivocada durante a reconstrução. Esta é uma das limitações da abordagem proposta, uma vez que somente foram consideradas as características geradas pela aproximação poligonal. Esse tipo de problema poderia ser atenuado pelo uso de algumas características contextuais, como textura e cor, por exemplo. Finalmente, a coluna do erro, que representa os fragmentos que não foram utilizados porque a LCS não encontrou nenhuma correspondência plausível. Uma fonte comum de erro neste caso é a falta de características discriminantes devido às bordas longas e lineares e também pelo reduzido número de vertices gerados neste caso. Este fato produz um número insuficuente de características para a associação adequada entre os vertices de cada fragmento.

Tabela 2. Resultados Comparativos.

Método	Reconstrução Correta (%)	Falso Positivo (%)	Erro (%)
Proposto	75	14	11
Justino et al [8]	57	24	19

4. Conclusão e Trabalhos Futuros

Neste artigo, foi proposto um método para reconstrução de documentos baseada em programação dinâmica e o algoritmo de Prim modificado. Através da aproximação poligonal foi possível reduzir a complexidade dos contornos dos fragmentos e extrair as características usadas como entrada para o algoritmo da programação dinâmica. A partir daí, uma estratégia baseada no algoritmo de Prim modificado é utilizada para realizar a reconstrução da imagem, podendo esse reconstrução ocorrer totalmente ou parcialmente.

Os resultados relatados neste trabalho mostram um aumento expressivo na taxa de reconstrução. No entanto, há muitas possibilidades de melhoria. Ainda considerando a estratégia proposta, acredita-se que o uso de outras cadeias de sequência produzidas pela matriz LCS, pode ser usada para resolver alguns problemas de falso positivo. Além disso, pode-se adicionar outro tipo de características, tais como o grau de curvatura, análise testura do fragmento, entre arestas.

5. Referências

[1] F.-H. Yao and G.-F. Shao, "A shape and image mergingtechnique to solve jigsaw puzzles," Pattern RecognitionLetters, vol. 24, pp. 1819–1835, 2003.

[2] H. Wolfson, "On curve matching," IEEE Trans. Pattern Anal. and Machine Intell., vol. 12, pp. 483–489, 1990.

[3] W. Kong and B. Kimia, "On solving 2D and 3D puzzles under curve matching," in CVPR, 2001, pp. 583–590.

[4] A. R. Willis and D. B. Cooper, "Computational reconstruction of ancient artifacts," IEEE Signal Processing Magazine, vol. 25, pp. 65–83, 2008.

[5] H. C. G. Leitao and J. Stolfi, "A multiscale method for the reassembly of two-dimensional fragmented objects," IEEE Trans. Pattern Anal. and Machine Intell., vol. 24, pp. 1239–1251, 2002.

[6] C. Papaodysseus, T. Panagopoulos, M. Exarhos, C. Triantafillou, D. Fragoulis, and C. Doumas, "Contourshape based reconstruction of fragmented," IEEE Trans. Signal Processing, vol. 50, pp. 1277–1288, 2002.

[7] P. De Smet, "Reconstruction of ripped-up documents using fragment stack analysis procedures," Forensic Science Intern., vol. 176, pp. 124–136, 2008.

[8] E. Justino, L. S. Oliveira, and C. Freitas, Reconstructing shredded documents through feature matching," Forensic Science Intern., vol. 160, 2005.

[9] David Douglas and Thomas Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," The Canadian Cartographer, vol. 10, pp. 112–122, 1973.

[10] R. C. Prim, "Shortest connection networks and some generalisations," Bell System Technical Journal, vol. 36, pp. 1389–1401, 1957.

[11] H. W. Chan, E. Gillespie, D. Leong, "Design and Implementation of a Paper De-shredder", ECE 412 Term Project Report, University of British Columbia Vancouver, Canada 2011.

Abordagem Morfológica de Segmentação do Melanoma

Arlete Teresinha Beuren[1], Rodrigo Janasieivicz Gomes Pinheiro[2], Jacques Facon[3]

[1]*CESUFOZ Centro de Ensino Superior de Foz do Iguaçu – Av. Paraná, 3695 - Jardim Central, CEP: 85863-720 - Foz do Iguaçu-Pr, Brasil;*

[1]*FAESI Faculdade de Ensino Superior de São Miguel do Iguaçu – Rua Valentim Celeste Palavro, 1501, 85877-000 Conjunto Panorama – São Miguel do Iguaçu-Pr, Brasil*

[2] *INVISYS Sistemas de Visão Computacional Ltda, Invisys, Curitiba-Pr, Brasil*

[3] *PPGIA-PUCPR Pontifícia Universidade Católica do Paraná – Rua Imaculada Conceição 1155, 80215-901 Curitiba-Pr, Brasil*

arlete_pr@yahoo.com.br, rjgpinheiro@gmail.com, facon@ppgia.pucpr.br

Abstract

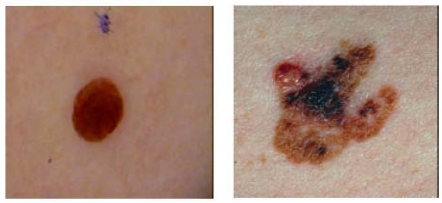
This paper presents a morphological-based approach for melanoma segmentation. By using color mathematical tools based on HSI lexicographic order, we show that it is possible to detect the melanoma lesion with accuracy. Experimental results have shown the efficiency of the methodology onto benign and malignant melanoma databases.

Resumo

O artigo apresenta uma abordagem morfológica de extração automática do câncer do melanoma. Operadores morfológicos baseados numa abordagem lexicográfica HSI são aplicados para filtrar e definir a região do melanoma que, em seguida, é segmentada por limiarização global. Exceto o tamanho do elemento estruturante e o número de iterações que são heuristicamente definidos, nenhum conhecimento a priori adicional sobre o tipo de melanoma é empregado. As potencialidades desta abordagem são descritas e ilustradas através de testes realizados em duas bases de imagens de melanoma benigno e maligno.

1. Introdução

A grande maioria das pintas ou nevos aumenta de volume durante a vida. Isso não significa que elas irão transformar-se em melanoma. Por isso, é muito importante fazer um exame regular dessas pintas consultando um dermatologista. As lesões benignas geralmente têm a forma arredondada ou oval com um contorno regular e cor uniforme. A figura 1-(a) mostra um típico nevo benigno. Uma transformação maligna é indicada como uma lesão grande com alteração na cor, aumento do diâmetro ou irregularidade das bordas, como mostra a figura 1-(b).



(a) **(b)**

Figura 1. Exemplos de pintas: (a) Nevo benigno, (b) Melanoma maligno

O melanoma maligno nada mais é do que um tumor de etiologia desconhecida [7]. Vários fatores estão relacionados, envolvendo genética, exposição ao sol e lesões traumáticas. Entretanto, a importância maior é a exposição ao sol, pois as pessoas que se expõe ao sol desde a infância têm maior chance de desenvolver o melanoma na idade adulta.

A incidência do melanoma maligno é crescente em todo o mundo, e segundo dados do GBM (Grupo Brasileiro do Melanoma) [7], esse tipo de câncer representa cerca de 4% das neoplasias malignas de pele. O melanoma é considerado o câncer menos comum e ao mesmo tempo o mais preocupante pela sua capacidade de disseminar-se a outras partes do corpo, mesmo quando a lesão ainda é pequena. Portanto os resultados do

tratamento do melanoma dependem essencialmente do diagnóstico precoce da doença, principalmente se detectado nos estágios iniciais.

Propõe-se uma abordagem morfológica de segmentação da lesão do melanoma. Uma ordem lexicográfica no espaço de cor HSI é empregada para filtrar e definir a região lesada pelo melanoma que é segmentada, em seguida, por uma técnica de limiarização global. Uma avaliação numérica da segmentação obtida e uma comparação com outra abordagem por aprendizagem *Fuzzy C-Means* são também realizadas para mensurar a eficiência da abordagem morfológica proposta. A estrutura do resto do artigo é a seguinte: a seção 2 irá detalhar a metodologia de filtragem e segmentação do melanoma. Na seção 3 serão apresentados os resultados obtidos a partir de duas bases de lesões benignas e malignas e as avaliações numéricas realizadas.

2. Metodologia

A metodologia proposta objetiva extrair a região da lesão de seu contexto por uma abordagem de filtragem morfológica em cor seguida de um processo de segmentação por limiarização. A metodologia proposta não envolve nenhuma técnica específica de remoção de pelos.

2.1. Morfologia matemática em cor por ordenação *HSI*

A técnica de filtragem baseia-se na morfologia matemática em cor onde a escolha do tipo de ordenação de cores é fundamental para o sucesso da abordagem. Decidiu-se adaptar a ordenação lexicográfica proposta em [8] inicialmente baseada no espaço de cor *HSV* para o espaço de cor *HSI*. Com base nessa abordagem, definiram-se os operadores básicos de erosão e dilatação em cor por ordenação lexicográfica *HSI*, conforme as fórmulas:

$$x_1(H_1,S_1,I_1) < x_2(H_2,S_2,I_2) \Leftrightarrow \begin{cases} I_1 < I_2 \\ ou \\ I_1 = I_2 ... e ... S_1 > S_2 \\ ou \\ I_1 = I_2 ... e ... S_1 = S_2 ... e ... H_1 < H_2, \end{cases}$$

$$x_1(H_1,S_1,I_1) = x_2(H_2,S_2,I_2) \Leftrightarrow I_1 = I_2 ... e ... S_1 = S_2 ... e ... H_1 = H_2, \quad (1)$$

A ordenação lexicográfica *HSI* sendo estabelecida, as definições da erosão e dilatação seguem a formulação tradicional. Define-se vetor de erosão de *f* em um ponto *x* pelo elemento estruturante *g* como segue:

$$(f \Theta g)(x) = \wedge\{f(z) - g_x(z)\} .. para .. z \in D[f] \cap D[g_x]. \quad (2)$$

E o vetor de dilatação de *f* em um ponto *x* pelo elemento estruturante *g* como segue:

$$(f \oplus g)(x) = \vee\{f(z) + g_{-x}(-z)\} .. para .. z \in D[f] \cap D[g_{-x}]. \quad (3)$$

2.2. Filtragem morfológica por cor

Através da filtragem morfológica por cor, objetiva-se aprimorar as cores da lesão no sentido de realçar a sua geometria perante o fundo, respeitando o seu formato sem deformá-la. Nenhuma técnica específica de remoção de pelos foi aplicada. Tanto a eliminação de pêlos quanto do ruído é implicitamente embutida nesta etapa. O processo morfológico adotado foi de abertura (erosão seguida de dilatação).

A figura 2 ilustra os resultados da abertura por ordenação lexicográfica HSI com elemento estruturante (EE) quadrado 3x3, 5x5 e 7x7. Testes exaustivos permitiram definir como melhor seqüência de filtragem o uso da abertura com elemento estruturante (EE) quadrado 7x7.

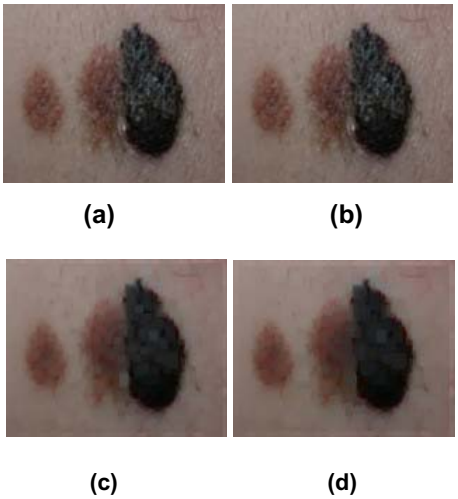


Figura 2. Exemplos de filtragem em cor: (a) Imagem original; (b) Abertura com EE Quadrado 3x3, (c) Abertura com EE Quadrado 5x5, (d) Abertura com EE Quadrado 7x7.

2.2. Segmentação por limiarização

A extração da região lesada faz-se por segmentação por limiarização global. Inicialmente as imagens filtradas coloridas são transformadas em imagens em níveis de cinza. Após ter testado vários métodos de binarização disponíveis na literatura, a limiarização global de Renyi [9] mostrou-se a mais eficiente no que diz respeito "ao resgate" das lesões sem prejudicar a geometria e o formato e sem adicionar ou remover partes das mesmas.

Através dos resultados da segmentação por limiarização mostrados na Figura 3, é possível observar alguns dos benefícios da filtragem morfológica por cor. Sem filtragem, as regiões segmentadas são imperfeitas. A abertura morfológica por cor propiciou a remoção do ruído colorido e dos brilhos excessivos preservando a geometria e forma das lesões. Resultando em regiões segmentadas mais perfeitas.

2.3. Filtragem morfológica binária

Esta etapa consiste no preenchimento de lacunas na região segmentada do melanoma e na remoção do ruído. O preenchimento foi realizado por reconstrução morfológica binária [5]. A remoção do ruído busca eliminar mínimos e máximos locais inferiores a um dado tamanho sem afetar as estruturas da região segmentada do melanoma. Essa etapa foi realizada por abertura binária com elemento estruturante quadrado 7x7.

A Figura 4 ilustra resultados obtidos com imagens apresentando grandes variações de cor, de bordas e de forma. Percebe-se a eficiência da abordagem proposta quando se compara as imagens obtidas pela abordagem proposta (Figura 4-(c)) com as imagens da segmentação ideal (ground-truth) da Figura 4-(d).

3. Experimentos

Para mensurar a eficiência da abordagem proposta, dois procedimentos foram realizados. O primeiro procedimento consistiu numa avaliação numérica com doze métricas a partir de 2 bases de imagens de melanoma benigno e maligno, cada uma contabilizando 100 imagens. As imagens foram coletadas através do link [11] que descreve e apresenta diferentes tipos de lesões benignas e malignas. Para cada imagem foi gerada, manualmente por um especialista em processamento de imagens, a imagem da segmentação ideal (ground-truth) [1] [2], como ilustrado na Figura 4-(d).

Doze métricas foram adotadas para avaliar a eficiência da metodologia proposta que são: o erro de má classificação *ME* [10], o erro relativo da área do primeiro plano *RAE* [10], a taxa de verdadeiros positivos *TP* e verdadeiros negativos *TN*, os erros de falsos positivos *FP* e falsos negativos *FN*, a precisão *P* [4], a acurácia *A* [4], o recall *R* [4], o erro *ER* [4], a métrica *F-Measure FM* [6] e a métrica da taxa negativa *NRM* [6].

O segundo procedimento foi comparar a metodologia proposta com a abordagem de segmentação do melanoma de Cheng e Umbaugh [3] que foi modificada e aprimorada. Substituiu-se o espaço de cor *RGB p*elo espaço de cor *TSL* considerado por Vezhnevets et al [12] como o espaço de cor mais apropriado para processar imagens de pele humana. O número de características foi reduzido de sete para cinco (média, desvio padrão, skewness e kurtosis) para cada um dos três canais Tintura, Saturação e Luminância do espaço *TSL*. Reduzindo assim o vetor de características em 12 características mo lugar de 17 na abordagem original de Cheng e Umbaugh [4]. Por fim, o classificador multi-layer perceptron (*MLP*) foi substituído pelo classificador *Fuzzy C-Means* [13].

As mesmas métricas e bases de imagens de melanoma acima citadas foram utilizadas para avaliar a abordagem modificada *Fuzzy C-means*. As tabelas 1 e 2 ilustram, através das doze métricas, os resultados das avaliações numéricas da abordagem proposta e da abordagem modificada *Fuzzy C-means* para as duas bases.

Em relação à abordagem original de Cheng e Umbaugh que obteve 78% de sucesso na segmentação do melanoma para uma base de 160 imagens coloridas de tumores [4], a abordagem modificada obteve uma maior taxa de segmentação com taxas de verdadeiros positivos *TP* (86,69±18,80)% e (89,33±18,14)% para as bases de melanoma benigno e maligno, respectivamente.

A abordagem morfológica proposta superou a abordagem modificada com taxas bem superiores de *TP* (95,67± 06,17)% e (97,22±04,96)% para as bases de melanoma benigno e maligno, respectivamente. Essa maior eficiência se reflete não somente em relação às taxas de verdadeiros positivos *TP*, mas também aos valores das métricas *ME, RAE, TN*, *FP, FN*, *P*, *A*, *R*, *ER*.. E também em relação às taxas de desvio padrão.

Podemos ainda enfatizar que as taxas *FM* e *NRM*, consideradas por [6] como sendo as mais representativas em termos de eficiência, são de (95,22±03,88)% e (04,79±03,79)% para a base de melanoma benigno e de (94,65± 03,78)% e (05,56 ± 03,96)% para a base de melanoma maligno. Podemos perceber ainda que os

valores de desvio padrão são aproximadamente quatro vezes inferiores aos valores de desvio padrão das mesmas taxas para a abordagem modificada *Fuzzy C-means*. O que demonstra a eficiência da metodologia proposta.

4. Conclusões

Neste artigo, uma metodologia de extração automática do melanoma baseada numa abordagem morfológica em cor foi apresentada. Usando pouco conhecimento à priori, a seqüência morfológica por ordenação lexicográfica *HSI* foi decisiva para filtrar corretamente a lesão do melanoma, quaisquer que sejam as variações de cor e/ou de formato. Os valores das métricas utilizadas para avaliar a eficiência da metodologia proposta comprovam que apesar da abordagem ser relativamente simples, ela é promissora.

5. Referências

[1] A. T. Beuren, J. Facon, "Segmentação automática do Melanoma benigno: Criação da base ground-truth", relatório técnico, PPGIa Pontifícia Universidade Católica do Paraná, 2010

[2] A. T. Beuren, J. Facon, "Segmentação automática do Melanoma maligno: Criação da base ground-truth", relatório técnico, PPGIa Pontifícia Universidade Católica do Paraná, 2010

[3] Y. Cheng, S. Umbaugh, "Color-based diagnosis: Clinical Images", Computer Vision and Image Processing Research Lab@ECE Dept., SIUE, https://www.ee.siue.edu/CVIPtools, 2005.

[4] A. Del Bimbo, "Visual Information Retrieval", Morgan Kaufmann 1999

[5] J. Facon, "Morfologia matemática: teoria e exemplos", Editor Jacques Facon, 196 páginas, (1996).

[6] B. Gatos, K. Ntirogiannis, I. Pratikakis, " ICDAR 2009 Document Image Binarization Contest (DIBCO 2009)", International Conference on Document Analysis and Recognition, 1375–1382 (2009).

[7] GBM Grupo Brasileiro Multidisciplinar e Multicêntrico de Estudo do Melanoma. Disponível em: http://www.gbm.org.br/gbm/. Acesso em 31 jan. 2011.

[8] G. Louverdis. M. I. Vardavoulia, I. Andreadis, P. Tsalides, "A New Approach To Morphological Color Image Processing", Pattern Recognition No 35, pp 1733-1741, July 2001.

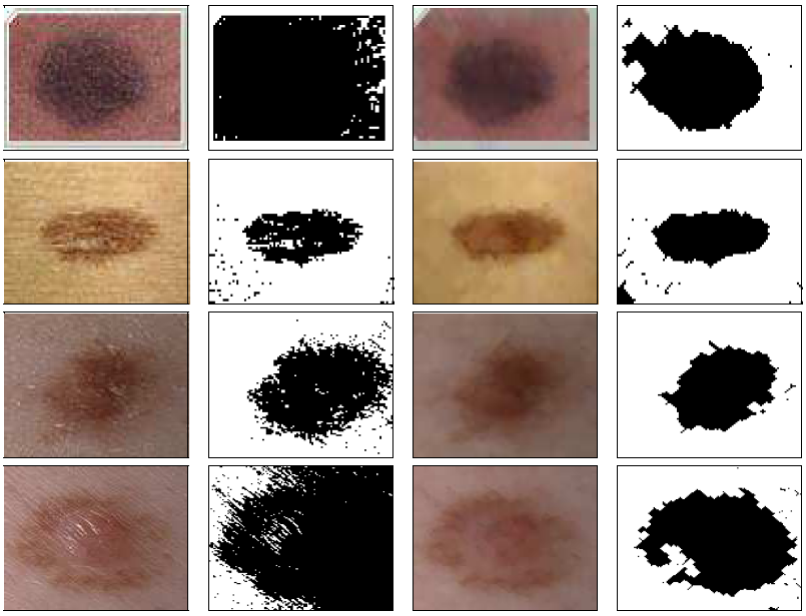
[9] P. Sahoo, C. Wilkins, J. Yeager, "Threshold Selection Using Renyi's Entropy", Pattern Recognition Vol 30, No 1, Pp 71-84 (1997)

[10] M. Sezgin, B. Sankur, "Selection of thresholding methods for non destructive testing application", International Conference on Image Processing, ICIP2001, Vol 3 p764–767 (2001).

[11] Testez-vous en images. ONCOPEAU. Disponível em http://info-melanome.net/pub/en_savoir_plus/testez_vous_en_images Acesso em 02 de março de 2011.

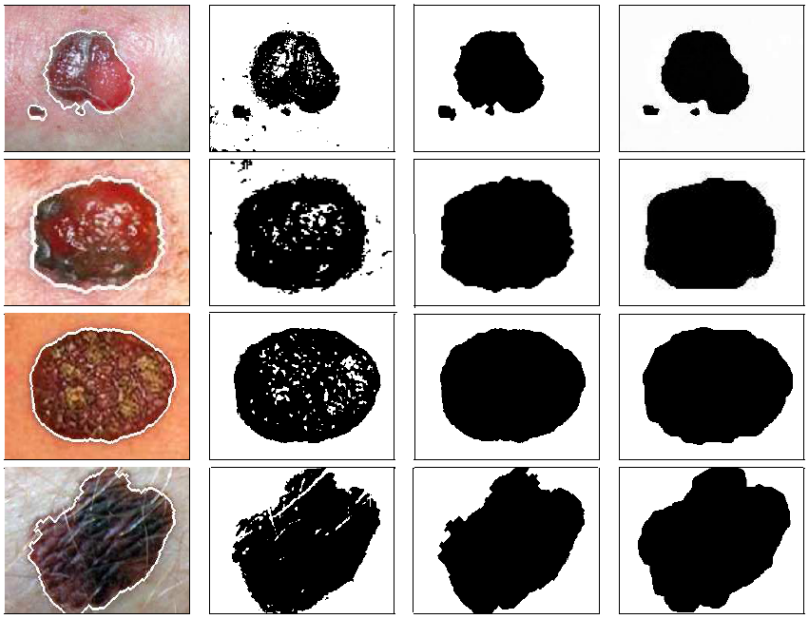
[12] V. Vezhnevets, V. Sazonov, A. Andreeva, "A Survey on Pixel-Based Skin Color Detection Techniques", Proc. of Graphicon p 85 -92 (2003)

[13] Y. Yong, Z Chongxun, L. Pan, "A novel Fuzzy C-Means Clustering Algorithm for Image Thresholding", Measurement Science Review, vol. 4, Section 1, 2004.



(a) (b) (c) (d)

Figura 3. Benefícios da filtragem morfológica por cor: (a) Imagem original (b) Binarização sem filtragem (c) Filtragem (d) Binarização após filtragem.



(a) (b) (c) (d)

Figura 4. Resultados da abordagem: (a) Imagem original com borda da região segmentada (b) Binarização (c) Filtragem morfológica binária (d) Segmentação ideal (ground-truth).

Tabela 1: Avaliação da abordagem proposta a partir das doze métricas (em %).

	ME	RAE	TP	TN	FP	FN	P	A	R	ER	FM	NRM
Benignas	04,35	12,37	95,67	94,75	05,25	04,33	95,26	95,21	95,67	04,79	95,22	04,79
Benignas Desvio Padrão	04,00	13,87	06,17	06,67	06,67	06,17	05,57	03,79	06,17	03,79	03,88	03,79
Malignas	05,84	12,30	97,22	91,66	08,34	02,78	92,62	94,44	97,22	05,56	94,65	05,56
Malignas Desvio Padrão	04,46	09,10	04,96	07,75	07,75	04,96	06,24	03,96	04,96	03,96	03,78	03,96

Tabela 2: Avaliação da abordagem *Fuzzy C-Means* a partir das doze métricas (em %).

	ME	RAE	TP	TN	FP	FN	P	A	R	ER	FM	NRM
Benignas	08,00	17,25	86,69	92,63	07,15	13,26	93,41	89,78	86,73	10,22	88,33	10,20
Benignas Desvio Padrão	08,00	17,37	18,80	09,22	09,15	18,81	07,27	09,12	18,81	09,12	12,52	09,12
Malignas	10,40	18,45	89,33	88,43	11,57	10,67	89,79	88,88	89,33	11,12	87,82	11,12
Malignas Desvio Padrão	07,04	16,24	18,14	10,73	10,73	18,14	07,86	08,84	18,14	08,84	13,33	08,84

Mosaico Automático de Imagens Agrícolas Aéreas

André de Souza Tarallo[1]; Maria Stela Veludo de Paiva[1]; Lúcio André de Castro Jorge[2]

[1]Escola de Engenharia de São Carlos - Universidade de São Paulo - USP, Depto. de Engenharia Elétrica, São Carlos-SP
[2]Pesquisador, Embrapa Instrumentação Agropecuária, São Carlos – SP.

andre.tarallo@gmail.com, mstela@sc.usp.br, lucio@cnpdia.embrapa.br

Abstract

The generation of automatic mosaics from high-resolution digital images is a very important tool, because it increases the efficiency in obtaining results and helps fast decision-making. In agriculture, the automatic image mosaics of forest and crop areas, helps different applications in precision farming that needs to answer in a very fast way. A problem of this area is joining two or more images in a new image without great distortion. In this paper it is presented a tool for automatic mosaic reconstruction, using a method based on SIFT transform to join sequential images. Satisfactory results were obtained using two image databases. Image pairs were bonded with 100% accuracy in one of the bases. At another base, this percentage was 66.6%, but the reasons are discussed. It was concluded that tool minimize the time to obtain results (compared with the manual method executed on the field).

Keywords: mosaics, SIFT, agriculture

1. Introdução

Até o presente momento, a construção de mosaicos na agricultura vem sendo feita de maneira semi-automática, necessitando obter o modelo digital do terreno, fazer a ortorretificação de imagens e colocação manual de bandeirinhas, que são usadas como marcadores para que um software possa construir um mosaico deste terreno. Utilizando este tipo de implementação, a construção de mosaicos é demorada e trabalhosa.

Com os mosaicos é possível direcionar vistorias de campo durante o ciclo do cultivo ou em datas posteriores à colheita, possibilitando fornecer um diagnóstico preciso da área de cultivo. A partir daí, podem ser elaborados os mapas de recomendações: descompactação, fertilidade e aplicação de insumos em taxa variável.

Mais recentemente, aplicações na agricultura passaram a exigir maior rapidez na construção destes mosaicos para possibilitar a obtenção mais rápida de informações para tomada de decisões relativas a controle de pragas, doenças ou queimadas.

Este fato levou à proposta deste projeto, da construção automática de mosaicos para a agricultura a partir de imagens digitais aéreas obtidas com uma câmera de alta resolução acoplada em um avião. A alta resolução da câmera minimiza possíveis problemas de distorção nas imagens, causados pela distância entre a aeronave e o solo.

O problema em utilizar imagens digitais de grande dimensão e alta qualidade de imagens, é que a geração dos mosaicos se torna lenta, devido ao custo computacional. Outro problema é que imagens sequenciais, com mesmas características (mesma textura), dão a impressão de uma única área contínua e, podem dificultar a detecção de pontos comuns entre pares de imagens para formar um mosaico.

Pode ser observado nos trabalhos de [1], [2], [3] e [4], exemplos de aplicações com mosaicos automáticos de imagens aéreas. Um dos problemas desta área é unir duas ou mais imagens, em uma nova imagem com a menor distorção possível.

A estratégia para desenvolver este projeto foi baseada na pesquisa de [5], [6], [7] e [8], com algumas modificações.

Na metodologia empregada foi utilizada a técnica SIFT (*Scale Invariant Features Transform*) [6] para extrair as características de cada uma das imagens (Seção 3), uma *K-d tree* modificada, denominada BBF (*Best-Bin-First*) para encontrar as características correspondentes entre pares de imagens (Seção 4) e o algoritmo RANSAC (*Random Sample Consensus*) para filtrar os pontos falsos correspondentes (Seção 5). Após estes passos, é realizada a junção entre um par de imagens sequenciais, fazendo uma transformada em perspectiva de uma imagem em relação à outra, ou seja, o cálculo da matriz homográfica.

2. Base de Imagens

As bases de imagens utilizadas neste projeto envolvem áreas de vegetação, pastagens e áreas urbanizadas. Foram utilizadas duas bases de imagens, com 48 imagens em cada base. Uma das bases de imagens possui padronização para a obtenção das imagens e a outra não.

As imagens da primeira base de imagens (base 1) foram capturadas em Santa Rita do Sapucaí – MG. As imagens possuem dimensão de 3504 x 2336 pixels, com 24 bits por pixel e 72 *dpi*.

A base 1 foi adquirida de acordo com o seguinte procedimento: a câmera foi acoplada ao suporte do avião de maneira bem horizontal à superfície da terra; o avião percorreu sempre a mesma distância, fazendo movimentos horizontais na área demarcada, para obtenção das imagens; a obtenção entre uma imagem e a próxima imagem foi feita com um tempo determinado e síncrono, entre a imagem anterior e a posterior (3 segundos). Na figura 1 é apresentada uma imagem exemplo da base 1.

Figura 1. Exemplo de imagem utilizada na base 1.

As imagens da segunda base de imagens (base 2) foram capturadas em Araraquara – SP. As imagens foram obtidas manualmente, por um engenheiro de dentro do avião. As imagens possuem dimensão de 3888 x 2592 pixels, com 24 bits por pixel e 72 *dpi*.

A base 2 não seguiu um padrão para a obtenção de suas imagens, como: as imagens não foram obtidas de maneira horizontal em relação à superfície da terra, não tendo como fixar um ângulo de obtenção, já que a obtenção foi manual. Não houve tempo sincronizado de obtenção entre a imagem anterior e posterior; o avião nem sempre fez os movimentos horizontais na área demarcada para obtenção das imagens.

Figura 2. Exemplo de imagem utilizada na base 2.

Estes fatores dificultaram a construção do mosaico, uma vez que muitas características da imagem anterior não aparecem na posterior, e também porque o ângulo de obtenção entre a imagem anterior e posterior era diferente, ocasionando a perda de informações nas imagens. Na figura 2 é possível observar uma imagem exemplo da base 2 de imagens.

Pode ser observada na Figura 1, a presença de uma borda preta ao redor da imagem. Esta borda é o suporte onde foi acoplada a câmera. Em todas as imagens existe a presença desta borda preta, mas através de um pré-processamento esta borda foi tratada para não afetar o resultado do mosaicamento das imagens.

3. SIFT

A SIFT [5], [6] é capaz de transformar uma imagem, em uma coleção de vetores de características locais (descritores de características), e cada um desses vetores são invariantes à escala, rotação e parcialmente invariante à mudanças de iluminação e ponto de vista.

As características fornecidas pela SIFT são bem localizadas em ambos os domínios, o da freqüência e o do espaço, reduzindo assim a probabilidade de não haver correspondência das características por oclusão ou ruído. As características são altamente distintas, permitindo que uma simples característica seja corretamente correspondida com alta probabilidade diante de um grande banco de dados de características, possibilitando assim, uma base para o reconhecimento de objetos e cenas.

O custo de extrair essas características é minimizado por meio de uma abordagem de filtragem em cascata, na qual as operações com maior custo operacional são executadas apenas em locais que passaram em testes iniciais.

Os quatro principais estágios que compõem a SIFT para gerar o conjunto de características de imagens são:

- **Detecção de extremos no Espaço Escala:** Neste primeiro estágio é feito a procura por todas as escalas e locais de uma imagem. Para isto é utilizada uma função conhecida como Diferença da Gaussiana, para identificação dos potenciais pontos de interesse que são invariantes à escala e orientação. Esta é a parte mais custosa do algoritmo.
- **Localização dos pontos chave:** Para cada local candidato, é determinada a sua posição e escala. Os pontos chave são selecionados baseados em medidas de sua estabilidade.
- **Definição da Orientação:** Uma ou mais orientações são atribuídas para cada ponto chave localizado, baseada em direções do gradiente. Todas as operações posteriores são realizadas sobre os dados da imagem que foram considerados ponto chave e que foram transformados em relação à orientação, escala e

localização, proporcionando invariância a estas transformações.

- **Descritor dos Pontos Chave:** Os gradientes da imagem são mensurados na escala selecionada, na região ao redor de cada ponto chave, sendo criados histogramas de orientações para compor o descritor.

Com as características extraídas a partir de todas as imagens, as mesmas devem ser pareadas. Na figura 3 pode ser observado um exemplo da localização de características através da SIFT.

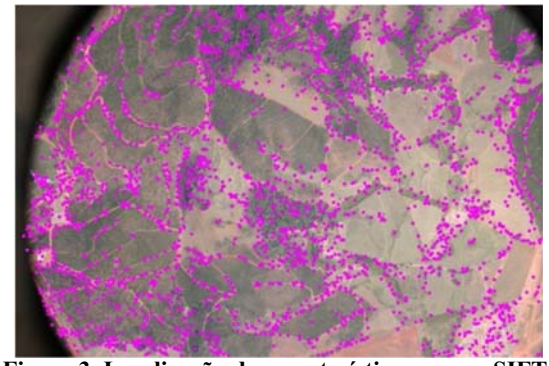


Figura 3. Localização de características com a SIFT.

A técnica SIFT transforma dados da imagem em coordenadas invariantes à escala, relativas às características locais. Um aspecto importante é o grande número de características geradas, que cobrem densamente toda a imagem.

A quantidade de características é importante para reconhecimento de objetos, sendo que para pequenos objetos em imagem de fundo complexo (imagem cujo fundo contém informação não relevante), são necessárias pelo menos três características corretamente correspondidas de cada objeto para uma identificação confiável.

4. BBF

Uma vez aplicada a técnica SIFT sobre as imagens, é possível encontrar a correspondência entre duas imagens de acordo com os pontos chaves detectados. Há a comparação dos descritores das duas imagens, encontrando os melhores candidatos a serem seus equivalentes na outra imagem.

O melhor candidato correspondente para cada ponto chave é encontrado, identificando os seus vizinhos mais próximos na base de dados dos pontos chave a partir de imagens de entrada. O vizinho mais próximo é definido como o ponto chave, com distância euclidiana mínima entre os descritores em questão.

A maneira mais eficaz de identificar o melhor ponto candidato é obtida através da comparação da distância do vizinho mais próximo ao de um segundo vizinho mais próximo.

Quando se procura classificar uma imagem em um extenso banco de dados de descritores para vários objetos, a busca exaustiva de vizinho mais próximo pode ser demorada e para tal existe a técnica BBF para acelerar a busca.

O algoritmo BBF usa uma busca ordenada modificada de um algoritmo *k-d tree* de modo que as posições no espaço das características são procuradas na ordem de suas distâncias mais próximas a partir do local investigado.

Uma razão para o algoritmo BBF [9] funcionar bem é que somente são consideradas correspondências nas quais o vizinho mais próximo é menor do que 0,8 vezes a distância do segundo vizinho mais próximo e, portanto não é necessário resolver os casos mais difíceis, nos quais muitos vizinhos tem muitas distâncias similares.

5. Ransac

Após a correspondência dos pontos chave, os mesmos são usados para encontrar uma transformada que mapeie as posições dos pontos de uma imagem para as posições dos pontos correspondentes, na outra imagem.

Primeiramente, é usado o RANSAC para excluir falsas correspondências, ou seja, selecionar um conjunto de *inliers* (características consistentes de um conjunto de características) que são compatíveis com a homografia entre as imagens. Após o RANSAC, haverá o cálculo das homografias entre esses pares sequenciais.

RANSAC [10] é um procedimento de estimação robusto que usa um conjunto mínimo de correspondências amostradas randomicamente para estimar os parâmetros de transformação da imagem e achar a solução que tem o melhor consenso com os dados. Para o cálculo das homografias [11] devem ser selecionados conjuntos de r=4 correspondências de características; o cálculo da homografia H é realizado utilizando o método da transformação linear direta (DLT). Isto deve ser repetido n vezes (n = 200) e após, selecionar a solução que tem o maior número de *inliers* (cujas projeções são consistentes com H com uma tolerância de x pixels) com o uso do RANSAC.

A figura 4 apresenta um exemplo de aquisição dos *inliers* entre duas imagens sequenciais.

5.1. Matriz Homográfica

Existem vários fatores a serem considerados ao fazer mosaicos de partes de imagem de um mesmo panorama. Entre esses fatores estão as transformações que podem ocorrer entre uma imagem e outra devido às distorções causadas pelas lentes da câmera. Para o mosaicamento de imagens, é então, necessário o tratamento dessas imagens, para a correção de tais distorções.

A matriz homográfica [11] mapeia uma transformação do plano projetivo real para o plano projetivo transformado, fazendo as correções geradas por essa transformação.

Um ponto 2D (x,y) em um imagem pode ser representado por um vetor 3D x=(x1,x2,x3), sendo x=x1/x3 e y=x2/x3. Isto é denominado de representação homogênea de um ponto e encontra-se no plano projetivo P^2. Uma homografia é um mapeamento invertível de pontos e linhas no plano projetivo P^2.

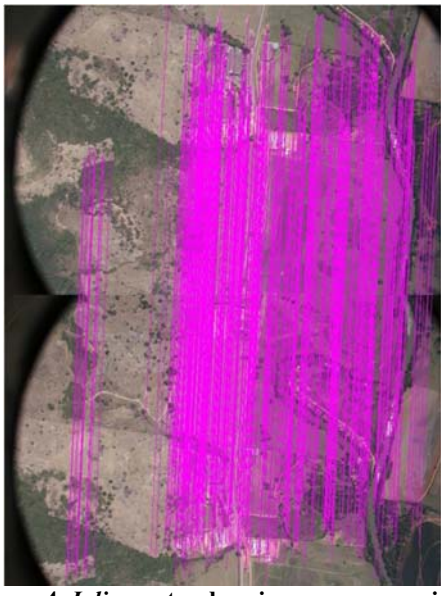


Figura 4. *Inliers* entre duas imagens sequenciais.

Um mapeamento a partir de $P^2 \rightarrow P^2$ é uma projetividade se e somente se existe uma matriz H 3x3 não singular tal que para qualquer ponto em P^2 representado por um vetor x é verdade que seu ponto mapeado é igual a Hx. Isso mostra que para calcular a homografia que mapeia cada xi com o seu correspondente xi', é suficiente calcular a matriz homografia H 3x3, conforme equação 1.

$$\begin{pmatrix} x1' \\ x2' \\ x3' \end{pmatrix} = \begin{bmatrix} h1 & h2 & h3 \\ h4 & h5 & h6 \\ h7 & h8 & h9 \end{bmatrix} \begin{pmatrix} x1 \\ x2 \\ x3 \end{pmatrix} \qquad (1)$$

Sendo x'=Hx. É possível notar que H pode ser modificado, multiplicando por uma constante não zero (fator de escala) sem alterar a transformação projetiva. Consequentemente é, possível dizer que H é uma matriz fundamental. Há oito independentes relações, mesmo H tendo nove elementos, formalizando oito graus de liberdade. Estes oito parâmetros precisam ser calculados.

A matriz homográfica possui 8 graus de liberdade no que diz respeito a isometria, transformação de similaridade, transformação afim, transformação projetiva e projeção em perspectiva. Há transformações simples que contém restrições específicas, reduzindo o número de graus de liberdade.

Homografias são estimadas entre imagens para detectar características correspondentes nessas imagens. Dentre os algoritmos capazes de estimar a homografia é possível citar o DLT (Transformação Linear Direta) e o RANSAC.

Neste projeto é utilizada a transformada em perspectiva, que transforma a imagem de origem através de uma matriz específica, conforme a equação 2.

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix} \qquad (2)$$

Utilizou-se a função *cvWarpPerspective* da biblioteca livre desenvolvida pela Intel, OpenCV [12], para realizar a transformação da imagem (projeção) e assim realizar o mosaicamento da imagem, através da seguinte estratégia:

$$x' = H * x \qquad (3)$$

sendo H a matriz homográfica, x a imagem no plano projetivo real e x' a imagem no plano projetivo.

Após ser realizada a transformação em perspectiva, as imagens estão prontas para serem unidas. Na figura 5 é possível visualizar um exemplo do pareamento entre duas imagens.

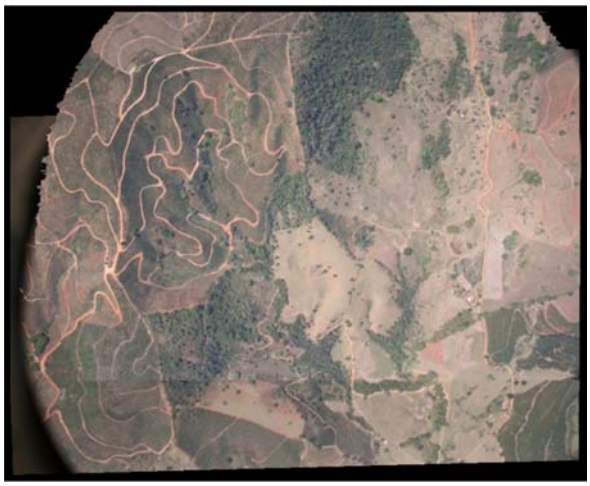


Figura 5. Exemplo de duas imagens unidas.

6. Resultados e Discussão

Para a geração dos resultados, por motivos de custo computacional as imagens da base 1 e base 2 tiverem suas dimensões reduzidas em cinco vezes.

De acordo com os padrões adotados na obtenção das imagens da base 1, foi possível parear todas as imagens desta base sem nenhuma falha (24 pares gerados). A tabela 1 mostra alguns valores médios obtidos com o total de imagens da base 1. Esses valores são: quantidade de pontos detectados na imagem 1 (img1) dos respectivos pares, quantidade de pontos detectados na imagem 2 (img2), quantidade de pares de pontos correspondentes detectados antes da aplicação do RANSAC (antes_RANSAC) e a quantidade de pontos correspondes (*inliers)* após aplicação do RANSAC (após_RANSAC) e o tempo médio gasto para gerar os 24 pares de imagens, em segundos.

Tabela 1 – Resultados da base 1 de imagens.

	img1	**img2**	**antes_ RANSAC**	**após_ RANSAC**	**tempo (s)**
média	4234,6	4182,5	628,5	419,3	840

Já para a base 2 de imagens, que não teve uma padronização na obtenção de suas imagens, foi possível fazer o pareamento de apenas 16 pares de imagens, de um total de 24 (66,6% de acerto). A tabela 2 mostra os resultados da base 2 de imagens, utilizando os mesmos parâmetros dos resultados da base 1.

Tabela 2 – Resultados da base 2 de imagens.

	img1	**img2**	**antes_ RANSAC**	**após_ RANSAC**	**tempo (s)**
média	3152,3	3189,8	505,6	478,1	612

É possível observar, comparando os resultados da tabela 1 com o da tabela 2, que os resultados da tabela 2 foram inferiores ao da tabela 1 devido à falta de padronização na obtenção das imagens que geraram os resultados da tabela 2 (base 2).

7. Conclusões

De acordo com os resultados obtidos, verifica-se a possibilidade desta ferramenta funcionar de maneira satisfatória, com geração de resultados confiáveis e de maneira automática, auxiliando com rapidez no controle de pragas, desmatamento e queimadas.

A ferramenta pode ser modificada para atender outras necessidades na área, necessitando para isso somente adaptações para a nova aplicação em questão, baseada na análise das características que se deseja analisar.

8. Referências Bibliográficas

[1]- Liu, J.; Chen, Z.; Guo, R. A Mosaic Method for Aerial Image Sequence by R/C Model. International Conference on Computer Science and Software Engineering – IEEE, 2008.

[2]- Abedini, A., Hahn, M., Samadzadegan, F. An Investigation Into the Registration of Lidar Intensity Data and Aerial Images Using the SIFT. Proceedings of The International Society for Photogrammetry, and Remote Sensing - ISPRS, 2008.

[3]- Goltz, G. A. M.; SHIGUEMORI, E. H. . Aplicação do Algoritmo SIFT em Imagens de Navegação Autônoma. Workshop Anual de Pesquisa e Desenvolvimento do IEAv, 2008.

[4]- Lin, Q. Yu, and G. Medioni, Map-Enhanced UAV Image Sequence Registration. IEEE Workshop on Applications of Computer Vision (WACV), 2007.

[5]- Lowe D. G. Object Recognition from Local Scale-Invariant Features. Proceeding of ICCV. Institute of Electrical and Electronics Engineers Inc, 1999.

[6]- Lowe D. G. Distinctive Image Features from Scale Invariant Keypoints. International Journal of Computer Vision, 2004.

[7] - Brown, M. and Lowe D. G. Recognising Panoramas, International Conference on Computer Vision, 2003.

[8] - Brown, M. and Lowe D. G. Automatic Panoramic Image Stitching using Invariant Features, International Conference on Computer Vision, 2007.

[9] - BEIS, J.; LOWE, D.G.. Shape indexing using approximate nearest-neighbour search in highdimensional spaces. Conference on Computer Vision and Pattern Recognition, Puerto Rico, pp. 1000-1006, 1997

[10] - FISCHLER, M. A.; BOLLES, R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM, v. 24, p. 381- 395, 1981.

[11]- Dubrofsky, E. Homography Estimation, dissertação de mestrado, Universidade de Columbia-Vancouver, 2009.

[12]- http://opencv.willowgarage.com/wiki/, acessado em 12 de Janeiro de 2011.

Segmentação de Imagens de Escoamentos Bifásicos Gás-líquido Baseada em Métodos Level Set e Morfologia Matemática

Leyza Baldo Dorini, Mauren Coelho de Andrade e Lucia Valeria Ramos de Arruda
Universidade Tecnológica Federal do Paraná
Departamento de Informática/Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial
{leyza, maurensguario, lvrarruda}@utfpr.edu.br

Resumo

O estudo de fluxos bifásicos é essencial para a modelagem de diversos sistemas. Neste contexto, técnicas de processamento de imagens tem sido cada vez mais utilizadas, pois possibilitam a identificação e avaliação visual de importantes características por meio de soluções relativamente simples para projetar/implementar e que não causam distúrbios no fluxo. Este artigo apresenta uma metodologia para medir os parâmetros de interesse baseada em métodos Level Set e em transformações de morfologia matemática. Devido a sua acurácia e versatilidade, métodos Level Set vêm sendo amplamente considerados, sendo uma de suas principais vantagens a sua capacidade de tratar eficazmente mudanças topológicas e/ou descontinuidades que podem eventualmente ocorrer durante a propagação de interfaces. O esquema proposto foi aplicado com sucesso em um grande número de imagens, conduzindo a resultados precisos que incentivam trabalhos futuros. Além disso, foram realizados testes de classificação de padrões de fluxo, os quais também geraram resultados promissores.

1. Introdução

O estudo de escoamentos bifásicos é relevante em diversas áreas, incluindo transporte de óleo/gás, energia nuclear e processamento químico. Medidas confiáveis sobre a distribuição das fases no canal de fluxo são cruciais para a modelagem precisa dos sistemas envolvidos. Sensores capacitivos tem sido considerados, sendo que o sinal de saída instantâneo também é utilizado para classificar o padrão do fluxo. Contudo, o efeito da interface e a velocidade relativa entre as duas fases dificulta tais estimativas [2] [14].

Por outro lado, técnicas de processamento de imagens constituem uma ferramenta poderosa para estudar fluxos bifásicos, sendo tipicamente não-intrusivas e relativamente simples para projetar e implementar. No que se refere a escoamentos bifásicos horizontais gás-líquido, foco deste trabalho, diversas abordagens utilizam tais técnicas para extrair a região de interesse para análises posteriores. Em Shi *et al.* [12], é proposta uma abordagem baseada em limiarização para segmentar as bolhas, mas esta mostrou-se extremamente sensível a presença de ruídos e variações nas condições de iluminação. Em Dihn et al. [3], é aplicada uma abordagem baseada em subtração de fundo seguida por detecção de bordas, que também apresenta problemas em imagens com ruídos, por exemplo.

A Transformada de Hough é utilizada por Yu *et al.* [15] como base em uma técnica de segmentação para escoamentos bifásicos sólido-líquido, mas esta é limitada a formas que podem ser parametrizadas. Em Grau e Heiskanen [5], a imagem pré-processada por um filtro passa-alta é utilizada para segmentação via transformada de *watershed*. Porém, assume-se que a bolha pode ser aproximada por formas esféricas, o que não ocorre para diversos padrões de escoamento. O método de detecção de partículas proposto por Zabulis et al. [16] é baseado em rastreamento de bordas e casamento de padrões. Entretanto, quando a densidade das partículas cai abaixo de um determinado nível, sua detecção e reconstrução é comprometida.

Contudo, embora existam diversos estudos experimentais neste campo de pesquisa, a análise automática de imagens para identificar e medir parâmetros de escoamentos bifásicos, bem como métodos para classificação de padrões de fluxo, são bastante limitados. Neste artigo, é proposta uma metodologia baseada em métodos *Level Set* e em operações de Morfologia Matemática. As imagens de escoamentos horizontais gás-líquido são capturadas por uma câmera CCD de alta velocidade e então processadas para extração de informações tais como fração de vazio, tamanho e velocidade de bolhas e classificação de padrões.

O artigo está organizado como segue. A próxima seção discute como métodos *Level Set* são aplicados na segmentação de imagens e a Seção 3 detalha a abordagem proposta. Os resultados de segmentação e classificação são discutidos na Seção 4. Por fim, a Seção 5 traz as conclusões e perspectivas de trabalhos futuros.

2. Aplicação de métodos Level Set para segmentação de imagens

Basicamente, técnicas de segmentação visam particionar uma imagem em um conjunto de regiões homogêneas e não-sobrepostas, as quais devem representar algo significativo para uma determinada aplicação. Como constituem a base para diversas tarefas de processamento de imagens, tais como reconhecimento e classificação, é essencial que os resultados obtidos sejam precisos.

A solução baseada no método *Level Set* é motivada pela abordagem de contornos ativos introduzida por Kass, Witkins e Terzopoulos [6]. Para que as curvas se movam em direção às bordas dos objetos de interesse sem ultrapassálas, será introduzido o seguinte termo:

$$g_I(x,y) = \frac{1}{1+|\nabla(G_\sigma * I(x,y))|} \tag{1}$$

em que a expressão $G_\sigma * I$ denota a convolução da imagem I com um núcleo Gaussiano com desvio padrão σ. O termo $\nabla(G_\sigma * I(x,y))$ é essencialmente zero, exceto próximo a variações significativas do gradiente, as quais tipicamente correspondem a bordas da imagem. Desse modo, o filtro $g_I(x,y)$ aproxima-se de um fora das bordas e tende a zero próximo a elas [11].

Considere uma função de velocidade da forma $F = \pm 1 - \epsilon\kappa$, em que ϵ é uma constante que atua como um termo de advecção [9] e κ um termo baseado na curvatura. Observe que, se F for multiplicada pelo termo definido pela Equação 1 acima, ocorrerá um retardamento da evolução do fluxo através das bordas.

O algoritmo básico é descrito a seguir. Um marcador, colocado dentro da região a ser segmentada, evolui de forma a parar nas bordas, caso em que o termo g_I reduz a função de velocidade F para próximo de zero. Tal abordagem tem diferentes aspectos desejáveis, tais como a possibilidade de utilização de mais de um marcador inicial (devido às propriedades topológicas do método *Level Set*, as regiões definidas por tais marcadores irão se fundir), a facilidade de extensão para extrair formas tri-dimensionais e a baixa sensibilidade a ruídos.

Entretanto, em métodos *Level Set* tradicionais, a função ϕ (veja Apêndice A) pode causar *shocks* e a planarização de regiões durante a evolução, conduzindo a erros em cálculos posteriores. Para evitar tais problemas, tal função precisa ser reinicializada periodicamente, o que representa um aumento no custo computacional. A abordagem proposta em [7], utilizada para os testes computacionais deste trabalho, torna desnecessário tal procedimento, e baseia-se em termos de energia interna (que penaliza o desvio da função *Level Set* de uma função de distância sinalizada) e externa (que conduz à propagação do *Level Set* zero ao redor das características de interesse) [10].

3. Abordagem proposta para segmentação de escoamentos bifásicos gás-líquido

A utilização da abordagem *Level Set* para segmentação de imagens requer a escolha de um marcador interno (ou externo) à região de interesse, o que pode ser realizado com operações simples (tais como limiarização e máximos extendidos) para as imagens aqui consideradas.

Além disso, é importante que o gradiente da imagem esteja bem definido, para assegurar que a função de velocidade tenda a zero próximo das bordas. Com este objetivo, as imagens originais foram pré-processadas pelo operador espaço-escala denominado SMMT (*Self-Dual Multiscale Morphological Toggle*). Tal operador possui propriedades interessantes de simplificação que, além de regularizar contornos, minimizam problemas relacionados à variação de iluminação [4]. Este operador é definido abaixo.

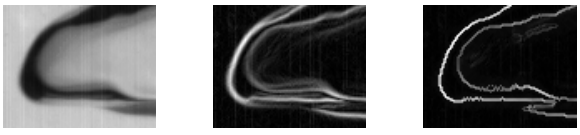
Definição 3.1 *(Operador SMMT) Seja* $\phi_1^k(\boldsymbol{x}) = [\delta_{B_\sigma}(f)]^k(\boldsymbol{x})$*, isto é, a dilatação de* $f(\boldsymbol{x})$ *com a função estruturante dependente de escala* B_σ k *vezes. De forma análoga,* $\phi_2^k(\boldsymbol{x}) = [\varepsilon_{B_\sigma}(f)]^k(\boldsymbol{x})$*, ou seja, a erosão de* $f(\boldsymbol{x})$ *com* B_σ k *vezes [13]. Denomina-se operador SMMT:*

$[\tau_{B_\sigma}(f)]^k(\boldsymbol{x}) =$

$$\begin{cases} \phi_1^k(\boldsymbol{x}), & \text{se } \phi_1^k(\boldsymbol{x}) - f(\boldsymbol{x}) < f(\boldsymbol{x}) - \phi_2^k(\boldsymbol{x}), \\ f(\boldsymbol{x}), & \text{se } \phi_1^k(\boldsymbol{x}) - f(\boldsymbol{x}) = f(\boldsymbol{x}) - \phi_2^k(\boldsymbol{x}), \\ \phi_2^k(\boldsymbol{x}), & \text{caso contrário.} \end{cases} \tag{2}$$

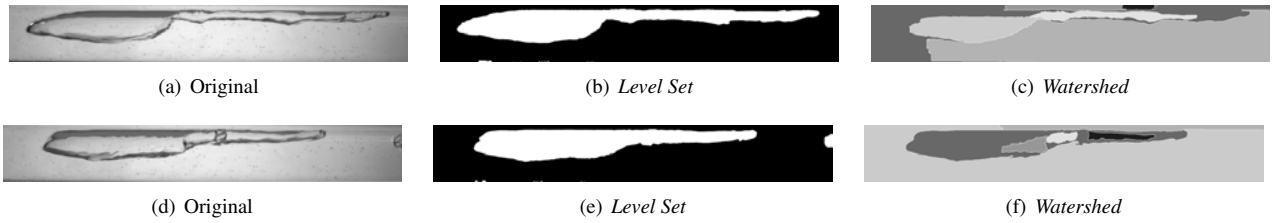
Observe que a regra de decisão analisa, em cada pixel, o valor transformado que mais se aproxima do original. As propriedades espaço-escala garantem que o número de extremos decai com o aumento da escala, diminuindo as características de interesse. Ao aumentar o número de iterações, os extremos criam regiões de influência com níveis de cinza similares [4].

A Figura 1 ilustra a melhoria na qualidade do gradiente após o processamento com o SMMT (escala 0.2 e 5 iterações), o que potencializa o resultado de tarefas de mais alto nível, conforme discutido a seguir.



(a) (b) (c)

Figura 1. Melhoria no gradiente da imagem após aplicar o operador SMMT: (a) imagem original e gradiente (b) da imagem original e (c) da imagem processada pelo SMMT.



(a) Original (b) *Level Set* (c) *Watershed*

(d) Original (e) *Level Set* (f) *Watershed*

Figura 2. Resultados de segmentação.

Abordagem de segmentação	*F-measure*	Revocação	Precisão
Watershed (imagem original)	0.7027	0.9989	0.5420
Watershed (processada pelo SMMT)	0.8023	0.9876	0.6755
Level Set (imagem original)	0.9713	0.9854	0.9576
Level Set (processada pelo SMMT)	0.9896	0.9978	0.9816

Tabela 1. *F-score* ao considerar o segmento que melhor se ajusta ao objeto a ser segmentado.

A metodologia proposta para a segmentação de imagens de escoamentos bifásicos gás-líquido é sumarizada no Algoritmo 1 abaixo:

Algoritmo 1: Esquema para segmentação de imagens de escoamentos bifásicos

1: dada uma imagem de entrada I;
2: aplicar o SMMT em I, gerando uma imagem simplificada, $I1$;
3: selecionar de um marcador (interno à bolha), M, em $I1$ usando técnicas de processamento de imagens;
4: aplicar o método *Level Set* em I1 usando o contorno de M como curva inicial.

4. Resultados experimentais

A seguir, são discutidos os resultados obtidos nos processos de segmentação e classificação de imagens de escoamentos bifásicos gás-líquido.

4.1. Segmentação de imagens de escoamentos bifásicos gás-líquido

A base de testes considerada neste experimento é composta por 40 imagens de escoamentos bifásicos horizontais gás-líquido pertencentes ao padrão pistonado (exemplificado na Figura 2(a) e (d)).

Visando evitar avaliações subjetivas, a análise aqui realizada compara a segmentação resultante da abordagem proposta com aquela obtida manualmente por um especialista (aqui denominada *ground-truth*). Para tal, a medida considerada é a *F-measure*, que consiste na média harmônica entre os índices de precisão, P, e revocação, R [1]:

$$F = \frac{2 \times (P \times R)}{P + R}, \tag{3}$$

em que F varia de 0 a 1 (o valor 1 representa a segmentação perfeita). Em resumo, a precisão indica a porcentagem dos pixels classificados como parte do objeto que são de fato relevantes, enquanto que a revocação representa a proporção do total de pixels pertencentes ao objeto que foram classificados corretamente como tal.

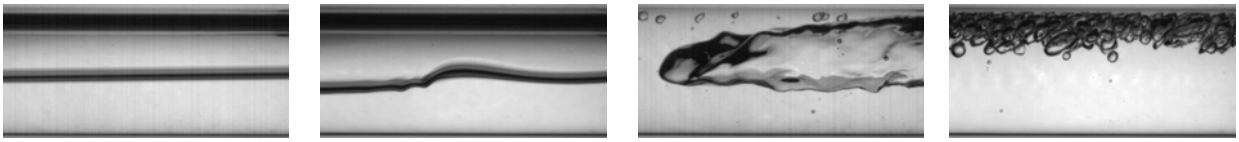
Para avaliar o desempenho da técnica proposta, as imagens também foram segmentadas utilizando a já amplamente conhecida transformada de *watershed* [13], uma abordagem baseada em regiões que possui duas etapas principais: seleção de marcadores das estruturas de interesse e detecção dos contornos mais significativos separando tais marcadores.

Além disso, para analisar o impacto da melhoria da qualidade do gradiente, são comparados os resultados obtidos utilizando como base as imagens originais e também aquelas pré-processadas pelo SMMT.

A Tabela 1 apresenta o índice *F-score* médio ao considerar, para cálculo das medidas de precisão e revocação, apenas os pixels pertencentes à região segmentada que melhor aproxima o objeto de interesse (identificado através da segmentação manual realizada por um especialista). Visando obter uma análise mais robusta, foram extraídos como marcadores máximos estendidos [13] para $h = 1, \ldots, 20$, tanto na imagem original quanto na processada pelo SMMT (para ambas as abordagens). Em ambos os casos, apenas o melhor resultado (representado pela união de segmentos que melhor se sobrepõe ao objeto de interesse) foi considerado para o cálculo do índice

Entrada	Fração Liq. (média)	Fração Vazio (média)	Tamanho (bolha - cm)
Abordagem proposta	0.6066	0.3922	12.9742
Ground truth	0.6023	0.3976	13.0220

Tabela 2. *F-score* ao considerar o segmento que melhor se ajusta ao objeto a ser segmentado.



(a) Estratificado (b) Ondulado (c) Pistonado (d) Borbulhado

Figura 3. Resultados de segmentação.

F-score médio da base. Desta forma, é possível comparar em termos quantitativos a melhor segmentação obtida a partir de cada abordagem.

Observe que a transformada de *watershed* conduz a resultados inferiores, principalmente no que se refere à precisão, indicando uma segmentação em que menos pixels significativos são corretamente classificados. Note também que a melhoria do gradiente após a aplicação do SMMT aumenta o índice para ambas as abordagens.

Em resumo, a transformada de *watershed* possui algumas desvantagens, incluindo o fato de ser mais dependente da precisão dos marcadores e a necessidade de um número maior de regiões (fragmentos) para compor a região de interesse, como ilustram as Figuras 2(c) e (f). A abordagem aqui proposta (Figuras 2(b) e (e)) conduz a resultados satisfatórios a partir de um único fragmento, mas possui um custo computacional maior.

Como mencionado anteriormente, medidas sobre o escoamento são fundamentais para a modelagem de sistemas envolvidos. Para ilustrar a usabilidade da abordagem proposta, a segmentação obtida é utilizada para calcular a fração de vazio e o tamanho da bolha.

Tais medidas também são extraídas das imagens segmentadas manualmente, para fins de análise da exatidão dos resultados. A Tabela 2 apresenta os resultados. As medidas obtidas possuem um erro menor que 1% (o que confirma o índice F-measure próxima de 0.99)

4.2. Classificação de padrões de escoamentos bifásicos

Escoamentos bifásicos gás-líquido, foco deste trabalho, são comumente encontrados em processos industriais que possuem ebulição de líquidos e condensação de vapores. Um exemplo típico acontece em reservatórios, onde petróleo e gás usualmente ocorrem associados nas fases de avaliação e produção dos poços. Neste contexto, onde gás e líquido ocorrem de forma simultânea em um tubo, diferentes padrões de escoamento podem ocorrer dependendo de fatores tais como condições de operação (temperatura e pressão, por exemplo), características do duto, propriedades dos fluidos (incluindo viscosidade e massa específica) e vazões do gás e do líquido. Aqui, são analisados os seguinte padrões de escoamento horizontais [14]:

1. Estratificado (*Stratified Flow*): ocorre quando há uma completa separação das fases devido às baixas velocidades do gás e do líquido. A interface que separa as fases é relativamente lisa e não apresenta distúrbios, sendo que o gás fica na parte superior. A Figura 3(a) ilustra um exemplo.
2. Estratificado ondulado (*Stratified-wavy flow*): com o aumento da velocidade do gás no fluxo estratificado, ondas são formadas na interface. Contudo, as fases permanecem separadas e suas velocidades relativas determinam a amplitude das ondas. Este regime é ilustrado na Figura 3(b).
3. Pistonado (*Slug Flow*): este padrão tem como característica principal a alternância entre bolhas e uma região de líquido entre estas, denominada pistão de líquido. Tal regime intermitente possui bolhas grandes, denominadas bolhas de Taylor. A Figura 3(c) ilustra um exemplo.
4. Borbulhado (*Bubble Flow*): nesse tipo de regime, as bolhas estão dispersas no líquido e a sua maior concentração ocorre na parte superior do tubo (devido à sua flutuabilidade). O padrão é ilustrado na Figura 3(d).

A classificação foi realizada em uma base de 600 imagens (20% foi utilizada para treinamento e 80% para testes) utilizando o algoritmo dos k vizinhos mais próximos para

Classe	Taxa VP	Taxa FP	Precisão	Revocação	F-Measure
Estratificado	0.664	0.008	0.963	0.664	0.786
Ondulado	0.975	0.115	0.739	0.975	0.841
Pistonado	0.992	0	1	0.992	0.996
Borbulhado	1	0	1	1	1
Média	0.908	0.031	0.926	0.908	0.906

Tabela 3. Taxas de classificação.

Classe real	Estratificado	Ondulado	Pistonado	Borbulhado
Estratificado	79	41	0	0
Ondulado	4	116	0	0
Pistonado	0	1	119	0
Borbulhado	0	0	0	120

Tabela 4. Matriz de confusão.

$k = 3$ (optou-se por utilizar um classificador simples devido à característica dos descritores). As imagens da base de treinamento foram manualmente classificadas por um especialista como pertencente à uma das quatro classes acima.

Basicamente, o processo envolveu os seguintes passos: segmentação das imagens da base de teste utilizando a técnica proposta neste trabalho; cálculo de descritores (área, excentricidade, orientação de eixos, tamanho de eixo, área da imagem preenchida (sem buracos) e solidez (proporção dos pixels do fecho convexo que também estão na região de interesse)); e classificação utilizando o algoritmo dos k vizinhos. A taxa de classificação obtida foi de 90.8%.

A Tabela 3 mostra os dados detalhados para cada uma das classes consideradas, em que VP representa taxa de verdadeiros positivos e FP a de falsos positivos (os demais índices já foram discutidos). Em todos os casos, os valores variam de 0 (erro) a 1 (acerto). As classes borbulhado e pistonado apresentam ótimos índices, pois possuem características bastante discriminativas com relação às demais. O padrão ondulado possui mais ocorrências de falsos positivos, o que afeta a taxa de precisão. O mesmo ocorrecomo o índice de verdadeiros positivos do padrão estratificado.

A matriz de confusão (Tabela 4) permite analisar de forma mais detalhada os erros de classificação. Pode-se observar que existem erros entre os padrões ondulado e estratificado, o que é justificado pela semelhança entre os valores de seus descritores. Resultados superiores podem ser obtidos com a utilização de descritores locais que considerem informações tal como a curvatura, por exemplo.

5. Conclusões

Técnicas de processamento de imagens, que permitem analisar informações de escoamentos bifásicos sem perturbar características do fluxo, tornaram-se essenciais em diferentes áreas, atuando inclusive como ferramentas de validação de novos equipamentos.

Neste artigo, é proposta uma abordagem baseada em métodos *Level Set*. Estes possuem inúmeras vantagens, tal como o fato de tratar eficazmente mudanças topológicas e/ou descontinuidades que possam surgir durante a propagação da curva. Os resultados de segmentação obtidos foram muito bons, atingindo taxas de precisão e revocação acima de 98% para determinados padrões.

O pré-processamento da imagem utilizando o operador espaço-escala SMMT simplifica a imagem, conduzindo a uma regularização do contorno que melhora significativamente o gradiente, evitando problemas tais como vazamento, por exemplo. Além disso, tal simplificação favorece uma melhor propagação da curva, exigindo menos iterações para segmentar a região de interesse. Os resultados de classificação também foram satisfatórios mesmo com a utilização de descritores globais.

Trabalhos futuros incluem a utilização de características que permitam discriminar melhor as classes estratificado e ondulado e a inclusão de outros padrões de escoamento.

Apêndice A - Detalhes da abordagem Level Set

Dada uma hipersuperfície $(N-1)$-dimensional fechada, $\Gamma(t=0)$, objetiva-se propor uma formulação Euleriana (ou seja, em que o sistema de coordenadas permanece fixo) para a propagação de $\Gamma(t)$ em direção à sua normal segundo uma função de velocidade, F, a qual pode ser uma função de argumentos tais como curvatura, por exemplo.

A idéia principal do método *Level Set* consiste em incorporar tal interface como o *Level Set* zero de uma função ϕ. Seja $\phi(x, t=0)$, em que x é um ponto no $\mathbb{R}^N$ definido

por [11]:

$$\phi(x, t = 0) = \pm d, \tag{4}$$

em que d é a distância de x para $\Gamma(t = 0)$, e o sinal de mais (menos) é escolhido se o ponto x está fora (dentro) da hipersuperfície inicial $\Gamma(t = 0)$. Desse modo, tem-se uma função inicial $\phi(x, t = 0) : \mathbb{R}^N \rightarrow \mathbb{R}$ com a seguinte propriedade:

$$\Gamma(t = 0) = \{x \mid \phi(x, t = 0) = 0\}. \tag{5}$$

Neste ponto, é preciso definir como ajustar o valor de ϕ para os demais instantes de tempo, de modo a representar corretamente a propagação da curva. Em suma, deve-se definir uma equação para a função $\phi(x, t)$ de tal forma que seu *level set*, $\phi = 0$, contenha a incorporação da propagação de $\Gamma(t)$. Para tal, assuma que $x(t)$ é o caminho de um ponto na interface sendo propagada, ou seja, $x(t = 0)$ é um ponto de $\Gamma(t = 0)$ e $x'(t) \cdot n = F(x(t))$, em que $x'(t)$ é o vetor normal à interface em $x(t)$ e $n = \nabla\phi/|\nabla\phi|$. A exigência de que o *level set* zero da função ϕ deva corresponder à hipersuperfície significa que $\phi(x(t), t) = 0$ [11] . Pela regra da cadeia:

$$\phi_t + \nabla\phi(x(t), t) \cdot x'(t) = 0. \tag{6}$$

Como F fornece a velocidade na direção normal externa, então $x'(t) \cdot n = F$. Obtem-se, assim, uma equação de evolução para ϕ, isto é

$$\phi_t + F|\nabla\phi| = 0, \tag{7}$$

dado $\phi(x, t = 0)$ [11].

Ao inspecionar a Equação 5, constata-se que ela está na forma de uma lei de conservação hiperbólica [8]. Para certas formas de F, esta é uma equação Hamilton-Jabobi padrão.

São diversas as vantagens desta formulação. Inicialmente, como a função $\phi(x, t)$ permanece uma função se F é suave, $\Gamma(t)$ pode mudar sua topologia no processo de propagação. Tal fato também permite que simulações numéricas possam ser desenvolvidas utilizando uma malha discreta no domínio de x, havendo substituição para derivadas espaciais e temporais. Por exemplo, ao usar uma malha uniforme de espaçamento h, nós da malha (i, j), e empregando a notação padrão que ϕ^n_{ij} é a aproximação da solução $\phi(ih, jh, n\Delta t)$, em que Δt é o passo de tempo, pode-se escrever [8][11]:

$$\frac{\phi^{n+1}_{i,j} - \phi^n_{i,j}}{\Delta t} + V|\nabla_{i,j}\phi^n_{i,j}| = 0. \tag{8}$$

Em suma, a Equação 5 foi discretizada por uma aproximação em diferenças finitas progressivas no tempo, e $|\nabla_{i,j}\phi^n_{i,j}|$ representa algum operador de diferenças finitas apropriado para a derivada espacial. Em princípio, o problema poderia ser solucionado calculando-se $\phi^{n+1}_{i,j}$ para cada ponto da malha em função de aproximações em diferenças finitas para a derivada espacial de ϕ. Contudo, esta idéia resulta em instabilidades, fazendo com que o comportamento da curva torne-se não suave no decorrer do processo iterativo. Entretanto, se forem utilizados esquemas hiperbólicos, tais instabilidades são consideravelmente minimizadas e o comportamento da curva estudada permanece suave e homogêneo [8].

Agradecimentos

Este trabalho foi financiado pelo CNPq (pr. nro. 554994/2009-7), ANP e FINEP (PRH-ANP / MCT PRH10 UTFPR), PETROBRAS (nro. 0050.0017859.05.3).

Referências

[1] A. Alpert, M. Galun, R. Basri, and A. Brandt. Image segmentation by probabilistic bottom-up aggregation and cue integration. In *CVIU*, June 2007.

[2] B. Azzopardi. *Gas-liquid Flows*. Begell House, 2006.

[3] T. B. Dinh, B. S. Kim, and T. S. Choi. Application of image processing techniques to air/water two-phase flow. In *SPIE Conf.Series*, pages 725–730, 1999.

[4] L. Dorini. *Transform. de imagens baseadas em Morfologia Matem.* PhD thesis, Unicamp, 2009.

[5] R. Grau and K. Heiskanen. Visual technique for measuring bubble size in flotation machines. *Minerals Engineering*, 15(1):507–513, 2002.

[6] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1987.

[7] C. Li, C. Xu, C. Gui, and M. D. Fox. Level set evolution without re-initialization. In *CVPR*, pages 430 – 436, 2005.

[8] L. M. B. Napolitano. Implementação numérica do método level set para propagação de curvas e superfícies. Master's thesis, Instituto de F?sica da USP, 2004.

[9] S. Osher and R. Fedkiw. *Level Set Meth. and Dynamic Impl. Surf.* Springer, 2003.

[10] T. R. and S. Osher. Level set methods and their applications in image science. *Comm. Math Sci*, 1:623–656, 2003.

[11] J. A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge, 1999.

[12] L. Shi, Z. Zhou, and R. Ren. Parameter measurements of two-phase bubbly flow using digital image processing. In *5th World Congress on Intelligent Control and Automation*, pages 3858 – 3861, 2004.

[13] P. Soille. *Morph. Image Analysis*. Springer-Verlag, 2003.

[14] P. Whalley. *Two-Phase Flow and Heat Transfer*. Oxford University Press, 1996.

[15] X. Yu, D. Xing, T. Hazuku, T. Takamasa, T. Ishimaru, Y. Tanaka, and T. Akiba. Measurement technique for solid-liquid two-phase flow using a normal-line hough transform method. *Journal of Physicsn*, 147(1):81–89, 2009.

[16] X. e. a. Zabulis. Detection of densely dispersed spherical bubbles in digital images based on a template matching technique. *Eng. Aspects*, pages 96–106, 2007.

Software Para Auxílio no Diagnóstico do Glaucoma

Virginia Ortiz Andersson
virginia.andersson@gmail.com
Universidade Federal de Pelotas
Programa de Pós-Graduação em Computação, 96010-610 - Pelotas, RS

Lucas Ferrari de Oliveira
lferrari@ufpr.br
Universidade Federal do Paraná
Programa de Pós-Graduação em Bioinformática, 81520-260 - Curitiba, PR

Resumo

Glaucoma is a disease capable of causing an injury in the optic nerve with progressive impact on the visual field. The diagnosis of the disease is based in the analysis of pathological excavation by inspection of the optic nerve through ophthalmoscopy. This paper shows the development of a CAD (computer aided diagnosis) tool, which calculates the area ratio between pathological excavation and optical disk and the diameter ratios as well. This value is extremely important in glaucoma diagnosis and indicates the excavated optic nerv rate. The software uses region growing as technique for image segmentation. The tests were considered satisfactory: those made with known objects (Brazilian currency coin, R$ 1,00) had relative error less than 0.09 and those performed with retinal images evaluated by specialist did not exceed the relative error (expected) of 0.3. Similar tools exist, but with inaccessible costs to users. The differential is based on greater accessibility that the built software will be in relation to existing ones.

1. Introdução

O glaucoma é uma doença grave, assintomática e é causa irreversível de cegueira. O diagnóstico precoce desta doença, feito pela análise do nervo óptico entre outros parâmetros, é de extrema importância e não está acessível a todas as pessoas devido ao alto custo que este exame pode ter. O nervo óptico é um grosso feixe de fibras nervosas originadas na retina que atravessam o orifício posterior do globo ocular entrando no crânio através do canal óptico. Cada nervo óptico liga-se com o lado oposto formando um cruzamento parcial das fibras chamado quiasma óptico [7]. O disco óptico (DO) designa a porção intraocular do nervo óptico visível à oftalmoscopia. A artéria e veias centrais da retina atravessam o disco e se bifurcam em sua superfície. Dentre os exames utilizados no diagnóstico do glaucoma, a análise do nervo óptico se baseia principalmente na obtenção da porcentagem de nervo óptico escavado. Esse valor é obtido tanto pela razão entre os diâmetros ou entre as áreas de escavação por disco óptico total [1]. A Figura 1 mostra duas imagens de fundo de olho apresentando discos ópticos de diferentes configurações: Em (a) escavação da papila óptica normal, moderada, ocupando no máximo 30% do disco e em (b) escavação no glaucoma avançado ocupando 80% a 90% do disco óptico. Na Figura 2 pode-se observar as regiões D e C marcadas, a razão é calculada pela fórmula C/D e é conhecida como *"Cup-to-Disc ratio"*, ou razão escavação por disco, abreviada como CDR ou E/DO, é a divisão entre os limites da escavação pelos limites do disco óptico [1].

O algoritmo escolhido para a segmentação das imagens neste projeto foi o de crescimento de região que será discutido na próxima Seção. Esse método de segmentação ainda é muito utilizado, normalmente em conjunto com outras técnicas de segmentação, como no trabalho de Jun Tang [8]. Ele propôs em seu trabalho o uso do algoritmo de *Watershed* na base do tradicional algoritmo: Onde antes as sementes para o crescimento de região eram escolhidas manualmente agora são geradas pelas regiões resultantes do algoritmo de Watershed. O uso combinado desses métodos resulta em um método cuja complexidade de computação diminui no que diz respeito ao crescimento de região original, porém utiliza mais um algoritmo. Onde antes era dependente do número de pixels da imagem, agora depende do número de regiões geradas pelo algoritmo de *Watershed*. Além disso, essa proposta automatiza o que seria um trabalho do usuário, na seleção das sementes [8]. A Figura 3 mostra o fluxograma do algoritmo de crescimento de região

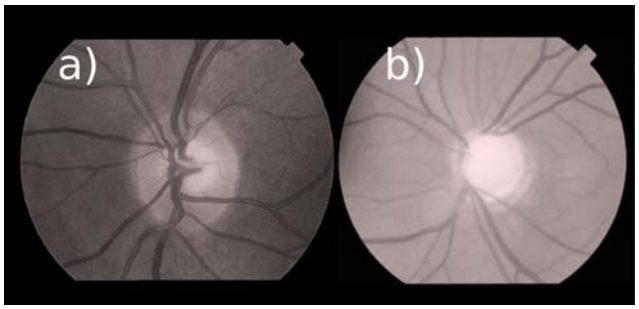


Figura 1. Em (a) uma Retinografia apresentando DO com escavação considerada normal e em (b) escavação em glaucoma avançado [9]

proposto por Tang, exibindo os métodos utilizados.

O propósito do trabalho foi desenvolver um software que auxilie no diagnóstico da neuropatia óptica glaucomatosa (danos no nervo óptico causados pelo glaucoma). Intitulado OnScope (*Optical Nerve Scope*), sua colaboração se dá no cálculo da razão entre áreas e diâmetros significativos do nervo óptico que fornecem ao especialista uma parcela importante no diagnóstico da doença. Esses valores são armazenados para acompanhamento da doença do paciente e seu histórico é de grande valor. O software fornece uma maneira de armazenar esses dados que auxiliam o oftalmologista.

O artigo está dividido nas seguintes Seções: A próxima Seção apresenta materiais e métodos utilizados no trabalho; a Seção 3, testes e validação do software; Seções 4 e 5, resultados e discussão sobre os mesmos e por fim, na Seção 6, uma conclusão sobre o trabalho realizado.

2. Materiais e Métodos

A proposta do software OnScope é calcular duas áreas e diâmetros em imagens de retinografia e apresentar uma razão entre elas para a utilização no diagnóstico clínico do glaucoma.

Inúmeras vezes, especialistas da área utilizam ferramentas como Adobe Photoshop [6] e GIMP [5] para elaborarem seus diagnósticos. Esses softwares não são adequados para essa finalidade, já que não foram desenvolvidos com o propósito de auxiliar na investigação de doenças. O uso de tais recursos pode despender boa parte do tempo dedicado a elaboração do diagnóstico, pois os resultados provenientes dos softwares devem ser anotados manualmente para sua utilização.

Optou-se por utilizar um algoritmo de crescimento de região para segmentação das imagens no nosso software,

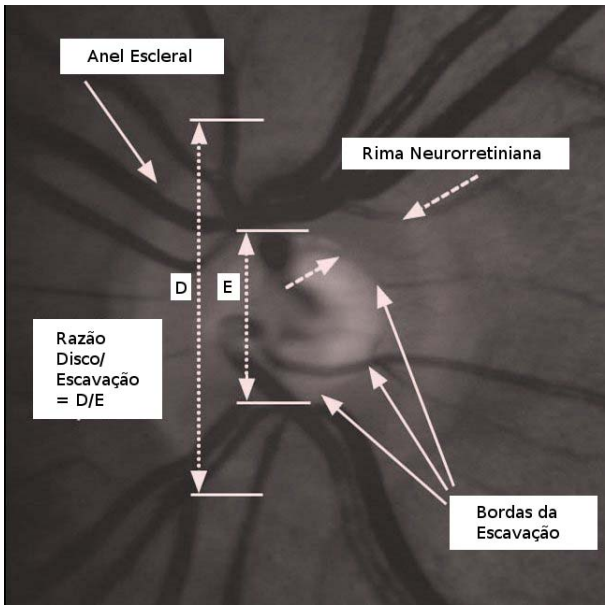


Figura 2. Marcação das regiões e suas relações entre o cálculo da razão entre elas [2]

pois o mesmo funciona para o dado problema e é similar a algoritmos encontrados em funções de seleção de regiões nos softwares citados e com isso, a adaptação do especialista pode ser mais fácil.

O processo de crescimento de região inicia com um conjunto de pixels chamados sementes, que devem crescer originando regiões uniformes e conectadas. Um pixel é adicionado à região se ele ainda não foi adicionado à outra, se ele é vizinho da mesma e se após sua adição, a nova região continua uniforme. Assumimos nesse algoritmo, oito pixels vizinhos ao redor do pixel atualmente analisado (*8-connectivity*) [4].

Cada pixel, ao ser adicionado à região, passa por um teste de uniformidade dado pela Equação 1 [4].

$$P(R) = \begin{cases} Verdadeiro & \text{se } |f(x,y) - \mu_R| \leq \Delta; \\ Falso & \text{caso contrário.} \end{cases} \quad (1)$$

assumindo que f(x,y) é o valor do pixel na posição x, y da imagem no espaço RGB, μ_R a cor média da região e Δ um limite escolhido. Para cada pixel avaliado no momento é calculado o valor f(x,y) e a diferença deste com o valor médio de cor de toda a região. Caso ela seja menor ou igual a um limite específico, o novo pixel é associado à região, caso contrário, ele é descartado. O algoritmo para

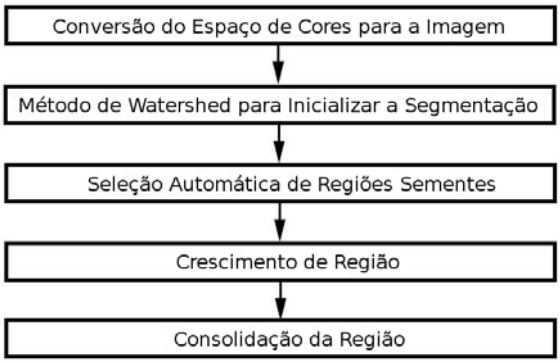


Figura 3. Fluxograma do algoritmo de Tang [8]

quando não encontra mais pixels que satisfaçam os critérios de adição para cada região [4].

A Figura 4 mostra, como é aplicado o algoritmo no que seria uma imagem muito simples. Neste exemplo é considerada uma vizinhança de oito pixels. Para cada iteração do algoritmo, é verificado se o valor do pixel é menor que um limite igual a 3. Como resultado, apenas pixels com valores 0, 1 ou 2 são adicionados.

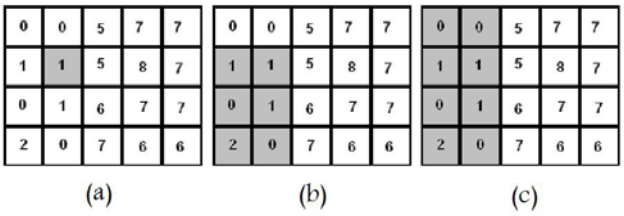


Figura 4. (a) Pixel semente. (b) Primeira iteração. (c) Iteração final [4]

No software, o especialista escolhe algumas sementes, na região que deseja seccionar e o limite Δ. Os diâmetros são calculados através da diferença entre dois pontos, também escolhidos pelo usuário.

Foram realizadas comparações entre nosso software, o software Adobe Photoshop e um diagnóstico clínico realizado por um especialista. O comparativo serviu para mostrar a semelhança entre os resultados e que em contrapartida, nossa ferramenta é específica para a detecção precoce do glaucoma auxiliando no armazenamento das informações relativas ao diagnóstico.

3. Testes e Validação do Software

Os testes e validação da ferramenta OnScope foram compostos de duas partes: A verificação da capacidade do software de calcular a razão entre duas áreas usando objetos com medidas e resultados conhecidos; e a verificação da capacidade do software ser usado no diagnóstico clínico. Para isso, os resultados foram comparados com resultados fornecidos pelo especialista.

3.1. Verificação com objetos de medidas e resultados conhecidos:

Cada paciente possui um tipo diferente de fundo de olho. Cada nervo óptico e escavação possuem diferentes medidas. Devido à essa falta de padrão nas medidas de nervos ópticos e escavações encontradas em indivíduos aleatórios, tornava-se impossível verificar se o software OnScope calculava corretamente a razão entre duas áreas em questão.

Para contornar esse problema, foi necessário escolher algum objeto que possuísse duas regiões distintas, cujo valor das áreas e a razão entre elas fosse conhecida ou facilmente verificável de maneira experimental (sem uso de software). Para esta finalidade foi escolhido uma moeda brasileira de 1 Real.

Os valores das áreas da moeda de 1 Real foram calculados utilizando uma régua milimetrada e Π = 3,14159265. Segundo o Banco Central do Brasil [3], o diâmetro do maior círculo mede 2,7 cm e o diâmetro do menor circulo (interno) mede 1,8 cm. Os resultados do cálculo manual pode ser visto na Tabela 1. Também foram realizadas 20 segmentações e medidas na moeda de 1 Real utilizando o softwares OnScope e logo após calculou-se a média e o desvio padrão dos resultados a fim de se obter um valor que correspondesse melhor ao desempenho do software, como pode ser visto na Tabela 2.

Área Maior (cm^2)	Área Menor (cm^2)	Razão das Áreas	Razão dos Diâmetros
5,7	2,5	0,4	0,6

Tabela 1. Valores reais da moeda de 1 Real

3.2. Caso Clínico

As diferenças na anatomia dos pacientes impediram a utilização de imagens de retinografia na etapa de verificação do cálculo da razão entre as regiões. Porém, foi necessário avaliar se as razões entre as áreas são consistentes para auxiliar no diagnóstico. Para isso, foram submetidas ao teste

Figura 5. Teste e validação utilizando moeda de 1 Real

algumas imagens de fundo de olho com os limites de nervo óptico e escavação bem definidos. O oftalmologista calculou os CDR's utilizando o software PhotoShop, e a razão entre as áreas foi calculada de maneira qualitativa, ou seja, valores atribuídos pelo oftalmologista, com base em sua experiência no diagnóstico do glaucoma. O software OnScope RG também calculou as áreas e suas razões e os CDR[V]s e CDR[H]s (razão de diâmetros vertical e horizontal) nas imagens.

Posteriormente, as mesmas imagens foram segmentadas no software Photoshop, utilizando a ferramenta de seleção por gama de cores. Essa etapa serve para compararmos os resultados obtidos na segmentação realizada pelo software Onscope com outra ferramenta que é utilizada em casos clínicos, mas não é específica para isso.

As Figuras 6 e 7 mostram o resultado de um caso segmentado e armazenado pelo software OnScope.

Teste	Razão do Diâmentro	Razão da Área
1	0,69	0,41
2	0,68	0,47
3	0,68	0,42
4	0,67	0,43
5	0,68	0,43
6	0,66	0,45
7	0,68	0,4
8	0,69	0,43
9	0,69	0,43
10	0,68	0,41
11	0,67	0,43
12	0,66	0,39
13	0,69	0,45
14	0,68	0,42
15	0,67	0,42
16	0,69	0,42
17	0,69	0,43
18	0,71	0,38
19	0,69	0,41
20	0,68	0,44
Média	0,68	0,42
Desvio Padrão	0,01	0,02

Tabela 2. Segmentações e medidas realizadas na moeda de 1 Real utilizando o software OnScope

4. Resultados

Um paralelo entre as razões das áreas das imagens de retinografia obtidas pelos softwares pode ser visto na Tabela 4, juntamente com os resultados fornecidos pelo especialista. O resultado fornecido pelo oftalmologista é um valor atribuído de acordo com a experiência do mesmo em avaliar o tamanho das escavações em relação ao disco óptico total nos pacientes.

Para averiguar a correlação entre os resultados, foram calculados os coeficientes de correlação de Pearson para os valores obtidos com o software OnScope e com o software Photoshop e posteriormente com os resultados do diagnóstico clínico, como mostra a Tabela 5.

5. Discussão

Neste trabalho mostramos a montagem de um sistema de auxílio ao diagnóstico do glaucoma e não foi encontrado na literatura científica nenhum trabalho que se tenha desenvolvido tal ferramenta. Um trabalho que pode ser implemen-

	OnScope	Photoshop
Média diâmetro	0,68	0,67
Desvio padrão diâmetro	0,01	0,01
Média área	0,42	0,43
Desvio padrão área	0,02	0,02

Tabela 3. Média e desvio padrão das razões do diâmetro e área da moeda calculados com o uso dos softwares

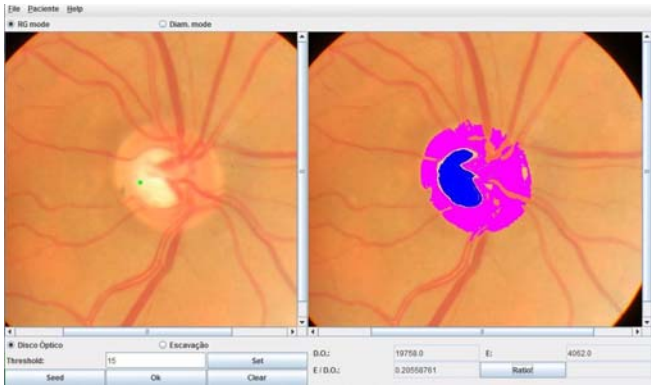


Figura 6. OnScope na modalidade de cálculo de áreas

tado como uma melhora do nosso sistema é o trabalho de Tang [8]. Como o especialista gasta muito tempo tabelando as informações nos softwares que não são próprios para isso e precisam realizar cálculos com os valores, a grande contribuição do nosso sistema é o cálculo da CDR otimizando o trabalho, bem como a tabulação dos dados clínicos.

Através do coeficiente de correlação de Pearson, podemos observar que existe uma forte correlação entre os resultados do OnScope e os do Phostoshop, bem como de ambos os softwares com os resultados fornecidos pelo especialista. Considerando que a correlação perfeita positiva entre duas variáveis é igual a 1, nosso software obteve 0.95 no ρ de Pearson mostrando a forte correlação existente entre as duas variáveis consideradas.

Se compararmos os resultados paralelamente, na Tabela 4 notamos uma diferença importante entre as razões do teste número 5. Isso se deve ao fato que o algoritmo de crescimento de região usado no software OnScope possibilita ao usuário uma maior liberadade na escolha das regiões, enquanto que a técnica de seleção usada no Photoshop restringe o usuário em uma determinada área contínua. Em contrapartida, algumas segmentações realizadas no Pho-

Figura 7. Dados do exame salvos no OnScope

Teste	OnScope	Photoshop	Oftalmologista
1	0.20	0.32	0.2
2	0.35	0.24	0.3
3	0.42	0.47	0.5
4	0.23	0.20	0.3
5	0.92	0.45	0.8
6	0.12	0.10	0.3
7	0.42	0.36	0.4
8	0.18	0.31	0.1
9	0.94	1.17	0.9

Tabela 4. Razões entre áreas obtidas nos softwares e no diagnóstico clínico, para as imagens de retinografia

toshop forneceram resultados mais satisfatórios para determinados tipos de imagem, onde as áreas de escavação e disco eram bem definidas.

As comparações entre os resultados fornecidos pelo especialista revelam que ambos possuem semelhanças. Porém, o OnScope leva vantagem nos testes realizados. Novos casos já estão sendo separados para futuros testes para que a validação seja mais completa e conclusiva. O software OnScope foi criado com o propósito de facilitar a mensuração de distâncias e a segmentação de imagens de retinografia por oftalmologistas. Oferece ao seu usuário, também, a possibilidade de armazenar os resultados para posterior consulta, condição importante para o acompanhamento da doença no paciente.

As Figuras 6 e 7 mostram um caso de fundo de olho

OnScope e Photoshop	0.78
Photoshop e Oftalmologista	0.78
OnScope e Oftalmologista	0.95

Tabela 5. Correlação de Pearson entre valores obtidos com OnScope, Photoshop e no diagnóstico clínico, para as imagens de retinografia

segmentado pelo software OnScope e sua imagem resultante, e as informações salvas para futuras consultas. A ferramenta apresenta ao médico o resultado das áreas calculadas, dos diâmetros e as razões importantes para o seu diagnóstico e salva todos esses dados juntamente com a imagem do exame, segmentada. A cor azul representa a porção de escavação encontrada, enquanto que a magenta designa a área representada pelo disco óptico total. O especialista pode refazer a segmentação escolhendo um novo limite de diferença entre as cores, para imagens de retinografia pouco nítidas ou com coloração muito semelhante nas áreas significativas.

6. Conclusão

O software OnScope funciona melhor para segmentações não-continuas, que exigem maior liberdade de escolha entre as regiões, enquanto que a segmentação realizada pelo Adobe Photoshop funciona bem para regiões bem definidas. A variação das estruturas do olho interferiram na escolha do algoritmo a ser utilizado, pois os nervos ópticos e suas escavações não possuem formatos uniformes e bem formados. Logo, o uso do software OnScope pode auxiliar na segmentação das áreas significativas nesses casos.

Referências

[1] G. V. d. Almeida. *Manual de semiologia do glaucoma*. Phoenix, São Paulo, 2004.

[2] R. R. Bourne. The optic nerve head in glaucoma. *Community Eye Health J*, 19(59):44–45, 2006.

[3] B. C. do Brasil. Disponível em:. http://www.bcb.gov.br/?MOEDAFAM2, 2011.

[4] N. Efford. *Digital Image Processing - a pratical introduction using Java*. Pearson Education Limited, Leeds (UK), 2004.

[5] S. Kimball and P. Mattis. Gimp - gnu image manipulation program. http://www.gimp.org/, 1996-2010.

[6] T. Knoll and J. Knoll. Adobe photoshop. http://www.photoshop.com/, 1987–2011.

[7] A. B. Machado. *Neuroanatomia funcional*. Livraria Atheneu, São Paulo, 1983.

[8] J. Tang. A color image segmentation algorithm based on region growing. *2nd International Conference on Computer Engineering and Technology*, 6:634–637, 16-18 April 2010.

[9] I. M. Tavares and P. A. d. Arruda Mello. Glaucoma de pressão normal. *Arquivos Brasileiros de Oftalmologia*, 68(4), July/Aug 2005.

Focus of Expansion Estimation for Motion Segmentation from a Single Camera

José Rosa Kuiaski, André Eugênio Lazzaretti and Hugo Vieira Neto
Graduate School of Electrical Engineering and Applied Computer Science
Federal University of Technology - Paraná, Brazil
jose.kuiaski@yahoo.com.br, andrelzt@yahoo.com.br, hvieir@utfpr.edu.br

Abstract

This paper proposes a new approach to motion segmentation from video sequences acquired using a single camera, whose aim is to identify which components are due to pure egomotion *and which components are due to* independent *moving objects within the observed motion field. The model has three main steps, namely computation of the optical flow field, estimation of the focus of expansion and classification under constraints. Preliminary results show that estimation of the focus of expansion followed by simple clustering techniques is promising for the achievement of motion segmentation when using a single camera with no* a priori *information about the environment.*

1. Introduction

Animal survival relies on the ability to perceive movements, distinguishing the static environment from moving agents, so that the animal can act appropriately. As the majority of objects in real environments is usually stationary, biological visual systems became sensitive to movement events and the environment itself can be considered stationary [14]. This ability is also very useful for artificial agents aiming to operate in dynamic environments.

Reproducing the ability of motion segmentation computationally is a complex task, which involves defining decision rules and handling noise. The computational problem is even more complicated when there is no *a priori* information about the *egomotion* (i.e. camera motion) or the behaviour of moving objects in the environment, which is almost always the case.

According to Gibson's Ecological Theory of Perception [4], the human visual system is able to extract all the necessary environmental information from the optical flow field at the viewer's retina. In his theory, Gibson states that the environmental optical information converges to a single point in space, forming the *Dynamic Ambient Optical Array*. Although this may be a controversial statement, it will be considered here when dealing computationally with movement analysis.

Let the focus of expansion be the representation of the *Dynamic Ambient Optical Array* on the 2D image projection of the environment, namely a single point in space where all optical flow vectors due to camera movement (image background optical flow vectors) intersect. So, estimating the focus of expansion may provide, according to Gibson's theory, all the *egomotion* information necessary to motion segmentation.

In a computational vision system, the information about movement is obtained from computing the optical flow field, which can be obtained in several ways, such as in Horn and Schunk [6], Lucas and Kanade [8], Shi and Tomasi [16], Nordberg and Farneback [13] or Fleet and Weiss [3]. Therefore, it may be possible to extract all – or almost all – information about the environment from the calculated optical flow field.

Video sequences acquired with stereo cameras simulate similar conditions as the ones present in the human visual system, but the availability of more than a single camera may not always be possible. With a single camera, depth information without *a priori* information is not available, but a possible solution for this problem is to model movement from image motion, rather than from object motion.

In this paper we compute image motion in the form of optical flow vectors obtained from a single camera and make some assumptions due to the unavailability of *a priori* information about *egomotion*:

1. The majority of optical flow vectors within the optical flow field are relative to *egomotion*, so the majority of pixels within a frame of a video sequence are related to the static environment (background).
2. The environment optical flow vectors theoretically converge to a single point in space – the *Dynamic Ambient Optical Array* – from now on called the focus of expansion (FoE) point.

3. The presence of independent moving objects and also noise in the optical flow computation will generate multiple additional FoE candidate locations, corresponding to different moving groups of pixels (objects).

The first step to consider is computing the optical flow field. Our current model is based on the algorithm by Nordberg and Farneback [13] for dense optical flow computation – in section 2, considerations about the use of dense or sparse algorithms will be made. Once the optical flow field is available, we estimate FoE locations for the optical flow vectors – this is a key step to our motion analysis approach and will be detailed in section 3. The proposed idea for the estimation of FoE points is to find the intersections of every pair of optical flow vectors within a pair of frames from a video sequence. Clustering of the resulting intersections and a decision rule are then necessary to determine the FoE location for the environment, which corresponds to *egomotion*, and for any moving objects that may be present in the scene.

2. Optical Flow Estimation

The image registration technique proposed by Lucas and Kanade [8] shows that it is possible to use the spatial intensity gradient to optimise the search for the position that yields the best match between a pair of images. This is particularly important when dealing with optical flow. In motion analysis, the optical flow field is the distribution of motion vectors, whose magnitudes mean the apparent velocity and whose directions indicate the relative trajectory between moving objects and the viewer.

Let us first consider image registration techniques as the mathematical foundation for optical flow understanding when dealing with two consecutive frames of a video sequence. Each pixel at the first frame possibly corresponds to a specific pixel at the second frame, as shown in figures 1 and 2. Figure 1 shows two consecutive frames of the optical_flow_input.avi video sequence, in which a moving car is followed by the camera [17], while figure 2 shows the corresponding optical flow field computed by the algorithm by Lucas and Kanade (the first frame is the reference for the optical flow vectors). Each optical flow vector shows the estimated movement for its corresponding underlying pixel.

In figure 2 one can clearly notice that the vast majority of optical flow vectors are due to the camera translation to the right, which results in apparent background image translation to the left.

One of the assumptions in computation of optical flow is that the brightness of an image feature does not change between frames, i.e. if $E(x, y, t)$ is the brightness of a given pixel at time t, its first partial derivative in time should be zero, as follows:

Figure 1. Two consecutive frames of the optical_flow_input.avi video sequence in which a moving car is followed by the camera [17].

$$\frac{\partial E}{\partial t} = 0. \tag{1}$$

Equation 1 can be rewritten as a function of the brightness gradient, leading to equation 2:

$$\frac{\partial E}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial E}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial E}{\partial t} = \nabla E^T \cdot \vec{v} = 0, \tag{2}$$

where $\vec{v} = (v_x, v_y, 1) = \left(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t}, 1\right)$ is the motion vector for a single point or region Ω within an image and ∇E^T is the spatio-temporal brightness gradient of the same region.

According to Nordberg and Farneback [13], there is no unique solution for $\vec{v}$ in equation 2. A possible solution for this equation would involve a constraint, say $\nabla E = 0$, and considering a region Ω where $\vec{v}$ could be assumed constant. Therefore, the local mean value over Ω carries all the spatio-temporal information of the object's orientation, as follows:

$$\left[\int_{\Omega} p(x)\left(\nabla E_x\right)\left(\nabla E_x\right)^T dx\right] \cdot \vec{v} = 0, \tag{3}$$

where p is a Gaussian function and the first term in the multiplication is a second-moment matrix derived from ∇E, the so-called *structure tensor*.

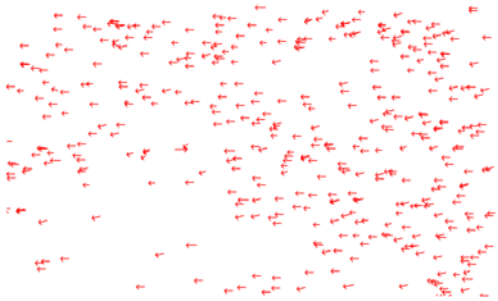


Figure 2. Optical flow field from matching points of the frames in figure 1 computed with the algorithm by Lucas and Kanade.

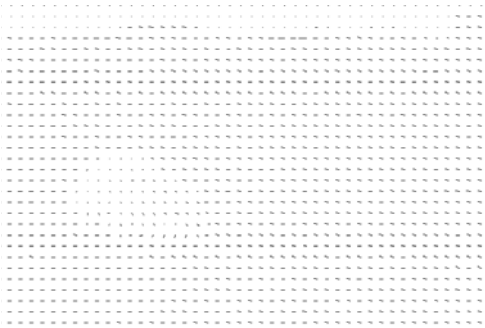


Figure 3. Dense optical flow field for the frames in figure 1 computed with the algorithm by Nordberg and Farneback.

Horn and Schunk [6] also suggest an additional smoothness constraint in which $\nabla^2\left(\frac{\partial x}{\partial t}\right) = \nabla^2 v_x$ and $\nabla^2\left(\frac{\partial y}{\partial t}\right) = \nabla^2 v_y$ (the Laplacians of velocity components) must be minimised.

2.1. Dense Optical Flow

Horn and Schunk's algorithm and also Nordberg and Farneback's algorithm consider an $m \times n$ window within the image to estimate the optical flow at its central point. Therefore, a uniformly sampled grid of points with their respective optical flow vectors is obtained. Figure 3 shows such a dense optical flow field, in which each vector connects the grid point to the calculated position for that same point in the consecutive frame. The use of dense optical flow in figure 3 clearly shows that there are noticeable differences in motion between object and background.

Dense optical flow may be more sensitive to the aperture problem, i.e. more prone to select edges from straight line segments in the input video sequence, which will result in the computation of ambiguous motion. Also, rigid moving objects that are larger than the considered analysis window may reflect in noisy components in the optical flow field, requiring a multiscale approach do be dealt with. However, the use of dense optical flow favours the analysis of optical flow vectors as stochastic processes. The static locations of flow vectors obtained in the dense approach do not depend on any tracking algorithm and ensure their behaviour as random variables for the stochastic process at time t.

The experiments reported in this paper were based on dense optical flow in order to model motion as a stochastic process, making it possible to use Independent Component Analysis [7] for motion classification and segmentation in future work, when we plan to work with optical flow vectors from two different observation points – right and left frames acquired from a stereo vision system, for example. The technique that involves estimation of the focus of expansion reported here is a preliminary approach to the problem.

2.2. Sparse Optical Flow

The most noticeable difference when dealing with sparse optical flow is the use of interest point detectors before computing the optical flow field. Figure 2 already showed an example of an sparse optical flow field in which the Harris interest point detector [5] was used to select which pixels within the image are those whose information is more relevant, and the optical flow field was computed using a multiscale version of Lucas and Kanade's algorithm [2]. This approach ensures greater robustness to the aperture problem, reducing noisy components. The sparse optical flow computation approach reduces wrong estimations of flow vectors when frames from the video sequence under analysis have a relevant proportion of homogeneous regions.

3. Focus of Expansion Estimation

The focus of expansion (FoE) corresponds to Gibson's *Dynamic Ambient Optical Array* [4], which is a single point in space where all the flow vectors should converge. Its main uses in vision applications are the estimation of the time-to-impact (TTI) in visual navigation and the 3D reconstruction of the environment. There are three usual approaches to FoE estimation in the literature [12]: discrete, differential and least-squares, all of them yielding good results for pure translational motion.

This papers adopts the differential approach, since our aim is to estimate the focus of expansion from the optical flow field. The differential approach is more robust, but

computationally heavy, as a consequence of not performing any least-squares minimisation [15].

The focus of expansion estimation plays an important role in motion segmentation. Our hypothesis is that it is possible to classify all optical flow vectors within the optical flow field as belonging to the background (environment) or to independent moving objects, from the estimated FoE, using a simple rule:

> "If a flow vector belongs to the environment, its backwards line segment should cross the FoE."

In this way, it seems to be possible to estimate all *egomotion* components for pure translation. Considering the general assumption that the environment is static and also assuming that the majority of pixels within an image frame represents the environment, it is possible to argue that the majority of flow vectors from a dense optical flow field are *egomotion* components due to camera translation. From these assumptions, it is possible to state the following:

1. For every pair of optical flow vectors, there may be an associated FoE. If there is not, one of the optical flow vectors is related to an independent moving object within the image frame.
2. FoE candidates originated from two environmental optical flow vectors will be concentrated within an small area in space, whose variational width will be a function of noise on the optical flow field computation due to numerical approximations.
3. FoE candidates originated from mixed optical flow vectors (one from the environment and one from an independent moving object), or from two optical flow vectors from independent moving objects, will result in sparsely and randomly distributed locations in space.

Therefore, it is assumed that the great majority of calculated FoE candidates will lie within a small area in space. A simple histogram analysis could inform precisely where the largest concentration of candidates lies. However, FoE candidates can be distributed across the infinite 2D space and computing such a histogram would be computationally infeasible. Therefore, clustering the resulting FoE candidates is necessary. Using a clustering approach, one can infer that the centroid of the cluster with the largest number of samples is the most likely FoE for the environment.

3.1. Naïve Estimation

A first approach to the focus of expansion estimation is to naïvely consider some *a priori* information about the optical flow vectors. If at least three of them are known to be from the environment, they can be simply triangulated into a small region and their average location will be considered the FoE. This approach works fine for simple video sequences with pure translational motion, but relies on *a priori* information.

3.2. Clustering by K-means

If no *a priori* information should be used, a possible solution is to perform clustering on the estimated FoE array. The K-means technique [9] can be used to split an array with n samples $(X_0, X_1, \ldots, X_n)$ into m clusters $(C_0, C_1, \ldots, C_m)$ whose total squared energy is minimal, as follows:

$$\text{Energy} = \arg\min \sum_{i=1}^{m} \sum_{\mathbf{X}_j \in \mathbf{C}_i} \parallel \mathbf{X}_j - \mu_i \parallel^2, \qquad (4)$$

where $\mathbf{X}_j$ is the jth sample and $\mathbf{C}_i$ is the ith cluster with centroid μ_i.

The most noticeable disadvantage in using K-means to cluster the FoE candidates is the need of *a priori* information about the number of clusters m. In most general cases, this information is not available.

4. Experimental Setup

After dense optical flow field computation, the first step is to count the total number of optical flow vectors that are parallel to the x or y axes. If the majority of optical flow vectors is parallel to one of these axes, it is inferred that the FoE lies in the positive or negative infinity for x or y, depending on the case. The second step is to count the number of parallel vectors lying in the same direction. This step informs whether the FoE lies at the positive or negative infinity for x or y. These two steps can be merged into a single step if a directional angle histogram is computed and the majority of directional angles lies within a small region of the histogram. However, this procedure is computationally more expensive than the previous two separate steps.

In both cases, we considered that two vectors are still considered parallel if they differ by a small value ϵ from each other. This approximation compensates for numerical errors in the optical flow field computation and yields good approximations for FoE candidates that lie at infinity. If all cases above do not apply, this implies that the FoE is at a finite location in 2D space and this is the case in which FoE estimation must take place. The model was implemented using OpenCV and the experiments were conducted using the optical_flow_input.avi video sequence [17]. Ground-truth data was obtained by counting correlated and uncorrelated optical flow vectors frame by frame.

5. Results

The naïve estimation used locations known to be background *a priori*, namely the coordinates (160, 160), (320, 560) and (320, 160). For the first 50 frames of the optical_flow_input.avi video sequence, the average error rate for the background was 13.8% and the average error rate for objects was as high as 90%. This large average error for objects led us to notice that if there are homogeneous regions in the frame (such as the sky in the optical_flow_input.avi video sequence), there will be no optical flow within these regions.

Since the approach being used was naïve, it was possible to deal with the sky region by labelling it *a priori* as background. By doing this, the average error rate for the background went down to 8.7% and the average error rate for objects went down to about 74%. A more careful observation showed that the optical_flow_input.avi video sequence yielded significantly small optical flow vectors in the majority of its frames – a consequence of having a camera following the car at low speed – which seemed to be the main reason for such a large average error rate for the object. Using another video sequence, Office_left.avi [11], with less camera movement and much more significant object translation – therefore resulting in larger optical flow vectors – yielded an average error rate for the background of 12% and an average error rate for the object of 22.4%, confirming our hypothesis.

Finally, the K-means clustering was applied to the optical_flow_input.avi video sequence. Since there is no *a priori* information about the spatial location of the FoE for every pair of optical flow vectors, the number of clusters had to be determined empirically and was set to seven. Figure 4 shows the spatial distribution of the estimated FoE candidates after clustering, where the x and y axis represent spatial coordinates in pixels.

In figure 4, the location most likely to be the FoE is the centroid of the cluster with the largest sample density. In this case, the FoE was considered to lie in coordinates (0, 250000), which is the centroid of the yellow cluster. In our implementation, this result corresponds to an FoE at positive infinity in x, which is coherent with the motion observed in the corresponding video sequence. The K-means clustering approach yielded an error rate of 5.9% for the background and 80% for objects, not dealing with the homogeneous sky region problem. After filtering the sky region using *a priori* information, the error rate for the background dropped to 5% and the error rate for objects dropped to 30%.

No K-means clustering was necessary in the experiments with the Office_left.avi video, due to the *a priori* information that no background movement, i.e. camera movement, was present.

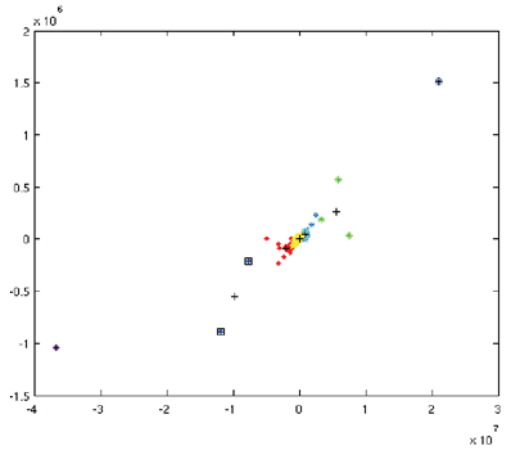


Figure 4. Spatial distribution of FoE candidates. The cluster with highest sample density lies in coordinates (0, 250000), which was considered the FoE.

6. Conclusions

The preliminary results shown in this paper, despite being beyond a reasonable error tolerance, show that it is possible to classify image motion through focus of expansion estimation. The large average error rates are attributed to the simplistic approaches for motion segmentation that were used (naïve estimation and K-means clustering). There seems to be great potential of use of the method in situations where a second camera or information about the camera's trajectory and speed is not available *a priori*.

Other issues, such as the interference of homogeneous regions, lack of optical flow vectors of significant magnitude and presence of noise in the optical flow computation suggest the use of adaptive clustering algorithms. Possible adaptive clustering algorithms to be used in future investigations include the GWR Neural Network [10] and Support Vector Machines [1], avoiding the need of *a priori* definition of the number of clusters.

The use of the approach presented here is intended to deal with pure camera translation and zooming when there are much more optical flow vectors due to *egomotion* than to object motion. Dealing with rotations would require the estimation of a focus of rotation in addition to the estimation of the focus of expansion. It should be noticed that the current approach is not well suited for real-time applications, as the amount of calculations slows down considerably the input video sequences. We expect that the use of neural network approaches for clustering will result in more robust classification for the determination of the focus of expansion.

References

[1] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik. Support vector clustering. *Journal of Machine Learning Research*, 2:125–137, December 2001.

[2] J.-Y. Bouguet. *Pyramidal Implementation of the Lucas Kanade Feature Tracker: Description of the Algorithm*. Intel Corporation, Microprocessor Research Labs, 2002.

[3] D. J. Fleet and Y. Weiss. *The Handbook of Mathematical Models in Computer Vision*, chapter Optical Flow Estimation, pages 239–258. Springer, 2005.

[4] J. J. Gibson. *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, Hillsdale, 1986.

[5] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, Manchester, 1988.

[6] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1–3):185–203, August 1981.

[7] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control. Wiley, New York, 2001.

[8] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference in Artificial Intelligence*, pages 674–679, Vancouver, Canada, August 1981.

[9] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, Berkeley, 1967. University of California Press.

[10] S. Marsland, J. Shapiro, and U. Nehmzow. A self-organising network that grows when required. *Neural Networks*, 15(8–9):1041–1058, October–November 2002.

[11] B. McCane, K. Novins, D. Crannitch, and B. Galvin. On benchmarking optical flow. *Computer Vision and Image Understanding*, 84(1):126–143, October 2001.

[12] S. Negahdaripour and B. K. P. Horn. A direct method for locating the focus of expansion. *Computer Vision, Graphics, and Image Processing*, 46(3):303–326, June 1989.

[13] K. Nordberg and G. Farneback. A framework for estimation of orientation and velocity. In *Proceedings of the 2003 IEEE International Conference on Image Processing*, volume 3, pages 57–60, 2003.

[14] S. E. Palmer. *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, 1999.

[15] D. Sazbon, H. Rotstein, and E. Rivlin. Finding the focus of expansion and estimating range using optical flow images and a matched filter. *Machine Vision and Applications*, 15(4):229–236, 2004.

[16] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the 1994 IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, Seattle, June 1994.

[17] D. Stavens. Lecture notes on optical flow and OpenCV. URL: http://robotics.stanford.edu/~dstavens/cs223b/, 2007. CS223-B Guest Lecture, Stanford University.

Sistema automático de diagnóstico de parásitos Helmintos

Gabriel Humpire
ICMC - Universidade de São Paulo
São Paulo - Brasil
ghumpire@icmc.usp.br

Juan Chuctaya
Universidad Nacional de San Agustín
Arequipa - Perú
juanherbert@gmail.com

Juan Gutierrez
Universidad Nacional de San Agustín
Arequipa - Perú
jcgutierrez@gmail.com

Odemir Bruno
IFSC - Universidade de São Paulo
São Paulo - Brasil
bruno@ifsc.usp.br

César Beltrán
Universidad Nacional de San Agustín
Arequipa - Perú
cesarbc@gmail.com

Resumen

The development of object recognition models in images is a challenging area because there is not any method to recognize all kinds of objects. In this paper, the problem is difficult because the objects in analysis are micro biological organisms called helminths which have morphological differences and extracted from human feces. We present a computational model that allows the detection, characterization and diagnosis of the parasites species presents in the fecal image micrography. For the detection of helminth eggs extraction techniques are applied based on the form over the segmentation realized. The techniques used for the segmentation were the salients regions, multithresholding and kmeans, the rate differentiation between eggs and artifacts achieving a 92 % success. The characterization of interest objects involves taking geometrical, morphological and texture features, making a total of 8 elements of the feature vector. For the classification process of the species, was applied supervised learning techniques, having experimented with Bayesian classifier with 93.5328 % success rate and Support Vector Machine with 94.0154 % success rate. Finally, it has developed a platform that allows diagnosis in real time via web, where Users can upload your pictures to our server, it is analyzed automatically obtaining the results of the species diagnosis under analysis. This is of particular interest to biologists professionals who have an additional tool for validate their diagnosis.

1. Introducción

Las infecciones parasitarias provocadas por Helmintos están vinculadas a condiciones socio-ambientales que aquejan a Peruanos de bajos recursos económicos que viven en condiciones favorables para el desarrollo de estos parásitos: cambios climáticos, alimentación insuficiente, vivienda precaria, tierra contaminada, ausencia de agua potable, etc [6]. Usualmente en puestos de salud publica de comunidades de bajos recursos del Perú no se dispone de personal capacitado para diagnosticar enfermedades provocadas por estos parásitos. Habitualmente para realizar el diagnóstico en este caso se recurre a recolectar muestras de heces de los pacientes y enviarlas a la ciudad céntrica más cercana para que realicen el diagnóstico.

La propuesta de realizar un sistema automático de diagnostico tiene por objetivo principal realizar un diagnostico en tiempo real tomando las muestras de las imágenes de las micrografías desde un microscopio y reducir el tiempo de espera para poder realizar un diagnostico rápido.

Cesar Beltrán en su tesis doctoral [3] realiza pruebas con parásitos de gallinas del género Eimeria mostrando una extracción de características geométricas, de textura y de curvatura, selecciona las características principales de forma estadística para usar los clasificadores bayesiano y estadístico obteniendo resultados del 80 % y 85 % respectivamente.

Yoon Seok [13] propone una identificación automática de huevos Helmintos en muestras fecales, realiza el diagnostico de parásitos en tres etapas: La primera etapa encargada del pre-procesamiento para descartar ruido de la imagen us-

ando binarización con umbral estático sobre las imágenes y obtener los candidatos segmentados, en la segunda etapa realiza la extracción de características usando la transformada de Fourier sobre el contorno en coordenadas polares del candidato (r, θ) considerando al origen como centro de masa siendo r la distancia hacia los contornos con $0 < \theta < 2\pi$ y finalmente la tercera etapa es la clasificación usando dos redes neuronales, una red dedicada exclusivamente para descartar artefactos no deseados, la siguiente recibe las salidas de la red anterior para clasificar los tipos de parásitos obteniendo un 83 % de precisión.
Derya [7] realiza un diagnóstico automático de 16 especies de huevos Helmintos. Para la detección del candidato realizó un pre-procesamiento de la imagen aplicando el filtro mediana para eliminar el ruido, operaciones morfológicas para obtener únicamente al candidato y un umbral para descartar objetos más grandes. Obtuvo características usando momentos de Hu [4] basado en la distribución de intensidad de la imagen, finalmente una multi-class Support Vector Machine (SVM) para la clasificación obteniendo resultados del 97.7 % . Posteriormente Chuctaya[10] se enfoca en detección de huevos helmintos usando técnicas de binarización adaptativa en imágenes globales sobre 4 especies de helmintos.

El presente trabajo esta estructurado de la siguiente manera: en la sección 1 explicaremos el fundamento e importancia de la propuesta, en la sección 2 discutiremos la propuesta que esta divida en tres principales etapas segmentación, extracción de características y clasificación, en la sección 3 discutiremos los experimentos sobre la base de datos SADPI8, resultados y conclusiones, finalmente la bibliografía.

2. Propuesta

Automatizar un sistema realizado por personal altamente capacitado como los Biólogos expertos en parasitología se dificulta al depender de procesos no automatizables variables como el procedimiento sedimentación y centrifugación de la muestra, incluso posterior a este procedimiento es necesario aplicar una solución química sobre la micrografía para resaltar los parásitos ante el ojo humano.
La presente propuesta automatiza el proceso de detección e identificación de parásitos Helmintos en micrografías. El diagnostico realiza una selección de candidatos de parásitos denominada detección de huevos, posteriormente se procede a la representación numérica llamada extracción de características y finalmente la etapa de clasificación que determina la pertenencia a una clase de parásito Helminto de los candidatos.
Todo el procedimiento anteriormente explicado es mostrado de forma gráfica en la figura 1 y sera detallado en las siguientes subsecciones.

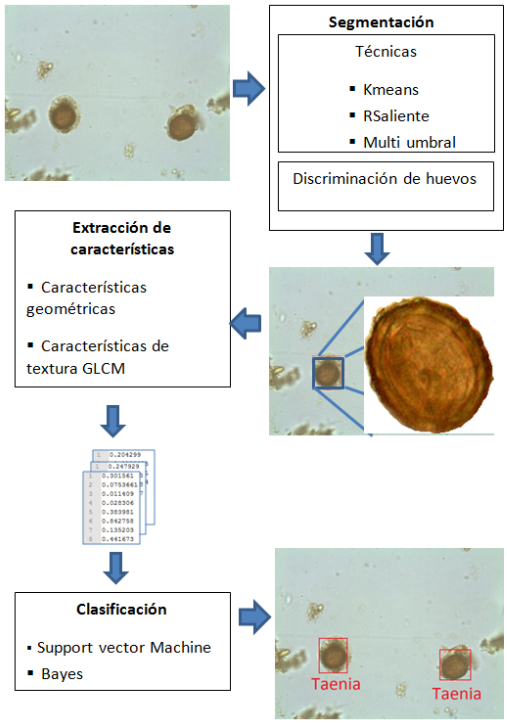


Figura 1: Esquema de diagnostico propuesto.

2.1. Detección de huevos

La segmentación de la imagen involucra determinar estrategias que permitan adecuarla para posteriormente diferenciar las zonas de interés. Este componente consiste en 2 operaciones:

2.1.1. Segmentación La característica de una micrografía fecal es que esta incluye residuos orgánicos que son productos de la digestión (tejidos epiteliales, microorganismos,etc), así como diferencias morfológicas que presentan las especies y degradación en la calidad de la imagen. Tomando en cuenta estos criterios procedemos a normalizar la imagen usando el filtro de reducción de ruido como el gaussiano y el filtro de la mediana con una vecindad [3 x 3]. Posteriormente tenemos tres técnicas utilizadas para la separación de pixeles.

- Region Saliente: La primera estrategia realizada es la segmentación por regiones salientes [1] [9]. Aplicamos esta técnica sobre la transformación RGB a CIELAB, el cual da mejor realce a características luminosas. Posteriormente calculamos la distancia de los

pixeles contra un valor promedio calculado por canal del modelo de color CIELAB (L,a,b) de esta manera mediante un procedimiento minmax normalizamos la imagen y obtenemos finalmente una imagen en grises con las intensidades salientes. En la ecuación 1 tenemos PM_k representa el valor promedio obtenido del k-ésimo canal, N_1 y N_2 son números de píxeles de las filas y columnas, $p_{i,j,k}$ es el valor de intensidad del pixel (i, j).

$$PM_k = \frac{1}{N_2}\sum_{j=1}^{N_2}\left[\frac{1}{N_1}\sum_{i=1}^{N_1} p_{i,j,k}\right] \qquad k = 1, 2, 3 \tag{1}$$

Hallado el valor promedio por canal, se normaliza los pixeles de un canal a el valor promedio calculado y combinado las diferencias entre mapas obtenemos una nueva matrix con las intensidades mas resaltantes de la imagen como se describe matemáticamente en la ecuación 2, donde $C_{i,j}$ es la matriz resultante del proceso, N_1 y N_2 son números de píxeles de las filas y columnas, $p_{i,j,k}$ es el píxel en (i, j) y k es el canal respectivo.

$$C_{i,j} = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \|(p_{i,j,1}) - (PM_1)\| + \|(p_{i,j,2}) - (PM_2)\| + \|(p_{i,j,3}) - (PM_3)\| \tag{2}$$

- Multi umbral : Se realiza el cálculo de múltiples umbrales teniendo en cuenta la distribución de intensidades del histograma dado un número de clústers. El procedimiento primero calcula valores máximos y mínimos del histograma definiendo limites de intensidades de una imagen en escalas grises hallada previamente de la conversión de una imagen a color a escalas grises posteriormente se calcula la longitud del histograma y se divide entre el numero de clústers obteniendo de esta manera los intervalos los cuales representan los umbrales hallados de manera que particionan las intensidades de la imagen en k secciones binarias o mascaras.
- Kmeans : Esta técnica clusteriza los pixeles en clases o conjuntos deacuerdo a la cantidad de cluster debido a que esta técnica utiliza centroides aleatorios representa un problema para la ubicación de los parásitos por lo que al igual que en la multiumbralización se normaliza previamente el histograma para obtener los valores mas intensidades mas resaltantes estableciendo de esta manera centroides deacuerdo al análisis de histograma.

2.1.2. Discriminación de huevos Dada una imagen binaria se extraen los bordes de cada uno de los objetos presentes posteriormente usamos la técnica *Least Square fitting* [8] el cual dada una distribución de coordenadas de los objetos encuadra una elipse que encaje a estos puntos utilizando internamente una función discriminante para emparejar. De esta manera obtenemos el borde y las coordenadas de la elipse por objeto en consecuencia para calcular su similitud se generan series del centroide del polígono a los bordes así generamos 2, posteriormente calculamos la distancia *Dynamic time warping* entre contornos finalmente usando este resultado y la información de los radios obtenida de formas empirica determinamos el descarta o no de una región para ser utilizado posteriormente en las siguientes etapas.

2.2. Extracción de características y clasificación

Actualmente los biólogos para distinguir a un parásito de un artefacto no deseado se basan en la morfología de los Helmintos que normalmente son circulares y ovoides, el siguiente paso para realizar el descarte es distinguir si la textura es rugosa o lisa; no consideran al color como una característica relevante por ser directamente dependiente de las soluciones químicas que aplican a la micrografía para que los candidatos puedan ser resaltados de mejor manera para el ojo humano.
La propuesta esta fundamentada en la simulación de los criterios de los Biólogos para distinguir las especies de los parásitos. Por lo que la dimension del vector de características es de 8 y esta compuesto por dos descriptores: 4 geométricas y 4 de textura *Gray Level Co-Ocurrence Matrix*(GLCM) que usan el contorno y la imagen en escala de grises del candidato respectivamente.

2.2.1. Características geométricas Definen la morfología del candidato:

- Diámetro máximo, es la mayor distancia existente entre dos puntos del contorno (figura 2). Es calculado usando *Principal Components Analysis*(PCA)[12].
- Diámetro mínimo, es la distancia ortogonal desde el punto medio de la linea generada por el diámetro máximo hacia el contorno.

Figura 2: Diámetros máximo y mínimo de un contorno

- Área, píxeles contenidos en el contorno del parásito. Logra distinguir los parásitos grandes de los pequeños

- Excentricidad, es la proporción existente entre el diámetro máximo y diámetro mínimo, esta característica ayuda a diferenciar a los huevos de parásitos que son ovoides de los circulares.

2.2.2. Características de textura interna GLCM Las características de textura GLCM[11] están basadas en la matriz de co-ocurrencia sobre la imagen segmentada en escala de grises, las pruebas fueron realizadas usando 8 niveles de umbral y desplazamientos $\delta x = 3$, $\delta y = 3$. Las características calculadas son: correlación, energía, contraste y homogeneidad. La etapa de segmentación retorna imágenes recortadas contedenedoras de candidatos con background blanco, esta particularidad es aprovechada para descartar cualquier pixel blanco del conteo para la matriz de co-ocurrencia garantizando así obtener características exclusivamente del candidato en escala de grises.

Para identificar la especie de los candidatos es necesario clasificarlos, fueron usados dos clasificadores: *support-vector machine*(SVM) y clasificador Bayesiano. SVM tiene varias funciones kernels propuestas[5], la forma genérica es $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$, se realizaron pruebas con los siguientes kernels:

- Linear: $K(x_i, x_j) = x_i^T x_j$
- Polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- Radial basis function(RBF): $K(x_i, x_j) = exp(-\gamma \parallel x_i - x_j \parallel^2), \gamma > 0$
- Sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

Los parámetros de configuración para el SVM fueron hallados por pruebas intensivas sobre la base de datos SADPI8 v2.0 obteniendo los siguientes parámetros: Kernel polinomial de grado $d = 1{,}8$, gamma $\gamma = 4$, $r = 0$ y penalty parameter of the error $C = 76$.

3. Experimentos y resultados

Los experimentos fueron realizados sobre la base de datos SADPI8 v2.0 proporcionada por el grupo PibapUnsa[1] que contiene imágenes microscópicas extraídas de dos laboratorios de diferentes ciudades del Perú mostrados en la figura 3. Las imágenes fueron obtenidas con un microscopio óptico (MOTIBAC) el cual esta acoplado a un ordenador por el cual se capturan imágenes con una magnificación objetiva 40 x en formato JPG. Los nombres de los parásitos están listados en el cuadro 1 y pueden ser vistos en la figura 4.

El conjunto de imágenes de micrografías por especie así como imágenes segmentadas de entrenamiento están puestos en el siguiente cuadro 1.

1 The website of this project is www.pibapunsa.org

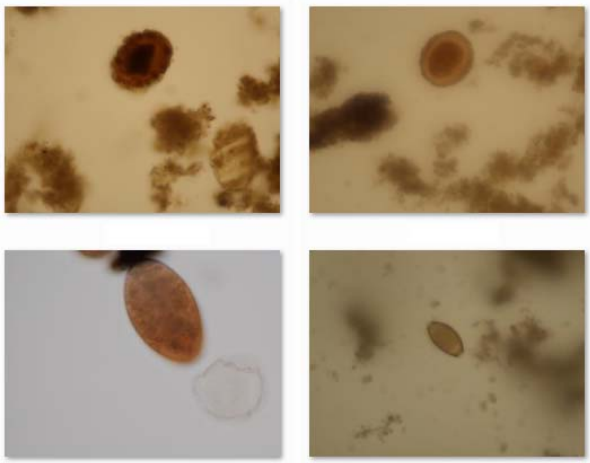


Figura 3: Base de datos SADPI8 v2.0

Especie	Num de micrografías	Num de muestras segmentadas
Ascaris Lumbroides	100	81
Uncinarias	75	51
Trichuris Trichiura	102	106
Hymenolepis Nana	115	74
Dyphillobothrium	103	164
Taenia Solium	211	231
Fasciola Hepatica	280	259
Enterobius Vermicularis	91	70
Total	**866**	**1036**

Cuadro 1: Cantidad de micrografías por especie y parásitos segmentados encontrados en las micrografías

3.1. Validacion de detección de huevos de parásitos

Los resultados porcentuales obtenidos en detección de parásitos se denotan en el cuadro 2 en el cual se emplearon las técnicas de regiones salientes y Multi umbralización. Como se puede apreciar las técnicas tienen sus ventajas y desventajas en la detección de parásitos de acuerdo a la naturaleza morfológica de estos.Anteriormente nosotros probamos con la técnica contornos activos dinámicos pero este requería de muchos parámetros de configuración para establecer la detección de bordes en vista que esta técnica utiliza las fuerzas internas y externas de la imagen. También se probo con técnicas como *Watershed* pero se tubo el problema de la sobre segmentación debido a el entorno variable que presentan las imágenes así también se utilizo la *transformada de Hough* pero el uso de acumuladores hacia que el proceso no sea óptimo debido a la búsqueda de condiciones

Especies	Técnicas de procesamiento	
	Regiones Salientes %	Multi umbral %
Ascaris Lumbroides	95.3	96.0
Uncinarias	90.0	90.0
Trichuris Trichiura	68.3	92.0
Hymenolepis Nana	60.0	89
Dyphillobothrium	90.0	92.3
Taenia Solium	89.5	86.6
Fasciola Hepatica	90.0	90.0
Enterobius Vermicularis	40.0	93.0

Cuadro 2: Cuadro de medidas de prescisión de detección de huevos por especie.

que asemejen los puntos a una circunferencia o elipse.

3.2. Validacion de la clasificación de parásitos

Por otro lado, la validación para la etapa de clasificación fue realizada con *leave one out cross-validation* en prueba intensiva sobre toda la base de datos y en el caso de SVM para cada uno de los kernels con los mismos parámetros, los resultados de precision para cada uno de ellos son mostrados en el cuadro 3.

La discusión de resultados se realizará por medio de matrices de confusión para ambos clasificadores, la diferencia de precisión es pequeña pero cabe resaltar la robustez que muestran los resultados ofrecidos por el clasificador SVM al mostrar errores que incluso los biólogos suelen cometer por tener características similares entre parásitos comprobando así la eficiencia del SVM sobre muchos clasificadores existentes actualmente[2].

Otro punto importante a discutir es el tiempo de entrenamiento de los clasificadores, el clasificador Bayesiano fue mucho más rápido que el SVM, pero para la etapa de testeo ambos obtuvieron igual performance en tiempo de respuesta.

Precisión para clasificador SVM		
Kernel	**Precision(%)**	**Error(%)**
Polynomial	94.0154	5.98456
RBF	93.9189	6.08108
Linear	93.2432	6.75676

Cuadro 3: Resultados de precision para los diferentes kernels del clasificador SVM

Para el clasificador Bayesiano se obtuvo una precision 93.5328 % y error 6.46718 % los resultados son mostrados en una matriz de confusión mostrada en el cuadro 4 donde las clases del 1 al 8 son las listadas en la figura 4.

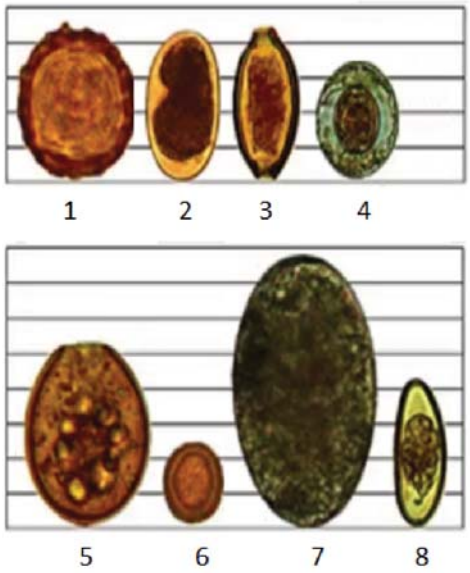


Figura 4: Nombres de parásitos: 1) Ascaris 2) Uncinarias 3) Trichuris trichuria 4) Hymenolepis nana 5) Dyphillobothrium pacificum 6) Taenia solium 7) Fasciola hepática 8) Enterobius vermicularis

Mientras que para el clasificador SVM se obtuvo una precision 94.0154 % y error 5.98456 % los resultados son mostrados en una matriz de confusión mostrada en el cuadro 5 con kernel Polynomial por haber obtenido mejor precision en el cuadro 3. El proyecto PIBAPUNSA ha desarrollado un sistema con interfaz web en tiempo real en la url http://www.pibapunsa.org que permite al usuario registrarse y subir sus micrografías.

Matriz de confusión para el clasificador Bayesiano									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**#**
1	74	0	0	6	1	0	0	0	**81**
2	0	50	0	0	1	0	0	0	**51**
3	0	0	94	1	0	0	0	11	**106**
4	3	0	0	65	6	0	0	0	**74**
5	5	5	0	10	144	0	0	0	**164**
6	0	0	2	4	7	218	0	0	**231**
7	0	0	0	0	0	0	259	0	**259**
8	0	0	5	0	0	0	0	65	**70**
	82	**55**	**101**	**86**	**159**	**218**	**259**	**76**	**1036**

Cuadro 4: Matriz de confusión para el clasificador Bayes : Error 6.46718 % Precision 93.5328 %

Matriz de confusión para el clasificador SVM									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**#**
1	78	0	0	2	0	1	0	0	**81**
2	1	47	0	0	3	0	0	0	**51**
3	1	1	91	0	0	1	0	12	**106**
4	4	0	0	58	10	2	0	0	**74**
5	0	4	0	6	152	2	0	0	**164**
6	0	0	0	5	4	222	0	0	**231**
7	0	0	0	0	0	0	259	0	**259**
8	0	0	3	0	0	0	0	67	**70**
	84	**52**	**94**	**71**	**169**	**228**	**259**	**79**	**1036**

Cuadro 5: Matriz de confusión para el clasificador SVM : Error 5.98456 % Precision 94.0154 %

El sistema se encarga automáticamente de realizar el preprocesamiento necesario para la detección de los huevos, extracción de características y clasificación de patrones. Todos los módulos fueron desarrollados en $C++$, por lo que el tiempo de respuesta es rápido en cuanto al procesamiento y diagnóstico de la micrografía.

3.3. Conclusiones

Nuestra propuesta es innovadora desde que con técnicas computacionales simples se consigue resolver un problema que tiene un impacto en la salud pública. Además, nuestra propuesta comprende el desarrollo de todas las etapas del proceso de diagnóstico por imágenes, en vista que los trabajos encontrados en la literatura no abordan el problema en forma conjunta, sino que lo analizan de forma separada. Esto no permite determinar el aporte real de esas investigaciones. Básicamente nuestro aporte consiste en las tres etapas más importantes de reconocimiento de objetos, como son: Detección de los huevos de helmintos, este trabajo no ha sido estudiado por otros grupos en vista que las micrografías muestran una alta presencia de basura o ruido, en la mayoría de las veces esta basura está unida a los objetos de interés. En ese sentido, nuestro modelo consigue muy bien diferenciar los huevos de helmintos de la basura. Caracterización y clasificación de objetos, considerando que existe cierta similitud entre algunas de las especies, nuestra caracterización de especies ha conseguido distinguir una de otra con una alta tasa de acierto. Para ello se ha conseguido caracterizar aspectos morfológicos que identifiquen la geometría, textura y forma de los huevos de helmintos. En resumen, nuestro enfoque muestra resultados muy alentadores que posibilitan resolver el problema de diagnóstico de especies, consiguiendo una tasa casi al nivel de los mayores expertos en diagnóstico de huevos de Helmintos. Adicionalmente, nuestra propuesta se muestra adecuada para desarrollar un sistema de diagnóstico remoto, donde sólo sea necesario que se nos envíe la imagen microscópica para poder realizar el análisis a través de nuestro sistema.

Referencias

[1] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk. Salient region detection and segmentation. Technical report, School of Computer and Communication Sciences and Ecole Polytechnique Féderale de Lausanne, 2008.

[2] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2:121–167, June 1998.

[3] C. B. Castañon. *Análise e reconhecimento digital de formas biológicas para o diagnóstico automático de parasitas do gênero Eimeria*. PhD thesis, IME-Universidade de Sao Paulo, 2007.

[4] P. E. Y. X. Chen, Q.a. A comparative study of fourier descriptors and hu's seven moment invariants for image recognition. *Canadian Conference on Electrical and Computer Engineering*, 2004.

[5] C.-C. C. Chih-Wei Hsu and C.-J. Lin. A practical guide to support vector classification. Technical report, National Taiwan University, 2010.

[6] I. de Higiene. Generalidades de helmintos. Technical report, Universidad de la República, Facultad de Medicina, Uruguay.

[7] A. Derya and V. Asaf. An expert diagnosis system for classification of human parasite eggs based on multi-class svm. *Expert Syst. Appl.*, 2009.

[8] A. Fitzgibbon, M. Pilu, and R. B. Fisher. Direct least-squares fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:476 – 480, 1999.

[9] G. H. A. R.-C. B. J. Chuctaya, J. P. Mena-Chalco and R. Patiño. Detección de huevos helmintos mediante plantillas dinámicas. *Conferencia Latinoamericana de Informática - CLEI*, pages 1–12, 2010.

[10] C. B. Juan Chuctaya, Raquel Esperanza Patino Escarcina. Detection of helminth parasite eggs in digital biological micrographs. *SIBGRAPI 2010*, pages 689–690, 2010.

[11] S. W.-C. Lam. Texture feature extraction using gray level gradient based co-occurrence matrices. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 1996.

[12] K. E. Svante Wold and P. Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 1987.

[13] H. C. K. M.-H. C. Yoon Seok Yang, Duck Kun Park and J.-Y. Chai. Automatic identification of human helminth eggs on microscopic fecal specimens using digitalautomatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network. *IEEE*, 2001.

Seleção Automática de Parâmetros de Detectores de Bordas de Imagens Digitais Baseado em Algoritmos Genéticos

Cristiane Leal da Costa
crislealcosta@globo.com

Inês Aparecida Gasparotto Boaventura
ines@ibilce.unesp.br

DCCE - IBILCE - UNESP
Rua Cristovão Colombo, 2265
15054-000 - São José do Rio Preto, SP, Brasil

Resumo

The aim of this paper is to present a new method of automatic selection of edge detector parameters. The developed system automates the process of parameters adjustment through a intelligent system based on genetic algorithm (GA). The adjustment of these parameters uses the information of quantitative assessment of the edges obtained in different steps, regarded here as fitness function of the GA. The quantitative assessment is performed using quantitative measures for the optimal adjustment of the edges. These measures are obtained by comparing the edges generated by the edge detector with the ideals edges, also called ground truth image, which is considered absolute truth. The space ROC (Receiver Operating Characteristics) is used to classify the edges in different steps of the algorithm application. The edge detector FUNED (Fuzzy Number Edge Detector) is used to show the feasibility of the proposed method. The results obtained so far show that is feasible select optimized parameters and the precision of edge detection is improved adaptively.

1. Introdução

Borda é a característica básica de um objeto em imagens digitais. O processo de detecção de bordas reduz uma imagem aos detalhes de contorno dos objetos que estão presentes na imagem e que, frequentemente, são usados em operações de análise de imagem para obtenção de características e reconhecimento de objetos. Os pixels de borda são definidos como localizações na imagem onde existe uma variação significativa nos níveis de cinza em uma direção fixa através de alguns pixels. O desempenho de um detector de bordas é o elemento chave de algoritmos bem sucedido para a análise e percepção de imagens.

Existem diferentes métodos de detecção de bordas e, de maneira geral, são baseados em duas abordagens principais: análise do gradiente e análise de regiões [9]. Dentre os métodos baseados em gradientes, encontram-se o filtro de Sobel, Prewitt, Robert, Canny, e outros [4] Esses métodos são baseados na descontinuidade dos gradientes de uma imagem e, embora os algoritmos destes métodos sejam simples e implementados facilmente, são sensíveis à ruídos e possuem dificuldade em extrair bordas de imagens com iluminação não uniforme, presentes na maioria das imagens naturais. Por outro lado, existem técnicas que caracterizam a detecção de bordas como um problema de raciocínio fuzzy. Estas técnicas são baseadas em regiões da imagem e utilizam a teoria fuzzy de diversas maneiras [1]. Dentre os detectores fuzzy, tem-se o detector de bordas FUNED (Fuzzy Number Edge Detector) [3], o qual considera uma vizinhança local dos pixels da imagem, definida pelo usuário e, baseado no conceito de números fuzzy, verifica se um pixel pertence ou não àquela região da imagem, com base na intensidade dos tons de cinza que compõem a região. O pixel que não pertence a região é classificado como um possível pixel de borda. Por meio de uma função de pertinência, a técnica fornece uma matriz de pertinência em tons de cinza e, pela escolha de um limiar, as bodas da imagem são segmentadas. Para a modelagem do problema, os tons de cinza são considerados como números fuzzy e, para cada pixel da imagem, calcula-se a sua pertinência em relação a uma determinada região, considerando os vizinhos que possuem níveis de cinza próximos do pixel em análise. Quando se considera os valores de cinza, incorpora-se a variabilidade inerente dos valores de cinza da imagem, proporcionando assim uma abordagem mais adequada ao tratamento de imagens digitais em comparação ao tratamento clássico, baseado em uma formulação analítica. A técnica permite ao usuário o ajuste manual desses parâmetros: o tamanho da vizinhança local a ser considerada; o suporte de um número fuzzy

e o limiar. Embora o ajuste manual destes parâmetros proporciona diversas possibilidades de visualização das bordas de uma imagem, permitindo ao usuário a escolha de detalhes da imagem, pode dificultar o uso do detector FUNED, quando o usuário está interessado simplesmente nas bordas ótimas da imagem.

Por outro lado, Boaventura [1] desenvolveu uma metodologia de avaliação de detectores de bordas baseada na análise ROC. O processo de avaliação considera diferentes medidas, que são tomadas comparando-se as bordas obtidas com as bordas ideais de uma imagem. A técnica calcula, para as bordas obtidas, um ponto no espaço e com essa informação, pode-se comparar desempenhos de diferentes detectores de bordas ou, comparar o desempenho de resultados diferentes de um mesmo detector de bordas, quando são variados os seus parâmetros.

O Ajuste automático de parâmetros de detectores de borda é um assunto de grande interesse e tem sido abordado com frequência na literatura científica. A maior parte das técnicas apresentadas até agora são baseadas na teoria estatística [9, 6, 10]. Neste artigo é proposto um método de seleção automática de parâmetros de detectores de bordas para melhorar a habilidade adaptativa e determinativa de detectores de bordas. A abordagem proposta utiliza algoritmos genéticos (AG) [5, 7], os quais possuem características de alta eficiência e otimização global e podem ser aplicados em áreas onde a busca heurística se faz necessária. A definição de uma função objetivo apropriada ao problema de detecção de bordas, faz com que o AG opere adequadamente, levando a uma estratégia efetiva para seleção automática de parâmetros para detectores de borda, com boa performance e facilidade de implementação. Na seção 2 apresenta-se brevemente o detector de bordas FUNED. A seção 3 é dedicada ao método de avaliação de detecção de bordas baseado no espaço ROC. O método de seleção automática de parâmetros baseado em algoritmo genético é discutido na seção 4. Alguns resultados experimentais são mostrados na seção 5. As conclusões são apresentadas na seção 6.

2. O Detector de Bordas FUNED

Os níveis de cinza de uma imagem são interpretados como números fuzzy. Com essa interpretação, incorpora-se a variabilidade inerente aos valores de cinza de imagens, proporcionando assim uma abordagem mais adequada ao tratamento de imagens digitais, em comparação ao tratamento clássico, baseado em uma formulação analítica. Para esta formulação foram adotados os números fuzzy simétricos e triangulares definidos pelas equações (1) e (2):

$$\mu_A(x) = f(|x - g|, \delta), \tag{1}$$

sendo que $f(.)$ controla a forma do número fuzzy e $\delta \in R^+$, chamado de suporte, controla o espalhamento do número fuzzy. O número fuzzy representado pela equação (2) é um número fuzzy triangular e simétrico.

$$\mu_A(x) = MAX(0, 1 - \frac{|x - g|}{\delta}). \tag{2}$$

A função de pertinência utilizada pelo detector de bordas tem a seguinte formulação [2, 3]: para cada pixel $g_{i,j}$ da imagem, interpretado como um número fuzzy, calcula-se a pertinência desse pixel em relação à região determinada por uma vizinhança local. Assim, seja uma imagem $A_{N\times M}$ em que, para cada pixel $g(i, j)$, tem-se uma janela de vizinhança espacial $W \times W$. A função de pertinência, $\hat{\mu}_{g(i,j)}$, de cada pixel $g(i, j)$ ao conjunto fuzzy *Região Homogênea* da imagem é definida pela equação:

$$\hat{\mu}_{g(i,j)} = \frac{\sum_{k=1}^{W}\sum_{l=1}^{W} \max(0, 1 - \frac{|g(i,j)-A(k,l)|}{\delta}) - 1}{W^2 - 1}, \tag{3}$$

$$i = 1 \ldots N, j = 1 \ldots M,$$

sendo que $\delta \in R$ representa o parâmetro de espalhamento do número fuzzy. Quanto menor δ, menor é a base do número triangular, ou seja, menor é o intervalo onde se considera a pertinência dos vizinhos.

A matriz resultante desta operação é chamada de matriz de pertinências ou imagem pertinência e é definida pela equação:

$$A' = \sum_{i=1}^{M}\sum_{j=1}^{N} \frac{\hat{\mu}_{ij}}{g_{ij}}, \tag{4}$$

sendo que, $\hat{\mu}_{ij}$ indica o grau de pertinência de cada pixel à região $W \times W$ pré-definida da imagem. Os valores de pertinências próximos de 1 significam uma maior pertinência dos respectivos pixels a uma região homogênea. Para valores de pertinências próximos de 0, tem-se que os pixels avaliados são diferentes dos pixels da região, ou melhor, podem pertencer à borda de uma região e não a uma região homogênea. Os pixels mais escuros que aparecem na matriz de pertinências são aqueles com menor pertinência a uma região homogênea, ou seja, com maior possibilidade de representar as bordas da imagem.

3. Método para Avaliação de Detectores de Bordas

Nesta abordagem, um detector de bordas é considerado como um classificador de padrões. Cada pixel da imagem pode ser mapeado para um elemento do conjunto de rótulos de classe {1,0}, em que 1 significa pertencente à classe borda (positivo) e 0 significa não pertencente à classe borda (negativo).

Dado um detector de bordas, a matriz de pixels que representa a imagem é o conjunto de amostras, em que cada pixel pode assumir valores *positivo* e *negativo*, respectivamente. Ao aplicar o detector de bordas a um pixel da imagem (elemento de entrada), tem-se quatro situações distintas. Se o elemento de entrada é um pixel de borda, considerado *genuíno* (positivo), e o detector de bordas o classifica como *positivo*, conta-se como *Verdadeiro Positivo* (VP); se é classificado como *negativo*, conta-se como *Falso Negativo* (FN). Se o elemento de entrada não é um pixel de borda, ou seja é um *impostor* (negativo) e, é classificado como tal, conta-se como *Verdadeiro Negativo* (VN); se é classificado como *positivo*, conta-se como *Falso Positivo* (FP).

Para quantificar a eficácia de um detector de bordas são tomadas as seguintes medidas: razão de verdadeiros positivos (sensibilidade), razão de falsos positivos. A razão de verdadeiros positivos, também chamada de sensibilidade, é a proporção de verdadeiros positivos classificados corretamente como genuínos, ou seja, é a proporção de genuínos de uma classe que são classificados corretamente como genuínos. A razão de falsos positivos, ou taxa de falso alarme, é a proporção de falsos positivos classificados relativamente ao total de impostores existentes, ou seja, é a proporção de impostores erroneamente classificados como genuínos.

Para a análise do desempenho dos métodos de detecção de bordas, as imagens de bordas obtidas devem ser comparadas com as respectivas imagens ideais (*GT*). A correspondência entre bordas obtidas e bordas ideais é feita pixel a pixel. Assim, seja a matriz B o mapa de bordas obtido a partir de uma imagem G e seja a matriz GT a imagem de bordas ideal correspondente. Cada elemento $b_{i,j}$ em B é avaliado em relação ao elemento $gt_{i,j}$ em GT e essa correspondência pode ser não exata. Neste caso, considera-se um limiar T_{mj}, chamado de limiar de tolerância, o qual permite fazer a correspondência de um pixel de borda $b_{i,j}$ em B com um pixel de borda ideal p em GT, mesmo que p esteja a uma pequena distância ao redor de uma vizinhança $W_{T_{mj}\times T_{mj}}$ centralizada em $gt_{i,j}$. Dada essa tolerância, em geral, um pixel de borda em B pode ter múltiplos casamentos potenciais a um pixel de borda correspondente em GT, em torno dessa vizinhança, neste caso, toma-se o pixel com distância mínima. Quando a correspondência é estabelecida, marca-se o pixel correspondente com zero na matriz GT.

O cálculo dos verdadeiros positivos (VP) e falsos positivos (FP) é dado pela seguinte expressão: se $b_{i,j} = 1$ e ($gt_{i,j} = 1$ ou $\exists\ p \in W_{T_{mj}\times T_{mj}}$ centralizada em $gt_{i,j}$ tal que $p = 1$) então considera-se verdadeiro positivo (VP), caso contrário, considera-se falso positivo (FP).

O cálculo dos verdadeiros negativos (VN) e falsos negativos (FN) é feito da seguinte maneira: se $b_{i,j} = 0$ e $gt_{i,j} = 0$ então considera-se verdadeiro negativo (VN), caso contrário, considera-se falso negativo (FN).

Dada uma imagem de bordas B, sua correspondente imagem ideal GT e uma janela de vizinhança $W_{T_{mj}\times T_{mj}}$ centralizada em $gt_{i,j}$, a menor distância entre um determinado pixel $p \in W_{T_{mj}\times T_{mj}}$ e o pixel $gt_{i,j}$ é zero, ou seja $p = gt_{i,j}$ e a maior distância entre p e $gt_{i,j}$ é $\lfloor T_{mj}/2 \rfloor\sqrt{2}$. Uma vez que o processo de comparação entre as bordas obtidas e bordas ideais considera a janela $W_{T_{mj}\times T_{mj}}$, essa busca fica restrita à distância euclidiana em R^2 $d_{\pounds 2} \in [0, \lfloor T_{mj}/2 \rfloor\sqrt{2}]$. Para a avaliação dos detectores de bordas, é considerada a variação de distâncias dentro desse intervalo. Por exemplo, se for tomado o limiar $T_{mj} = 5$ então, em uma janela de vizinhança $W_{5\times 5}$, tem-se as seguintes distâncias entre o pixel central $g_{i,j}$ e os demais pixels de W: {0, 1, 1.4, 2, 2.2, 2.8}. A avaliação pode ser restringida à busca por bordas com distância 0, ou seja, o casamento de pixels somente é aceito quanto $b_{i,j} = gt_{i,j}$. A avaliação pode ser estendida aos pixels com distância de no máximo 1.4, em que a busca por pixels de borda é estendida à vizinhança $W_{3\times 3}$ do pixel central $gt_{i,j}$. Conforme a distância é aumentada, flexibiliza-se a busca por bordas em uma região próxima ao pixel em análise.

A partir das medidas de sensibilidade e taxa de falso alarme, o espaço ROC (*Receiver Operating Characteristics*) é usado para avaliar os resultados de classificação de um detector de bordas. A figura 1 ilustra o espaço ROC de um detector de bordas, o qual é um classificador discreto. O espaço ROC de um detector de bordas mostra um ponto, o par ordenado (Razão de Falso Positivo (RFP), Razão de Verdadeiro Positivo (RVP)). Deve-se observar na figura 1 alguns pontos indicados com significados importantes. O ponto (0,1) representa um resultado de detecção de bordas perfeito, ou seja, nenhum falso positivo foi encontrado e todos os verdadeiros positivos foram classificados. Este caso pode ser representado pelo detector D. O ponto (x,0) representa o pior caso, ou seja, o detector de bordas não apresenta nenhum verdadeiro positivo. No ponto inferior esquerdo (0,0) o detector de bordas não apresenta nenhuma saída. A linha diagonal que vai do ponto (0,0) ao ponto (1,1) divide o espaço ROC em áreas de classificação boas ou ruins e é também chamada de linha de não discriminação (*random guess*). De maneira geral, um ponto no espaço ROC é melhor que outro se ele está mais à noroeste, em que a razão dos verdadeiros positivos é maior e/ou a razão de falsos positivos é menor. Classificadores no lado esquerdo do gráfico ROC (perto do eixo Y) são ditos "conservadores", pois fazem classificações positivas somente com uma evidência forte, portanto comentem poucos *falsos positivos*. Classificadores no lado direito são ditos "liberais", pois fazem classificações positivas com pouca evidência, mas cometem muitos erros "falsos positivos". Na Figura 1, A

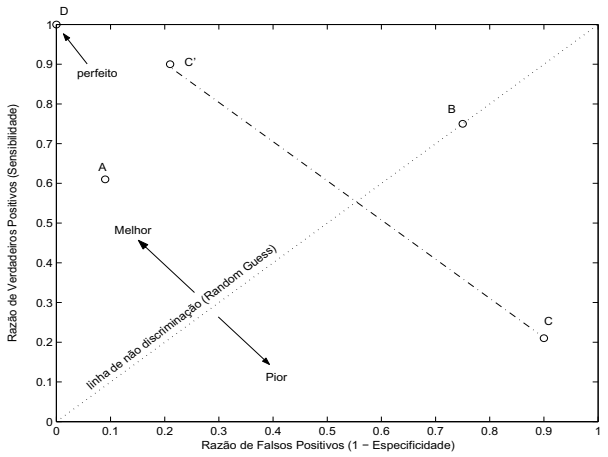


Figura 1. Espaço ROC de um detector de bordas.

Início

1. Gerar a população inicial
2. Repita até obter 200 gerações sem evolução
   (a) Aplicar o Detector de Bordas FUNED e Avaliar a População;
   (b) Selecionar a Elite;
   (c) Repita até gerar nova população
      i. Selecionar dois indivíduos;
      ii. Cruzar esses indivíduos gerando dois novos indivíduos;
   (d) Realizar a mutação da população gerada;
   (e) Verifica se houve evolução;

Fim

Figura 2. Algoritmo Genético Implementado

é mais conservador do que C'. Classificadores sobre a linha diagonal são considerados classificadores "aleatórios" (Na Figura 1, B é um classificador aleatório)

4. Método de Seleção Automática de Parâmetros de Detectores de Bordas

A geração manual de parâmetros para obtenção de bordas ótimas de uma imagem é uma tarefa que demanda bastante tempo do usuário. O uso de uma técnica de otimização evolucionária, tais como algoritmos genéticos, é bastante útil, uma vez que o problema pode ser visto como um problema de otimização combinatorial. Assim, a automatização do processo de ajuste dos parâmetros do detector de bordas FUNED é feita por meio de um algoritmo genético. O ajuste dos parâmetros utiliza as informações de avaliação quantitativa das bordas obtidas em diferentes passos. A figura 2 apresenta o algoritmo correspondente ao algoritmo genético implementado. Nas subseções seguintes são apresentados os passos do algoritmo desenvolvido, considerando as seguintes questões: o esquema de codificação dos cromossomos; a metodologia empregada para produzir a população inicial; a função objetivo; os operadores genéticos de seleção, cruzamento, mutação, elitismo; e o critério de parada.

4.1. O esquema de Codificação

A codificação de uma solução do problema em um cromossomo é uma questão crucial para os algoritmos genéticos. Considerando que a seleção de parâmetros do detector FUNED é um problema de otimização restrita, o esquema de codificação real de variáveis e um operador de busca, que respeite as regiões contínuas no espaço de busca, tem desempenho melhor do que o esquema de codificação binária [8]. Portanto, utiliza-se nesta implementação o esquema de codificação real dos cromossomos . O espaço de busca do problema em questão estão delimitados dentro das seguintes faixas de valores: $W \in [2, 5]$, $Threshold(Thr) \in [0, 1]$ e $\sigma \in [1, 100]$. Observa-se que os valores para W e σ foram restringidos por meio de avaliações empíricas.

4.2. Produção da População Inicial

Pelo esquema de codificação escolhido, o espaço da solução coincide com o espaço de cromossomo. O tamanho da população pode influenciar no tempo de processamento do algoritmo e uma população inicial pequena pode prejudicar a convergência do algoritmo. Assim, foi escolhido para esse problema uma população inicial de 100 indivíduos. Para a produção dos indivíduos da população inicial, os valores iniciais dos parâmetros projetados são distribuídos no espaço de solução de maneira mais uniforme possível por meio de cálculos que consideram números aleatórios em combinação com os espaços de busca de cada um dos parâmetros em questão. Cada cromossomo, C_i, da população inicial tem a forma $C_i = (W_i, Th_i, \sigma_i)$.

4.3. Função Objetivo

Conforme apresentado na seção 3, os resultados de classificação de um detector de bordas pode ser avaliado pelo espaço ROC. Assim, a técnica de avaliação quantitativa de detectores de bordas é utilizada nesta implementação como a função objetivo à qual cada indivíduo da população

é submetido. Portanto, dada uma imagem de entrada, cada indivíduo da população inicial é submetido ao detector de bordas FUNED e a imagem borda resultante é submetida à avaliação quantitativa. Dado o resultado desta avaliação, o qual é um ponto no espaço ROC, calcula-se a distância entre este ponto ao ponto (O,1), que é o resultado ótimo para uma detecção de bordas perfeita. Após os cálculos, os indivíduos são ordenados conforme sua aptidão, onde as menores distâncias representam as melhores soluções para o problema.

4.4. Seleção

Após a aplicação da função objetivo, seleciona-se os 45 melhores indivíduos da população corrente para o processo de reprodução, permitindo que estes indivíduos possam passar as suas características às próximas gerações. Essa escolha é feita por amostragem direta em que os indivíduos que possuem as menores distâncias do ponto ótimo são selecionados para gerar a nova população.

4.5. Cruzamento

A implementação do cruzamento está intimamente relacionada ao esquema de codificação empregado no problema, que no caso desta implementação é a codificação real. Portanto, neste trabalho adota-se a técnica de cruzamento de média simples, que consiste em gerar um novo cromossomo usando a média simples de dois cromossomos pais, conforme é mostrado pela equação 5:

$$C_1 = \frac{(P_1 + P_2)}{2}. \tag{5}$$

4.6. Mutação

Na fase de mutação é preciso estabelecer uma porcentagem da população a ser mutada. Para a taxa de mutação é preciso escolher o operador de mutação de maneira a manter a diversidade da população nos primeiros passos do algoritmo genético e evitar a precocidade. Nesta implementação adota-se a probabilidade de mutação adaptativa, conforme a equação 6, em que nos passos iniciais do algoritmo, tem-se uma taxa de mutação maior e a medida que o algoritmo caminha para a solução ótima o valor da taxa de mutação vai diminuindo.

$$P_m = \frac{\exp(-1.5 \times t/2)}{tam_{pop} \times \sqrt{L}} \tag{6}$$

em que t é a geração da iteração genética, tam_{pop} é o tamanho da população, $L = 3$ é o comprimento do indivíduo.

4.7. Elitismo

O elitismo é uma técnica utilizada para melhorar a convergência do algoritmo, em que se escolhe o melhor (ou melhores) indivíduos de cada geração, passando-o para a geração seguinte, sem que sofra cruzamento e mutação. Desta forma, mantém-se indivíduos cujas características são consideradas boas para a solução do problema. A utilização do elitismo faz com que o algoritmo convirja mais rápido, pois evita que os elementos mais aptos sejam perdidos ou modificados. Nesta implementação, são escolhidos os dois melhores indivíduos para formarem a próxima geração.

4.8. Critério de Parada

O algoritmo termina quando o melhor resultado obtido sobre as últimas gerações não se altera. Assim, para o algoritmo implementado, a diferença entre a menor distância do passo corrente e a menor distância do passo anterior deve ser menor do que 10^{-5} em 200 gerações.

5. Resultados e Discussões

Um conjunto de imagens sintéticas e suas respectivas imagens ideais foram utilizadas para a validação do sistema automático para seleção automática de parâmetros do detector de bordas FUNED. Os resultados obtidos até o momento mostram a viabilidade da técnica desenvolvida no sentido de auxiliar o usuário na tarefa automática de seleção dos parâmetros. Sabe-se que o comportamento dos algoritmos genéticos é influenciado pelos parâmetros utilizados, para tanto, se faz necessário ainda estudar de que forma os parâmetros estabelecidos na implementação afetam o comportamento dos algoritmos genéticos.

6. Conclusões

Este trabalho apresenta um novo método para seleção automática de parâmetros de detector de bordas. Os resultados da avaliação quantitativa podem ser usados como a função objetivo de um algoritmo genético que procura pelo melhor conjunto de parâmetros de acordo com a correspondência entre diferentes resultados de detecção de bordas. O propósito do algoritmo desenvolvido é melhorar o uso prático dos detectores de bordas paramétricos e ajudar com a incerteza associada com a seleção dos parâmetros. O método proposto é baseado em uma imagem *ground truth*, o que leva a necessidade da existência de uma imagem *ground truth* traçada manualmente. Para um conjunto de imagens com características semelhantes às imagens testadas, a técnica oferece uma maneira de ajustar os parâmetros automaticamente. Como trabalho futuro, pretende-se investigar uma maneira de construir mapas de

bordas denominados potenciais *ground truth*, produzidos por um conjunto de resultados de detecção por diferentes conjuntos de parâmetros. Neste caso, ao considerar um conjunto inicial de parâmetros do detector, de tamanho N por exemplo, pode-se utilizar essas informações para construir N diferentes resultados de detecção de bordas. Um pixel identificado como uma borda por todos os resultados de detecção terá uma grande chance de ser um pixel de borda, ou seja terá a mais alta correspondência a uma borda, e uma localização identificada como borda por poucos resultados de detecção terá baixa correspondência a uma borda. A teoria fuzzy será investigada para a modelagem deste problema, em que, a partir de diversos resultados de detecção de bordas, em que são variados os parâmetros do detector, constrói-se potenciais *ground truth*, o que torna a técnica de ajuste automático de bordas independente da existência de imagens ideais traçadas manualmente.

Referências

[1] I. A. G. Boaventura. *Números Fuzzy em Processamento de Imagens Digitais e suas Aplicações na Detecção de Bordas*. PhD thesis, Escola de Engenharia de São Carlos, USP, 2010.

[2] I. A. G. Boaventura and A. Gonzaga. Border detection in digital images: An approach by fuzzy numbers. In *Seventh International Conference on Intelligent Systems Design and Applications*, pages 341–346. IEEE Computer Society, 2007.

[3] I. A. G. Boaventura and A. Gonzaga. Edge detection in digital images using fuzzy numbers. *International Journal of Innovative Computing and Applications*, 2(1):1–12, 2009.

[4] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice Hall, 2nd edition, 2002.

[5] J. H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, 1975.

[6] A. Koren and Y. Yitzhaky. Automatic selection of edge detector parameters based on spatial and statistical measures. *Computer VIsion and Image Understanding*, 102:204–213, 2006.

[7] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, 1999.

[8] A. Santa Catarina and S. L. Bach. Estudo do efeito dos parâmetros genéticos sobre a solução otimizada e sobre o tempo de convergência em algoritmos genéticos com codificações binária e real. *Acta Scientiarum. Technology*, 2003.

[9] L. Xia Xinghua and W. Chengdong. Automatic parameters selection method of edge detector in thr unstruturec environment. In *Proceedings of the IEEE International Conference on Robotics and Biomimetics, Barcelona, Spain*, pages 1740–1743, 2008.

[10] Y. Yitzhaky. A method for objective edge detection evaluation and detector parameter selection. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 8:1027–1033, 2003.

Detecção automática de manchas solares utilizando técnicas de Visão Computacional

Costa, W. L.[(a)] (wander.lairson@gmail.com) – Santos, I.[(a)] (learsi.isr@gmail.com)
Spagiari, A. E. [(a)] (eduardo@spagiari.net) – Válio, A. [(a, b)] (avalio@craam.mackenzie.br)
Marengoni, M.[(a, c)] (mmarengoni@mackenzie.br)
Universidade Presbiteriana Mackenzie
[(a)] PPGEE – Programa de Pós Graduação em Engenharia Elétrica
[(b)] CRAAM – Centro de Rádio Astronomia e Astrofísica Mackenzie
[(c)] FCI – Faculdade de Computação e Informática

Abstract

This article describes the use of the Curto et al. algorithm [5] for sunspot detection on the solar disk applied to SOHO images. The importance of sunspot detection and its related difficulties are presented followed by a description of the Curto's algorithm as well as the changes performed in SOHO images. The results obtained with the Curto et al. algorithm [5] are compared with the published data by the NOAA Institute and SIDC. This comparison showed the results consistency, having reached a high data correlation with the observations. Furthermore, the changes made to the original algorithm seemed to improve the method robustness.

1. Introdução

A previsão do clima espacial busca elucidar diferentes processos físicos solares e suas consequências para a vida na Terra. Campos elétricos induzidos e correntes podem afetar a operação de sistemas terrestres, dentre eles, redes de transmissão de alta tensão, gasodutos, cabos de telecomunicações, sinalização ferroviária, sistemas de comunicação sem fio e satélites [1], os quais podem colocar em risco a saúde e vidas humanas.

A identificação e a classificação de manchas solares, incluindo sua localização, tempo de vida, contraste, entre outros, constituem elementos essenciais na modelagem da irradiância total durante o ciclo solar, os quais são requisitos para um estudo quantitativo.

A análise do comportamento das manchas solares também é utilizada no estudo de regiões ativas e na previsão da atividade de explosão solar [2]. Logo,0 o aumento substancial de dados contendo informações referentes às imagens solares, à detecção automatizada e à verificação de diferentes características de interesses, dentre outras aplicações, permitem uma previsão confiável da atividade solar e, portanto, da previsão de clima espacial [3].

As manchas solares são áreas escuras que crescem e decaem na fotosfera, camada mais baixa do Sol visível a partir da Terra. As manchas solares são mais escuras que os seus arredores porque são mais frias em relação à temperatura média da superfície solar. O seu aparecimento e desaparecimento ocorrem devido às mudanças subjacentes nos campos magnéticos que existem no Sol. A presença destes fortes campos magnéticos revela a existência de grandes quantidades de energia que potencialmente podem ser liberadas [4]. Geralmente, as manchas solares são primeiramente observadas como pequenos pontos escuros denominados poros, os quais podem se desenvolver como regiões de manchas solares, evoluindo em horas ou em dias. Ocasionalmente, quando um ponto se torna mais escuro e maior, ele pode se romper do ponto original, neste cenário, apresenta-se a razão para a contagem total de manchas solares e a classificação dessas em grupos [5].

Em física solar terrestre e, especialmente em geofísica, índices solares são de vital importância para avaliar o impacto potencial da atividade solar na Terra e em veículos espaciais. O número de manchas solares é o índice utilizado no estudo a longo prazo das variações de atividades solares, assim como nos estudos das relações solares terrestres. O número relativo de manchas solares foi definido por Rudolf Wolf [6], sendo $R = k(10g + s)$ onde, g é o número de grupos de manchas solares, s é o número total de 'pontos' em todos os grupos no disco visível, e k é um fator de correção de erros de observação em função do equipamento utilizado [7].

Em 2005 Zharkov, Zharkova e Benkhalil apresentaram técnicas para reconhecimento automático de manchas presentes no disco em imagens solares. Em 2007, Colak e Qahwaji discutiram a classificação automática das manchas solares em tempo real para previsão de atividade solar. No ano seguinte, em 2008, Curto, Blanca e Martinez apresentaram suas técnicas com o uso de morfologia matemática para detecção automática de manchas solares em imagens na faixa visível do disco solar, aplicando seus algoritmos às imagens do

observatório Ebro e comparando os resultados obtidos com os dados disponibilizados pelo SFC (*Solar Feature Catalog*), instituto europeu de controle de atividades solares.

Este trabalho apresenta uma versão modificada do algoritmo de detecção de Curto et al. [5] adaptado para as imagens solares na faixa visível do observatório SOHO (*Solar and Heliospheric Observatory*), denominado algoritmo Mack. As operações morfológicas relevantes para este artigo são brevemente discutidas na seção 2. A técnica de Curto et al. para descrição e agrupamento e as modificações feitas em seu algoritmo são descritas nas seções 3 e 4, respectivamente. A seção 5 apresenta os resultados da comparação dos dados com os apresentados por Curto et al. [5] para detecção, assim como expõe os resultados de atividade solar com os dados obtidos ao longo do ano de 2001, comparando-os com os dados oficiais dos Institutos NOAA (*National Oceanic and Atmospheric Administration*) e SIDC (*Solar Influences Data Center*). Por fim, a seção 6 apresenta a conclusão deste trabalho.

2. Operações Morfológicas

A Morfologia Matemática, no contexto de processamento de imagens, é utilizada para analisar e extrair características geométricas das imagens, além de permitir operações de pré e pós processamentos de imagem, como por exemplo, a filtragem morfológica. As operações morfológicas são baseadas na teoria dos conjuntos e, a seguir, serão apresentadas as operações relevantes para o algoritmo Mack. Embora a morfologia tenha sido inicialmente aplicada em imagens binárias, as operações aqui descritas foram aplicadas em imagens em nível de cinza [9].

2.1 Erosão e dilatação

Seja *f(x,y)* uma imagem bidimensional e *b* o elemento estruturante, a erosão de *f* por *b* é dada por:

$$f \ominus b(x,y) = \min_{(s,t)\in b_N}\{f(x+s,y+t)\} \quad (1)$$

Operação de erosão

A erosão de um ponto (*x*,*y*) é o valor mínimo dentre os pontos vizinhos de (*x*,*y*) que estão dentro do elemento estruturante *b*. A dilatação, por outro lado, sendo uma operação dual à erosão [8], tem o efeito de diminuir as partes mais claras da imagem. A dilatação é dada por:

$$f \oplus b(x,y) = \max_{(s,t)\in b_N}\{f(x+s,y+t)\} \quad (2)$$

Operação de dilatação

A dilatação de um ponto (*x*,*y*) é o valor máximo dentre os pontos vizinhos de (*x*,*y*) que estão dentro do elemento estruturante *b*.

2.2 Abertura e fechamento

Sendo f(*x*,*y*) a imagem e *b* o elemento estruturante, a abertura e fechamento são dadas, respectivamente, por:

$$f \circ b = (f \ominus b) \oplus b \quad (3)$$

Operação de abertura

$$f \bullet b = (f \oplus b) \ominus b \quad (4)$$

Operação de fechamento

A abertura é uma erosão seguida de uma dilatação, enquanto o fechamento é uma dilatação seguida de erosão. A abertura remove pequenas partes claras da imagem com mínima distorção das partes escuras. De modo análogo, o fechamento remove partes escuras com mínima distorção das partes claras.

2.3 Transformadas *top-hat* e *bottom-hat*

As transformadas *top-hat* e *bottom-hat* são geralmente utilizadas na remoção de objetos da imagem com base no elemento estruturante. A *top-hat* atua sobre objetos claros com fundo escuro, enquanto a *bottom-hat* atua sobre objetos escuros com fundo claro. Abaixo são apresentadas as descrições das transformadas *top-hat* e *bottom-hat*:

$$T(f) = f - (f \circ b) \quad (5)$$

Transformada *top-hat*

$$B(f) = f \bullet b - f \quad (6)$$

Transformada *bottom-hat*

3. Detecção de manchas solares

A superfície do Sol apresenta distribuição de estruturas com diferentes níveis de intensidade e padrões regulares indefinidos, conforme se apresenta na Figura 1.

Geralmente cada estrutura escura é considerada como uma mancha solar e, sua identificação, envolve um processo de segmentação de imagem, mostrado nas Figuras 1 e 2. Segundo Curto et al. [5], em um primeiro momento, existem três processos de segmentação: detecção de borda, crescimento de região e binarização. A técnica a ser descrita neste artigo, algoritmo Mack, utiliza o processo de binarização. Entretanto, ao aplicar um sistema global de binarização, este produz histogramas não particionáveis em decorrência da variação de

intensidade da imagem do disco solar, ou seja, o escurecimento do limbo solar [2]. Este fenômeno, o escurecimento do limbo solar, ocorre porque o brilho é máximo, na faixa visível, no centro do disco, ou seja, oriundo das camadas mais profundas da fotosfera, e decai aproximadamente 20% nas bordas em função da projeção de uma superfície esférica, logo, camadas mais superficiais, as quais são mais frias por isso menos brilhantes [4].

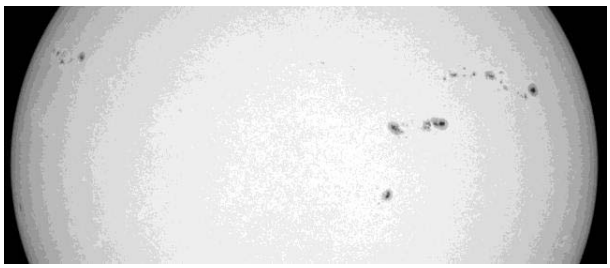


Figura 1. Imagem original com manchas, SOHO

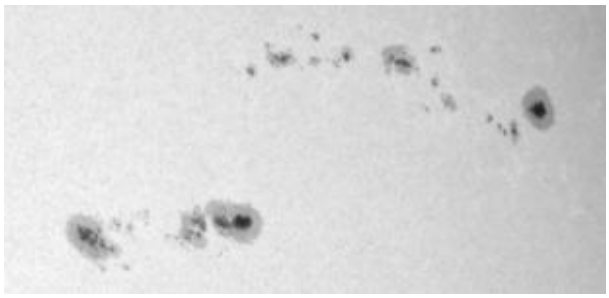


Figura 2. Manchas solares ampliadas

A técnica descrita por Curto et al. [5] consiste na aplicação de um procedimento utilizando a binarização iterativa, assim como em Zharkov et al. [2]. Contudo, em Curto et al. [5], as manchas solares candidatas são obtidas aplicando um procedimento que envolve somente dois laços iterativos na imagem original, sendo um incrementando o tamanho do elemento estruturante da operação de fechamento e o outro incrementando o nível de intensidade, enquanto o crescimento da população dos pixels da mancha solar está sendo controlado.

A detecção da totalidade de pixels verdadeiros pertencentes a uma mancha solar utiliza um método iterativo, o qual incrementa a escala de escopo a cada repetição. Novamente uma dificuldade surge. Existe uma larga variedade de tamanhos nas manchas solares. A priori, não se sabe o tamanho da maior mancha do disco solar em uma imagem em particular. Assim, as iterações devem continuar até a população de pixels a serem detectados estabilizar. Neste ponto, quando estabilizados, todos os pixels pertencentes às manchas solares estão detectados.

Para ser aplicada às imagens do observatório SOHO, a técnica de Curto et al. [5] foi modificada para eliminar ruídos e produzir maior robustez contra o efeito de escurecimento do limbo. Aplica-se a operação *bottom-hat*, Figura 3, em detrimento da operação *top-hat* utilizada por Curto et al. [5]. Este método inverte a imagem, sendo que o fundo se torna escuro e as manchas tendem ao branco. No processo original de binarização, inicia-se com um valor de corte que no início somente detecta os pixels mais escuros da imagem, e o valor de corte é iterativamente incrementado até ocorrer um crescimento abrupto no número de pixels, quando é detectado o fundo da imagem.

Figura 3. Imagem após operação *bottom-hat*

Figura 4. Imagem binarizada

A presente proposta é controlar os pixels detectados como pertencentes às manchas solares e interromper o processo quando o número de pixels a serem detectados estabilizar, ou seja, dado que o número de pixels com valor branco após a binarização na iteração n é P_n, então interrompe-se o processo quando P_n -1 $- P_n < q$, onde q representa um ponto de corte na taxa de decaimento de pixels. Diante disso, o processo finaliza quando a taxa de decaimento de pixels detectados tende a zero (Figura 4). Porém, impondo o valor zero, é possível que eventuais manchas que se sobreponham a alguns pontos de ruído no histograma de cor sejam eliminadas durante a binarização. Logo, a solução encontrada para este problema foi relaxar esta imposição, tornando q um valor maior que zero. Isso evita falsos negativos durante o processo de binarização, mas aumenta a probabilidade de ocorrência de falsos positivos. Entretanto, os resultados obtidos indicam que o aumento desta probabilidade de falsos positivos é, em geral, desprezível, mas a diminuição de falsos negativos é significante. Ruídos são em geral espacialmente menores que as manchas, logo, após o processo de segmentação completo, efetua-se uma abertura morfológica na imagem para eliminar ruídos remanescentes.

Após a binarização, o tamanho do elemento estruturante é incrementado e o processo se repete. O elemento estruturante de tamanho ótimo é determinado quando a população de pixels detectada se estabiliza [5] ou, para alguns casos observados, decresce. Curto et al. utilizaram um tamanho inicial de elemento estruturante 7x7. Para as imagens do SOHO foi utilizado neste trabalho um tamanho 3x3, com a âncora centralizada e formato de elipse. Para manter a âncora do elemento estruturante centralizada, os incrementos de tamanho são feitos em passos de duas unidades. No final, conforme mencionado, aplica-se uma abertura morfológica com tamanho do elemento estruturante igual a dois.

Em seguida à segmentação, há a necessidade de contar a quantidade de manchas detectadas. Neste trabalho, foi utilizado o algoritmo de crescimento de região para identificação dos *blobs* [9], sendo que cada *blob* tem seu centro geométrico calculado e todas as informações extraídas dos *blobs* são armazenadas em estruturas de dados vetorizadas, na qual cada *blob* possui um identificador (ID). O algoritmo completo de detecção de manchas está resumido na figura 6.

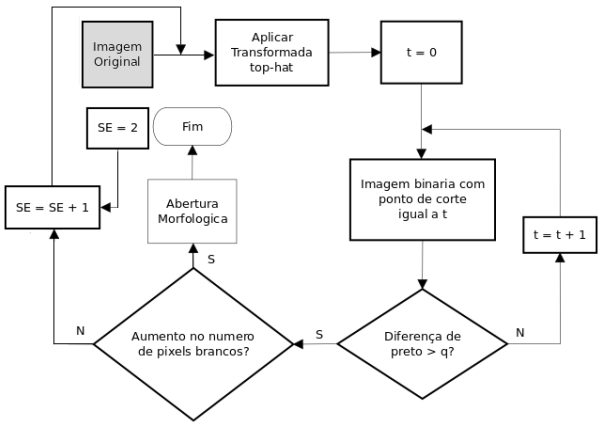


Figura 6. Algoritmo de detecção de manchas

O valor 2 para a abertura morfológica ao final foi determinado experimentalmente, sendo observado que tamanhos estruturantes maiores tendiam a eliminar manchas de tamanho reduzido. Diante disso, não foi utilizado um tamanho ímpar para o elemento estruturante.

Para este trabalho, utilizou-se o valor 40 para o parâmetro q, determinado experimentalmente. Uma vez detectadas as manchas, é necessário determinar sua localização na imagem. Para isso, conforme mencionado, aplica-se um algoritmo de detecção de *blobs* utilizando crescimento de região. O algoritmo funciona percorrendo a imagem linha a linha, procurando por pixels brancos e vizinhos conexos. Quando um pixel branco é encontrado, verifica-se os pixels imediatamente à esquerda e superior, na estrutura da imagem, para determinar se este pixel pertence a um *blob* já existente. Se não for, adiciona-se um novo *blob* ao conjunto de *blobs*. Os pixels pertencentes a cada *blob* são armazenados em uma estrutura de dados do tipo vetor para localização de manchas no algoritmo de agrupamento.

4. Agrupamento de manchas solares

No contexto de visão computacional, os resultados obtidos pelo algoritmo de detecção de manchas solares, tornam-se subsídios para um novo processamento fundamentado nos conceitos de física solar e técnicas de astrofísica, a fim de se determinar o número de grupos e, por sua vez, o número de Wolf.

As manchas solares formam grupos que compartilham propriedades físicas, tais quais as presentes na base de arcos coronais, os quais são estruturas curvilíneas alinhadas ao campo magnético [4].

O número de manchas individuais e o número de grupos são necessários para se determinar o nível de atividade solar, conforme o índice de Wolf [6]. Duas manchas solares fazem parte do mesmo grupo se estão localizadas próximas e compartilham do mesmo arco de fluxo magnético, para isso utiliza-se o critério de proximidade espacial [5]. Neste contexto, estatisticamente define-se quais manchas solares fazem parte do mesmo grupo se estiverem até seis graus heliográficos eqüidistantes [5].

Para se determinar as distâncias entre as diferentes manchas solares, torna-se necessário converter a imagem de coordenadas cartesianas para coordenadas heliográficas [8]. A detecção de grupos por proximidade se caracteriza por ser um processo computacionalmente simples e eficiente na classificação dos grupos, porém há certa dificuldade quando diferentes grupos estão dispostos próximos uns dos outros, principalmente nas extremidades do limbo solar.

Neste trabalho, define-se o raio e o centro do limbo solar utilizando a imagem original posicionada sobre um círculo maior. Iterativamente, por meio da detecção de transições abruptas, centraliza e ajusta o raio do círculo maior até suas características se aproximem das encontradas no limbo solar, respeitando uma tolerância (Δ) de no máximo três pixels. O algoritmo converge para a solução em no máximo três iterações.

Definindo-se os valores de centro e raio do limbo solar (r), transforma-se as coordenadas cartesianas (x,y) das manchas solares em coordenadas heliográficas. Aplica-se as equações 7 e 8 [8] para o centro de cada mancha, a fim de encontrar sua localização longitude e latitudinal na esfera solar, sendo que o meridiano principal e do equador solar são respectivamente as linhas que cortam o limbo solar na vertical e na horizontal em referência ao observador. Com a conversão das coordenadas, a distância em graus pode ser aferida por meio da aplicação

da equação 9 [8] nas manchas encontradas.

$$Lat = arcsen\left(\frac{y}{r}\right) \quad (7)$$

Latitude de um ponto

$$Lon = arcsen\left(\frac{x}{\sqrt{r^2 - y^2}}\right) \quad (8)$$

Longitude de um ponto

$$Lamb = \arccos(\sin(lat_1).\sin(lat_2) + \cos(lat_1).\cos(lat_2).\cos(lon_2 - lon_1)) \quad (9)$$

Distância entre dois pontos em uma esfera

Abaixo é apresentado o fluxograma do algoritmo de detecção de grupos de manchas.

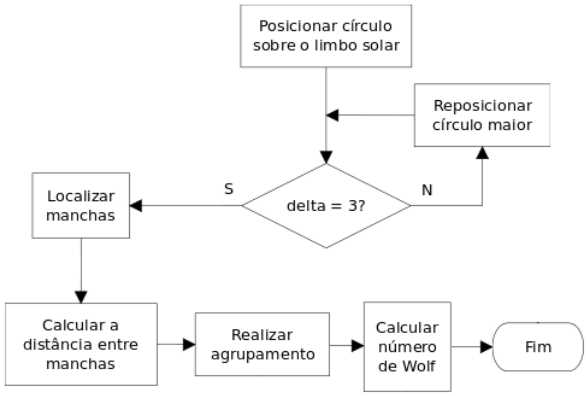


Figura 7. Algoritmo de detecção de grupos

5. Resultados

Este trabalho apresenta uma versão modificada do algoritmo de detecção de Curto et al. [5], implementado em linguagem C++, utilizando a biblioteca de algoritmos de Visão Computacional OpenCV, adaptado para as imagens solares na faixa do visível do observatório SOHO.

Curto et al. [5], para verificação dos resultados, aplicou seu algoritmo às imagens do observatório Ebro e comparou os resultados de detecção aos dados disponibilizados pelo *Solar Feature Catalog* (SFC), sendo que as imagens utilizadas datam de Maio de 2004. Neste trabalho foram utilizadas as imagens do satélite SOHO com as mesmas datas definidas por Curto et al. Convencionou-se para este trabalho que, quando eram disponibilizadas mais de uma imagem no mesmo dia no observatório SOHO, seria utilizada a imagem com o horário mais próximo ao meio dia. A Tabela 1 exibe os resultados desta comparação. Algumas imagens presentes na tabela original de Curto et al. [5] não estavam disponíveis no site do laboratório SOHO.

Tabela 1. Número de manchas detectadas.

Data	Mack	EBRO	SFC
04/05/04	8	9	9
05/05/04	11	9	3
06/05/04	9	9	4
08/05/04	5	3	7
09/05/04	5	4	6
10/05/04	6	9	7
18/05/04	19	17	24
19/05/04	17	17	17
20/05/04	14	15	17
21/05/04	14	15	13
24/05/04	26	28	13
27/05/04	14	27	12
28/05/04	12	12	6
29/05/04	12	11	18
30/05/04	12	12	22
31/05/04	13	13	10
Total	**197**	**210**	**188**

Nota-se, pelos resultados obtidos, que de forma geral os dados apresentados por Curto et al. [5] e este trabalho são consistentes. Em alguns pontos, por exemplo o dia 27/05/2004, Curto et al. obtiveram um resultado discrepante com o SFC, enquanto nós obtivemos um resultado mais próximo ao do SFC. Em outros casos, Curto et al. e este trabalho apresentaram discrepância com relação ao SFC, embora o grau de discrepância apresentado pelo algoritmo Mack seja ligeiramente menor que o de Curto et al.. O coeficiente de correlação entre os dados do laboratório Ebro, utilizando o método original de Curto et al. [5], e os resultados do SFC, foi de 46%. Utilizando as imagens do observatório SOHO com as alterações realizadas no algoritmo de Curto et al. [5], foi obtida uma correlação de 59%. A correlação entre os resultados do algoritmo Mack e o de Curto et al. [5] foi de 92%.

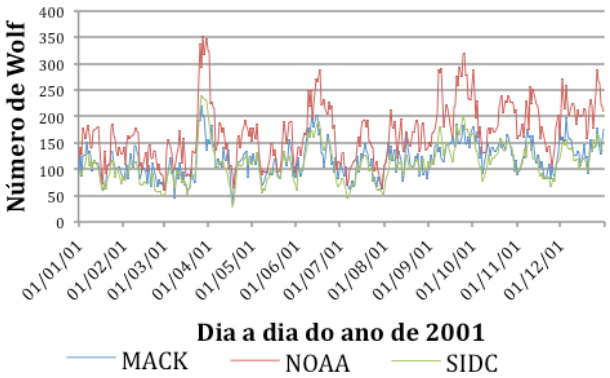


Gráfico 1. Comparação entre os algoritmos.

Finalmente, foram comparados os resultados de agrupamento utilizando trezentas e cinquenta imagens do ano de 2001 do observatório SOHO com os dados

divulgados pelos Institutos NOAA e o SIDC, em relação ao número de Wolf, os quais apresentaram correlação de aproximadamente 79% e 93%, respectivamente. O gráfico 1 apresenta a comparação dos resultados deste trabalho, com os resultados dos institutos NOAA e SIDC.

Nota-se, pelo gráfico, que o algoritmo deste trabalho detecta menos grupos em relação aos dados divulgados pelo Instituto NOAA, porém os resultados estão muito próximos aos apresentados pelo SIDC. Tais diferenças, em relação ao Instituto NOAA, podem ser vistas com maior intensidade nos resultados dos meses de Abril, Outubro, Novembro e Dezembro.

6. Conclusões

A aplicação de Visão Computacional para a automatização do processo de detecção de manchas solares gera uma ferramenta que objetiva facilitar o estudo do ciclo solar e suas influências na Terra, fornecendo subsídios relevantes para o estudo de previsão de clima espacial com a manipulação do numeroso acervo disponibilizado no Observatório SOHO.

A detecção de manchas solares com o apoio de ferramentas morfológicas foi utilizada para analisar e extrair características geométricas das imagens, mediante o uso de operações morfológicas de pré e pós processamentos. Tais operações são baseadas na teoria dos conjuntos, dentre elas foram utilizadas as de abertura e *bottom-hat* para o processamento da imagem. A operação de abertura foi utilizada para remover possíveis ruídos remanescente da imagem, após a aplicação do algoritmo de detecção de manchas. A transformada *bottom-hat* foi utilizada para separar as manchas solares do fundo da imagem.

Os resultados obtidos apresentaram robustez e demonstraram alta consistência com os resultados descritos por Curto et al. [5] e a correlação com os dados divulgados pelos institutos NOAA e SIDC apontaram que a adaptação realizada pode contribuir para a detecção automática de manchas solares com o algoritmo Mack desenvolvido neste trabalho.

Uma dificuldade encontrada foi a determinação de falsos positivos ou negativos, tanto na detecção de manchas quanto no agrupamento. Isto ocorreu em função da necessidade de acompanhamento de um profissional especializado em segmentação e agrupamento manual de manchas, para indicar em quais pontos o algoritmo falhou.

Para futuros trabalhos, sugere-se avaliar a possibilidade de agrupamento de manchas analisando o fluxo magnético das manchas candidatas através do magnetograma, análise da área das manchas e média de intensidade de brilho, conforme apresentado por Zhakov et al. [3], além de uma melhor análise dos parâmetros do algoritmo apresentado.

7. Referências bibliográficas

[1] T. Colak e R. Qahwaji, "Automatic Sunspot Classification for Real-Time Forecasting of Solar Activities". Congresso, ICEEE conference on Recent Advances in Space Technologies, Istanbul, Junho de 2007, pp: 733-738.

[2] S. Zharkov, V. Zharkova, S. Ipson, A. Benkhalil. "Technique for Automated Recognition of Sunspots on Full-Disk Solar Images". Revista, EURASIP Journal on Applied Signal Processing. Volume 2005, 1 January 2005, pp. 2573-2584.

[3] S. Zharkov, V. Zharkova, S. Ipson, A. Benkhalil. "Automated Recognition of Sunspots on the SOHO/MDI White Light Solar Images". Congresso, 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems. Nova Zelândia, Setembro de 2004.

[4] A. V. R. Silva "Nossa estrela: o Sol". Livro. 1a. Edição, Ed. Livraria da Física, São Paulo, SP, 2006.

[5] J. J. Curto, M. Blanca, E. Martínez. "Automatic Sunspots Detection on Full-Disk Solar Images using Mathematical Morphology". Revista, Solar Physics, Volume 250, Number 2, p. 411-429, Agosto de 2008.

[6] Wolf, J. R., Mittheilungen uber der Sonnenflecken, Astron. Mitt. Eidg. Sternwarte Zürich, 1856, pp13.

[7] Svalgaard, L.. "Objective Calibration of Sunspot Numbers". Congresso, SIDC - Securities Industry Dev elopment Corporation. Setembro de 2010.

[8] Green, R.M., "Spherical Astronomy", Livro. Cambridge University Press, Cambridge. 1985.

[9] Gonzalez,R.C., Woods R.C., "Processamento Digital de Imagens", Livro, 3ª edição, Editora Pearson Prentice Hall, São Paulo, 2010, PP. 415 a 447.

[10] V.V.Zharkova, S.S.Ipson, S.I.Zharkov, A. Benkhalil, J.Aboudarham e R. D. Bentley. "A full disk image standardisation of the synoptic solar observations at the Meudon Observatory". Department of Cybernetics, Bradford University, Bradford BD7 1DP, UK. LESIA, The Paris-Meudon Observatory, F92195 Meudon Principal Cedex, FRANCE. MSSL, University College London, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK.

The Power of Kullback-Leibler Divergence Extendable for Tsallis Entropy in Comparing Cone-Beam and Multi-Slice Tomography Images

A. Sobiecki and C. D. Gallão and D. C. Cosme and P.S. Rodrigues
Artificial Intelligence Group
Centro Universitário da FEI
São Bernardo do Campo, São Paulo, Brazil
sobiecki@fei.edu.br celsogallao@hotmail.com dcardosocosme@gmail.com psergio@fei.edu.br

Abstract

Image registration is a technique that compares two images for purposes of alignment and calibration. A well known problem underlining medical imaging is that the alignment of different exams for the same patient and same anomalies may not match. So we can not make reliable conclusions based on direct comparison of both exams. Generally, we can accomplish alignment of geometric features as well as luminance profile. Besides, there are a vast literature demonstrating their importance and challenges. The luminance alignment is close related to the amount of information between two images. Until the end of 90's the most known strategy to measure transfer information between two probability distributions of physical systems was through the classical Shannon entropy. A further improvement of this kind of formalism is now the so called Tsallis entropy, which has been proved to be a generalization of Shannon theory. The recent related literature has been shown that Tsallis entropy is suitable for medical imaging with many works proving its performance in several applications. This paper presents an analysis of the probability luminance distribution between images acquired from Cone-Beam and Multi-Slice techniques based on Kullback-Leibler divergence extendable for Tsallis statistics.

1. Introduction

Currently, it is vital to use images for evaluation and monitoring patients in the area of medicine [10], whose methods of acquisition and their technology have been so much improved. A common example is the Computed Tomography (CT), one of the most used and important test today. Different acquisition techniques are developed to improve the quality of the exam with small damage as possible to the patient. Despite their advantages, methods such as Multi-Slice Computerized Tomography (MSCT) have a high dose of radiation. In contrast, studies has shown that the use of techniques such as Cone-Beam Computerized Tomography (CBCT) can be less harmfull to the patient [8] [9]. Its importance comes from its frequent use in orthodontics. Then there is a need of evaluating differences between both techniques, taking into consideration the underlining mistakes or difficulties in diagnosis.

In order to minimize the differences between these two types of images, we first perform the registration between them, which consists of luminance and spatial alignment [12]. On the other hand, the spatial alignment can be determined under techniques of geometric transformations, such as in [16], [12],[13]. Once computed these transformation, the differences between the images can be computed by matching their luminance distribution. The most common techniques to accomplish this task is through matching their luminance histograms.

It is well known from Information Theory Area that the probability distribution of an image luminance can carry information related to its semantic content. Other features such as color, texture and spatial relationship between dominant regions also can carry similar information. However, due to the high degree of correlation between pixel features it can be difficult measuring such properties.

The traditional way to measure the amount of information between two images is by computing their relative entropies or Kullback-Leibler divergence. Recently, works of [1],[4], [13], [12] and [16] provided evidences that medical images can be better explained if their physical systems are considered as having non-extensive behavior, which means that they have long-range spatial and temporal interactions. This issue has been little explored in literature for image registration yet.

This paper presents a study of techniques based on non-extensive entropy for comparison of Cone-Beam (CB) and Multi-Slice (MS) CT images used in orthodontics. We carried out our studies under the Kullback-Leibler divergence

extendible for Tsallis entropy. The results show the power of this recent methodology for measuring relation between two probability distributions, and suggest that the pixels interactions of Cone-Beam and Multi-Slice may have sub-extensive behavior.

2. Non-Extensive Entropy

The term entropy appeared first in the field of thermodynamics, where it was used to demonstrate microscopic behaviors under macroscopic physical processes. Initially, it was considered merely a physical property that was thought to be applicable only in the context of thermodynamics. Later, E. Boltzmann and W. Gibbs showed entropy as a statistical measure for various other applications, resulting in the celebrated formula of Boltsmann-Gibbs, where the entropy (S) is the product of the Boltzmann constant (k) by the logarithm of a state W [18][4], given by the following equation:

$$S = k \log W \tag{1}$$

Later, C. Shannon gave an important contribution using the concept of entropy in the light of a new contexts, where in his work [17] he derived the celebrated Equation (2), known as Boltzmann-Gibbs-Shannon (BGS) entropy. So, p_i is the probability of finding the system in state i, and $P = [p_1, \ldots, p_k]$, $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$ [13].

$$S = -\sum_i p_i \ln(p_i) \tag{2}$$

Systems that can be described by the BGS entropy are called extensive systems and have the following important property, called additive property:

$$S(A * B) = S(A) + S(B) \tag{3}$$

where $S(A * B)$ is the entropy of a composed system of two independent variables, S(A) and S(B), calculated according to Equation (2). However, this formalism does not describe physical systems that behave as non-extensive systems, requiring a generalization. Then, Constantino Tsallis defined in his work [19], the following equation:

$$S_q(p_1, \ldots p_k) = \frac{1 - \sum_{i=1}^{k} p_i^q}{q - 1} \tag{4}$$

Tsallis entropy is a generalized formulation for Boltzmann-Gibbs-Shannon Entropy, where k is the number of possibilities of the system and q is the entropy index that characterizes the degree of non-extensivity. Note that, when $q \to 1$, Equation (4) meets the traditional Equation (2).

The additive property can explain better extensive systems, but fails in explains non-extensive ones, where Tsallis has proposed the following generalization, called pseudo-additive property:

$$S_q(A * B) = S_q(A) + S_q(B) + (1 - q) S_q(A) S_q(B) \tag{5}$$

Both additive and pseudo-additive equations are widely used in image thresholding methodologies for background extraction. However, achieving an optimal threshold is still a challenge. In [6] was proposed a threshold method using the BGS entropy. The work considers two probability distributions, one for the foreground and one for the background. Then, the optimal threshold is computed using the Equation(3). In [1] the methodology proposed by [6] was improved by using non-extensive formalism. Rodrigues in [16] proposed an algorithm for classification of ultra-sound images of breast cancers. His strategy has used a non-extensive entropy. Also in [16] was proposed the first algorithm for automatic computation of the q index, optimizing results.

In [4] are discussed and presented results based on non-extensive relative entropy, or kullback-Leibler Divergence, which is detailed in the following section.

3. Kullback-Leibler Divergence

The Kullback-Leibler divergence or relative entropy, as it is also known, is similar to BGS entropy. Furthermore it considers two probability distributions and computes the divergence between them. With Kullback-Leibler divergence we can measure the gain of information between two regions from the same images. The traditional Kullback-Leibler divergence is defined by the following equation:

$$D_{KL}(P : P') = \sum_i^n p_i \cdot \log \frac{p_i}{p_i'} \tag{6}$$

where P and P' are two probability distributions: $P = (p_1, p_2, \ldots, p_n)$ and $P' = (p_1', p_2', \ldots, p_n')$.

The Kullback-Leibler divergence can be computed using the non-extensive formalism proposed by Tsallis, that has been defined as [16]:

$$D_{KL}(P : P') = \sum_i^n \frac{p^q}{q - 1} \cdot (p_i^{1-q} - {p_i'}^{1-q}) \tag{7}$$

This equation measures the divergence between two probability distributions for non-extensive systems.

So similar to Equation (6), which measures the divergence of extensive systems, Equation (7) measures the divergence of those non-extensive. In the limit when $q \to 1$ it can be shown that Equation (7) meets Equation (6), then Equation (7) is a generalization of Equation (6).

The works of [1], [5], [14] and [15] have suggested that Tsallis statistics is a powerfull tool for medical image segmentation and the works of [3], [7], [2] and [11] have shown

that the Kullback-Leibler Divergence for non-extensive systems has promisor results when applied to the problem of image classification.

Usually one can not know a priori the behavior of systems as being extensive or non-extensive. The usual solution is then to compute the divergence for various q values including $q = 1$ and choosing q values with best performance. When the value is less than 1.0 it is said that the system has a sub-extensive behavior; when $q = 1$ is observed, it is said that the system is extensive; and when $q > 1$ occurs, it is said that the system is super-extensive [20].

In our work, we propose the use of extensible Kullback-Leibler divergence and investigate the q value that minimizes the amount of information between the two types of images of CT: Cone-Beam and Multslice, both taken for orthodontics.

4. The Proposed Methodology

As we are interested in investigate the power of Kullback-Leibler divergence as a tool for measure luminance information only, the comparison proposed here is carried out only under a range of gray scale $L = [0 : 255]$. Then, the complete Hounsfild scale for CT images was mapped for this range. Figure 1(left) shows an example of CB image and Figure 1(right) shows an example of its corresponding MS image.

Figure 3 shows the diagram of the proposed methodology. In step 1 we have non-normalized input images. The manually normalization is accomplished in step 2, where the captured images by the Cone-Beam method were just cut and maintained without resize or rotations. On the other hand, the Multi-Slice images were cut but rotated in order to meet the CB images as well as possible. Since the probability distributions are invariant to rotation as well as translation, these geometric tasks do not have any influence over our results and are accomplished in order to better visualization purposes only. For the same reason, there is no need to match their respective geometric centers. Figure 2 shows the corresponding normalized images (step 3 in the diagram).

In step 4, the histograms are computed in order to get the probability distributions. Both histograms are shown in Figure 4. We can observe that they are visually similar as also show Figures 2(left) and 2(right), respectively.

With the histograms of each image computed, we calculate the Kullback-Leibler divergence, accomplished in step 6.

5. Experimental Results

In our experiments, first we show the behavior of Tsallis entropy in comparing the information content for 8 both CB

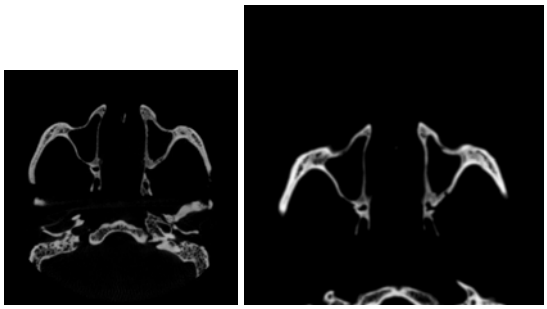


Figure 1. Example when using images in the tests before the normalization step. (left) image captured with the Cone-Beam technique; (right) image captured with the Multi-Slice technique.

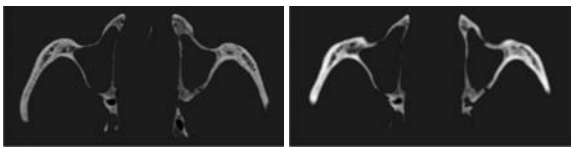


Figure 2. Example after normalization step. (left) normalized image captured with Cone-Bean technique; (right) normalized image captured with Multi-Slice technique.

and MS images. Table 1 shows the entropy difference between CB and MS images under increasing q values (0.1 to 0.9). Column 2 shows the entropy for CB, column 3 for MS and column 4 shows the corresponding relative difference. The lower the q value the larger is the non-extensive entropy intensity, enhancing their corresponding differences. That's because the smaller the q value the greater is the corresponding histogram powering, increasing the small histograms entries associated with low probability values.

In order to evaluate the robustness of Tsallis Entropy under changes of Signal-to-noise ratio (SNR), we have applied increasing gaussian noise with standard deviation varying from 2 to 14 over BC and MS images. These strategy simulates the cases when we have scattered histograms.

Figure 5 shows two images of CB, one with no noise (left) and other (right) which was applied a noise with standard deviation 14. Also, we have applied five values for q (0.1, 0.3, 0.5, 0.7, 0.9). In the graphic of Figure 6 and 7 we can see that for q lesser than 0.5 the value of entropy increases significantly; and for q greater than 0.5, approaching to 0.9, the entropy value decreases under increasing gaussian noise. It suggests that even under large gaussian noise,

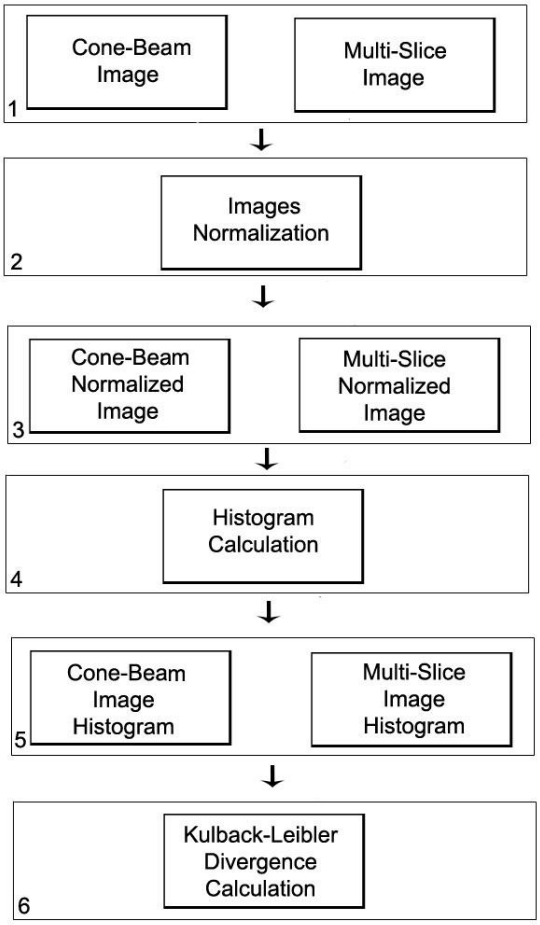


Figure 3. Proposed Methodology.

as that of Figure 5, the Tsallis entropy can generate observable values. Combining this information whit that from Table 1, we can conclude that the larger difference between two images occur for values around $q = 0.5$, suggesting a non-extensive system for the images used here. When we see the physical system as a traditional extensive system under severe gaussian noise, it is quite impossible to distinguish the differences between them.

The differences highlighted by Shannon and Tsallis entropies can be projected over their respective Kullback-Leibler divergencies. To show it we have applied Equation (7) to measure the relative information between CB and MS by interchanging the images. Table 2 shows the Kullback-Leibler divergence for CB under MS (column 1) and for MS under CB (column 2) showing its Euclidian distance in col-

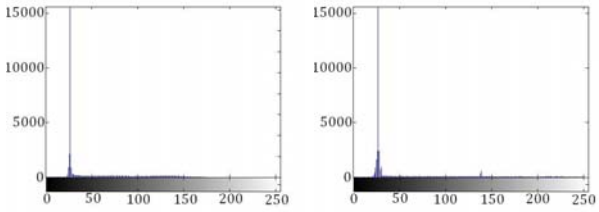


Figure 4. Histogram of normalized images. (left) Cone Beam histogram; (right) Multi-Slice histogram.

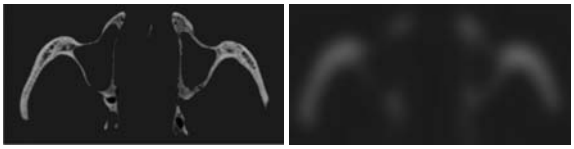


Figure 5. (left) Original image from Cone Beam; (right) image with gaussian noise with standard deviation 14.

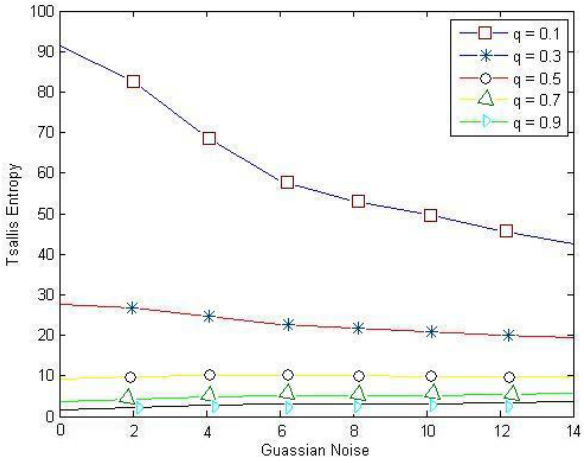


Figure 6. Tsallis entropy under increasing Gaussian noise for CB.

q	$s(mt)$	$s(cb)$	$\frac{\lvert s(mt),s(cb)\rvert}{max(s(mt),s(cb))}$
0.1	128.8339	91.3420	0.2910
0.2	68.3955	49.6353	0.2743
0.3	37.1733	27.6408	0.2564
0.4	20.7459	15.8134	0.2377
0.5	11.9378	9.3262	0.2188
0.6	7.1189	5.6942	0.2001
0.7	4.4240	3.6159	0.1827
0.8	2.8799	2.3984	0.1672
0.9	1.9711	1.6669	0.1543

Table 1. Table showing the differences in Tsallis Entropy between CB and MS images

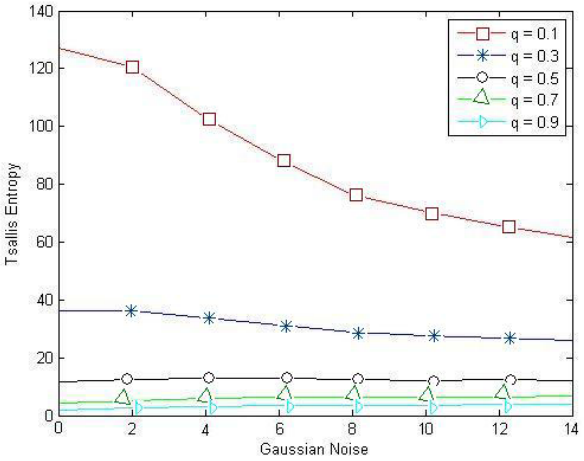


Figure 7. Tsallis entropy under increasing Gaussian noise for MS.

umn 3 for increasing q values. We can observe that as we approach to the traditional extensive systems by increasing q, greater is the Kullback-Leibler divergence. Then, we can conclude that, when images have significant differences and behave as non-extensive systems, it may be better to use the extendable Kullback-Leibler divergence in order to measure their relative content information.

6. Conclusions

This paper shows the sensitiveness of non-extensive Tsallis entropy in a comparative study between Cone-Beam and Multi-Slice CT images. By simple histogram matching it is possible to investigate the contribution of each bin related to each probability of the whole physical system but it is not possible to clearly see the histogram differences. Nevertheless, as the use of Cone-Beam exams has becoming popular and can substitute the Multi-Slice images for various exams, there is a growing need to investigate their differences. Then, the simple euclidian distance or traditional relative entropy may not enhance the main differences, at least when the important pixels are those related with the small probabilities.

q	$K(cb:mt)$	$K(mt:cb)$	Euclidian Diatance
0.1	0.3597	0.2862	0.4597
0.3	0.5642	0.5474	0.7861
0.5	0.9799	0.9508	1.3654
0.9	6.4464	6.2551	8.9823

Table 2. Kullback-Leibler Divergence with interchanged images. Column 4 shows the Euclidian distance between Columns 2 and 3 for different q values.

When we need to investigate the contribution of small probabilities the extendable Kullback-Leibler divergence may be the appropriate choice. The Kullback-Leibler divergence allows the fine-tuning of the entropic parameter q for matching two images as probability distribution histograms. This matching can computes the amount of relative information between samples and also allows the sensitiveness under noise and small probabilities when it has some demanded significance. In our results we show that the extendable Kullback-Leibler may be a powerful tool for investigate the differences of information under such situations, which is accomplished in step 6 of our methodology (Section 4).

We can note that, when the q parameter moves way from 1.0 towards 0 (zero) the entropy value becomes more observable. This behavior remained less invariant even under the strong presence of noise.

This behavior suggest that the Kullback-Leibler divergence is better used by considering the image as non-extensive systems; which means that the statistical states (or luminance probabilities) may have long-range spatial and temporal interactions, even for the small probabilities. It is interesting for systems where the small probabilities are semantically important.

The results presented here only reinforce those found in literature for other types of applications, requiring however, a larger investigation to prove its effectiveness in medical images.

References

[1] M. Albuquerque, M. Albuquerque, I. Esquef, and A. Mello. Image thresholding using tsallis entropy. *Pattern Recognition Letters*, 25:1059 –1065, 2004.

[2] M. P. de Albuquerque, M. P. de Albuquerque, and I. A. Esquef. Image segmentation using nonextensive relative entropy. *IEEE Latin America Transaction*, 6(5):477–483, 2008.

[3] H. Erdmann. Detecção e classificação de imagens sintéticas utilizando entropia não-extensiva. Master thesis, Centro Universitario da FEI, Sao Bernardo do Campo, Brazil, Dezembro 2009.

[4] I. A. Esquef. *Tecnicas de Entropia em processamento Digital de Imagens*. PhD thesis, Centro Brasileiro de Pesquisas Físicas, Dezembro 2002.

[5] G. A. Giraldi, P. S. Rodrigues, C. E. Thomaz, and E. Kitani. Statistical learning approaches for discriminate features selection. *Journal of Brazilian Computer Society*, 20:1–16, 2008.

[6] J. Kapur, P. Sahoo, and A. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29:273–285, 1985.

[7] V. Lessa. Classificação de imagens de ultrassom de câncer de mama baseada em informações híbridas utilizando teoria da informação. Master thesis, Centro Universitario da FEI, Sao Bernardo do Campo, Brazil, Dezembro 2010.

[8] M. Loubele, R. Bogaerts, E. V. Dijck, R. Pauwels, S. Vanheusden, P. Suetens, G. Marchal, G. Sanderink, and R. Jacobs. Comparison between effective radiation dose of cbct and msct scanners for dentomaxillofacial applications. *European Journal of Radiology*, 71:461 – 468, September 2009.

[9] M. Loubele, M. E. Guerero, R. Jacobs, P. Suetes, and D. van Steenberghe. A comparison of jaw dimensional and quality assessments of bone characteristics with cone-beam ct, spiral tomography, and multi-slice spiral ct. *The International Journal of Oral and MaxilloFacial Implants*.

[10] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2:1–36, March 1998.

[11] F. C. Olivio. Um modelo bayesiano com divergência de kulback-leibler estendida para reconhecimento de objetos 3d baseados em múltiplas visoes. Master thesis, Centro Universitario da FEI, Sao Bernardo do Campo, Brazil, Dezembro 2009.

[12] P. S. Rodrigues and G. A. Giraldi. Computing the q-index for tsallis nonextensive image segmentation. In I. C. Society, editor, *XXII Brazilian Symposium on Computer Graphics and Image Processing, SIBIGRAPI'09*, pages 232–237, 2009.

[13] P. S. Rodrigues and G. A. Giraldi. Improving the non-extensive medical image segmentation based on tsallis entropy. 2009.

[14] P. S. Rodrigues, G. A. Giraldi, and J. Suri. Implicit dual snakes for medical imaging. In *The 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, 2006.

[15] P. S. Rodrigues, G. A. Giraldi, J. Suri, and R. F. Chang. Non-extensive entropy for cad systems of breast cancer images. In I. C. Society, editor, *International Symposium on Computer Graphics and Image processing, SIBIGRAPI'06*, pages 121–128, Manaus, 8-11 October 2006.

[16] P. S. Rodrigues, C. E. Thomaz, G. A. Giraldi, M. D. Faria, and J. Suri. Computer aided diagnosis for breast lesions in ultrasound images under non-extensive tsallis entropy.

[17] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, January 2001.

[18] A. Tavares. Aspectos matemáticos da entropia. Master's thesis, Universidade de Aveiro, 2003.

[19] C. Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics*, 52:479–487, 1988.

[20] C. Tsallis. *Nonextensive Statistical Mechanics and its Applications*. Lecture Notes in Physics. Springer, Berlin, 2001.

Estudo Comparativo da Utilização de Filtros de Suavização para Reconstrução 3D de Superfícies

Sérgio Yoshimitsu Fujii[1] *sfujii@ufpr.br*
Lauro Wichert-Ana[2] *lwichert@fmrp.usp.br*
Paulo Mazzoncini de Azevedo Marques[2] *pmarques@fmrp.usp.br*
Lucas Ferrari de Oliveira[1] *lferrari@ufpr.br*

[1] *Universidade Federal do Paraná*
Programa de Pós-Graduação em Bioinformática, 81520-260 - Curitiba, PR

[2] *Universidade de São Paulo*
Faculdade de Medicina de Ribeirão Preto, 14048-900 - Ribeirão Preto, SP

Abstract

Computer Aided Diagnosis (CAD) tools have been widely used as a complementary method in medical diagnosis. Using Computer Graphics and Image Processing techniques, they give a complementary tool in addition to the typical side-by-side visualization. However, surface reconstruction associated with smooth filtering techniques generates output deformation. This issue compromises qualitative and quantitative evaluation, for instance, volumetric measurement and visual shape analysis. This paper proposes a volumetric measurement comparison among smooth filtering techniques applied to Marching Cubes generated surfaces. The combination of filters totalized 10 different samples. There were generated several distinct sized spheres to evaluate the behavior in each particular condition. Furthermore, a case study using real data from an epileptic patient was reported. The results revealed that small fragments are severely degenerated, negatively affecting the volumetric estimation. Thus, the filter's choice must be done carefully to avoid influencing the visualization and size evaluation of the lesion.

1. Introdução

Para pacientes em tratamento da epilepsia e que não se beneficiam do tratamento com medicamentos nos primeiros 12 meses (25 a 45% dos casos), a cirurgia é considerada uma terapêutica já que a opção não-invasiva não surte efeito. No mundo, estima-se que aproximadamente 50 milhões de pessoas tenham epilepsia. A questão da intratabilidade é um pré-requisito essencial para a indicação cirúrgica [4]. O diagnóstico médico muitas vezes se beneficia das informações complementares dos diferentes exames. Para análises das características cerebrais, a Ressonância Magnética (RM) fornece informação anatômica, enquanto a informação funcional pode ser obtida através de imagens de Medicina Nuclear, como a técnica conhecida por SPECT ("Single Photon Emission Computed Tomography", Tomografia Computadorizada por Emissão de Fóton Único). A subtração entre os SPECTs ictal (dentro da crise epiléptica) e o interictal (entre as crises) fornece informações extremamente relevantes sobre a localização do foco epiléptico ou zona epileptogênica (ZE). Essa técnica, visualizada em conjunto com o exame de RM, pode indicar anatomicamente a localização da(s) ZE(s) do paciente. Muitas técnicas foram desenvolvidas ao longo das décadas para executar esta tarefa de maneira eficiente e precisa [8][3][4][16]. A visualização tridimensional das ZEs facilita a localização da(s) área(s) ativadas durante a crise epiléptica, o que permite um melhor planejamento cirúrgico. Em imagens bidimensionais, os exames são colocados lado a lado, analisando-se os eixos axial, coronal e sagital. A reconstrução por superfícies, utilizando técnicas de computação gráfica, oferece noção espacial no auxílio ao diagnóstico. Aliada à reconstrução volumétrica, temos a possibilidade do cálculo do volume encefálico. Isso é possível pois toda a área dentro da superfície reconstruída faz parte da ZE e, com este valor, obtém-se mais um dado quantitativo. Porém, o cálculo do vo-

lume não é tarefa trivial e está propensa a erros devido a metodologia de reconstrução e aos filtros de suavização aplicados [2].

As ferramentas CADs ("Computer Aided Diagnosis") fornecem informações importantes para que o especialista tome a decisão mais apropriada em cada caso. Ou seja, servem como uma "segunda opinião"na tomada de decisão. Softwares como *Osirix* [13], *Viltrails* [1] e *3D Slicer* [10] realizam a visualização de imagens médicas utilizando técnicas de computação gráfica para a reconstrução 3D e processamento de imagens para a suavização do objeto reconstruído. Este tipo de ferramenta aumenta o grau de confiança do especialista na tomada de decisão. Porém, este tipo de aumento só será possível se a ferramenta fornecer os dados de forma correta e precisa. O ganho potencial da união da ferramenta com o conhecimento do especialista faz com que os sistemas CAD tenham um uso clínico bastante grande nos últimos tempos [5].

Neste trabalho é apresentado um estudo visando identificar quais os erros encontrados em alguns dos filtros comumente utilizados em ferramentas de CAD, visando contribuir na escolha da técnica que apresenta o melhor resultado no cálculo volumétrico.

2. Reconstrução 3D

Para a reconstrução de objetos 3D a partir de seções transversais, os métodos existentes são classificados em duas categorias: reconstrução volumétrica e reconstrução de superfície. A reconstrução volumétrica é representada por um conjunto finito de dados chamado voxel. O objeto é integralmente reconstruído nos três planos, formando uma matriz de voxels. Este método é interessante quando a reconstrução do objeto como um volume é necessária. Porém, devido à quantidade de dados a serem armazenados e processados, apresenta alto custo computacional. Na reconstrução de superfície, os dados são representados através de malhas poligonais, formadas a partir de contornos de fatias que apresentam valores volumétricos constantes [9]. Assim, o método gera apenas a superfície do objeto, de acordo com o parâmetro utilizado. Isovalores são pontos utilizados em dados volumétricos. O conjunto de isovalores formam a isosuperfície. O algoritmo Marching Cubes é a referência na literatura para reconstrução de isosuperfícies, sendo largamente utilizado em ferramentas de CAD.

2.1. *Marching Cubes*

Este algoritmo foi proposto por Lorensen e Cline [7] e é uma das técnicas mais utilizadas para a visualização de dados provenientes dos exames de TC e RM, pois os dados são do tipo amostrado. Essa característica do exame torna o algoritmo propício para a aplicação em imagens médicas. Este algoritmo baseia-se em dois passos: Localização da superfície correspondente ao valor especificado como parâmetro; Cálculo das normais nos vértices dos triângulos, a fim de criar uma superfície de alta qualidade visual. O Marching Cubes utiliza a técnica de divisão e conquista (divide and conquer) para localizar a superfície através de um cubo lógico, formado por oito pixels (quatro para cada fatia adjacente), onde cada pixel é representado por um vértice do cubo [7].

2.2. *Decimation*

O *Decimation* [12] foi desenvolvido a fim de diminuir o número de triângulos em uma malha sem alterar a topologia do objeto. Essa técnica é particularmente útil ao ser aplicada na superfície resultante do *Marching Cubes* devido ao seu elevado número de triângulos, o que dificulta o processamento de objetos em tempo real. O algoritmo segue três passos:

1. Caracterização da geometria e da topologia do vértice local;
2. Avaliação do critério de redução de triângulos (*Decimation*);
3. Triangulação do orifício resultante da remoção dos pontos preservando arestas e outras características importantes.

3. Filtros de Suavização

Em processamento de imagens, é comum encontramos valores que estão fora da frequência desejada, caracterizando artefatos indesejados. Para reduzir esse efeito, são utilizados filtros de suavização, que atenuam os picos de frequência e diminuem o ruído. Essa técnica é estendida para objetos tridimensionais obtendo-se os mesmos benefícios. No trabalho, foram utilizados os filtros Gaussiano, Laplaciano e Sinc Janelado.

3.1. Filtro Gaussiano

É um dos filtros mais utilizados para a remoção de ruídos, cuja função em três dimensões é definida na equação (1) [6].

$$g(x,y,z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2+z^2}{2\sigma^2}} \quad (1)$$

Na função, σ é o desvio padrão. Esse valor altera a largura da base da função gaussiana, tornando possível

determinar o nível de suavização da função. Pode-se considerar que a largura da base é relacionada com o tamanho espacial da função. Ou seja, quanto maior a largura, mais pontos serão utilizados na suavização, alterando a malha com maior intensidade.

3.2. Filtro Laplaciano

Ao ser aplicado em uma malha de triângulos, o filtro laplaciano reduz a curvatura da superfície, gerando uma suavização laplaciana. A equação (2) mostra a equação de Laplace para um ponto p na posição xi [Oliveira M].

$$\overrightarrow{x_i + 1} = \vec{x_i} + \lambda \vec{V_{ij}} = \vec{x_i} + \lambda \sum_{\forall j : 0 \leq j < n} \left(\vec{x_j} - \vec{x_i} \right) \quad (2)$$

Na equação, $x_i + 1$ é uma nova posição no ponto p, x_j são as posições dos pontos pj conectados a pi e é um valor especificado. Tipicamente, é um valor pequeno (ex.: 0,01) e a equação é processada repetidamente (entre 50 a 100 iterações). No entanto, o número excessivo de iterações pode gerar deformações na superfície.

3.3. Filtro de Sinc Janelado

Os filtros de sinc baseiam-se na implementação de um filtro passa-baixas ideal utilizando a função de sinc (3) [14].

$$sinc\,(x) = \frac{sen\,(\pi x)}{\pi x} \quad (3)$$

No entanto, essa função é definida entre limites infinitos e, para ser representado computacionalmente, é necessário truncar os valores limites. Esse procedimento gera valores instáveis no domínio da frequência. Para atenuar a instabilidade, deve-se aliar a função de sinc a uma janela de truncamento. A janela de Hamming [15], utilizada neste trabalho, apresenta boa atenuação na frequência de passagem. A suavização gerada por esse filtro oferece a manutenção topológica da superfície, porém, com alto custo computacional.

4. Metodologia de Teste

4.1. Imagem com Esferas

A metodologia utilizada nos testes compreendeu os seguintes critérios: erro volumétrico, qualidade da superfície e desempenho no cálculo de estruturas. Os testes foram realizados em um equipamento com a seguinte especificação: processador AMD Phenom X3 710 2,6 GHz, 4 GB de memória RAM DDR2 800MHz. Foram montadas 8 imagens com fundo preto em formato Analyze e a elas foram acrescentadas esferas totalmente brancas com os seguintes tamanhos de raio: 3,5, 5,0, 7,5, 10,0, 15,0, 20,0, 25,0 e 30,0 mm. As esferas foram reconstruídas utilizando a combinação de filtros conforme a Figura 1.

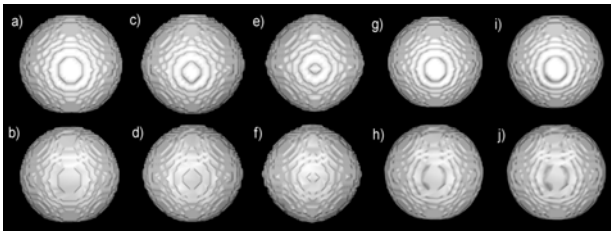


Figura 1. Esferas com raio de 15 mm em formato Analyze reconstruídas através do Marching Cubes com e sem aplicação de filtros: a)Sem filtro b)Decimate c)Gaussiano com fator de raio 0,5 d)Gaussiano com fator de raio 0,5 e Decimate e)Gaussiano com fator de raio 1,0 f)Gaussiano com fator de raio 1,0 e decimate g)Laplaciano h)Laplaciano e Decimate i)Sinc Janelado j)Sinc Janelado e Decimate

Com os tamanhos das esferas controlados e conhecidos, o cálculo do volume foi feito somando-se todos os pontos do interior de cada esfera e multiplicando-os pelo volume do voxel da imagem (1 mm^3). Com esta metodologia, foi possível saber o volume exato de cada uma das esferas criadas. A partir destes valores, foi feita a comparação entre as respostas do cálculo do volume após a reconstrução por superfície através do algoritmo Marching Cubes e a utilização de filtros de suavização para melhorar a visualização do objeto. Os filtros utilizados foram: Gaussiano, Laplaciano e Sinc Janelado com interpolação na janela de Hamming [11], para suavização, e Decimation para diminuição do número de triângulos na malha. Em todos os filtros foram utilizados valores padrão da biblioteca VTK. Porém, em um caso específico (filtro Gaussiano) foram utilizados os valores 0,5 e 1,0 para o fator de raio. No VTK, esse valor é alterado através do método *setRadiusFactor* da classe *vtkImageGaussianSmooth*.

4.2. Caso Clínico

Um paciente utilizado no estudo de [16] foi analisado através da metodologia de teste descrita na subseção

Raio (mm)	Gaussian 1.0 + Decimation	Gaussian 0.5 + Decimation	Laplacian + Decimation	Sinc Janelado + Decimation
3,5	100,00%	89,46%	22,83%	8,00%
5,5	93,26%	64,66%	12,94%	8,61%
7,5	81,61%	50,54%	8,10%	4,09%
10,5	65,68%	37,58%	5,57%	4,63%
15,5	48,65%	26,68%	2,68%	2,22%
20	39,45%	21,14%	1,42%	1,19%
25	32,47%	17,15%	0,97%	0,58%
30	27,54%	14,40%	0,71%	0,41%
Erro Médio	61,08%	40,20%	6,90%	3,72%

Tabela 1. Comparação dos erros gerados entre os filtros de suavização com Decimation.

4.1. O alinhamento e o processamento das imagens de RM e de medicina nuclear (SPECT) podem ser encontrados no referido trabalho. Para testar a influência dos filtros no cálculo volumétrico em imagens do tipo SISCOM, foi realizada a reconstrução por superfícies das estruturas como mostra a Figura 2. Foi utilizada a soma dos valores da média e mais dois desvios padrão da técnica descrita por [4] como valor de *threshold* para binarizar a imagem e possibilitar a contagem dos voxels brancos. Desse modo, a contagem é multiplicada pelo volume do voxel das imagems de RM, sendo utilizadas como padrão no cálculo do volume. Com esta metodologia, temos o volume real das imagens segundo um *threshold* (com contagem de voxels) e comparamos este valor com os obtidos pelo cálculo de volume após as reconstruções com e sem os filtros descritos anteriormente.

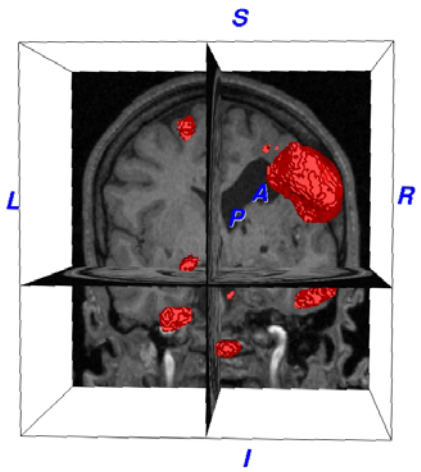


Figura 2. Reconstrução da superfície em um caso clínico utilizando a soma dos valores da média mais dois desvios padrão.

5. Resultados

5.1. Erro Volumétrico das Esferas

As Figuras 3 e 4 mostram o erro percentual gerado pela aplicação dos filtros de suavização com e sem decimation, respectivamente.

A Figura 3 mostra o erro percentual gerado somente com a aplicação dos filtros de suavização. O erro diminui à medida que o tamanho da esfera aumenta.

A Figura 4 mostra o erro percentual entre os volumes das esferas com a aplicação do algoritmo Decimation e os filtros de suavização. Nas esferas de menor raio, o erro é maior com o filtro Decimation.

5.2. Caso Clínico

Para testar o impacto dos filtros na reconstrução por superfícies em casos clínicos, foi selecionado um caso proveniente do trabalho de [16] que utiliza a metodologia proposta em [4]. Os casos foram escolhidos de forma aleatória. Neste caso fizemos o cálculo volumétrico das lesões epilépticas apontadas pelo B.R.A.S.I.L. [4]. e utilizamos os mesmos valores de referência (média + 2 desvios padrão) para as reconstruções e aplicações dos filtros. A Tabela 2 mostra os valores calculados para os seguintes parâmetros: volume total, erro volumétrico e tempo de processamento.

O filtro Decimation, em combinação com os filtros Laplaciano e Sinc Janelado, são os únicos casos em que há aumento no volume final. Ainda, nos filtros de suavização citados, o tempo de processamento é em média 2,5x maior em comparação com os demais casos.

6. Discussão

Os resultados mostram que, dependendo do filtro utilizado e do tamanho da estrutura reconstruída, o

	Volume (mm^3)	Erro	Tempo (s)
Original	47638	-	-
Marching Cubes	36053	24,32%	1,193
Decimation	38612	18,95%	2,128
Gaussian 0.5	22613	52,53%	1,387
Gaussian 0.5 + Decimation	22355	53,07%	1,768
Gaussian 1.0	17180	63,94%	1,425
Gaussian 1.0 + Decimation	16932	64,46%	1,660
Laplacian	34318	27,96%	4,458
Laplacian + Decimation	63253	32,78%	4,497
Sinc Janelado	34644	27,28%	4,373
Sinc Janelado + Decimation	75567	58,63%	6,068

Tabela 2. Volume total, erro volumétrico e tempo de processamento utilizando a média mais dois desvios padrão.

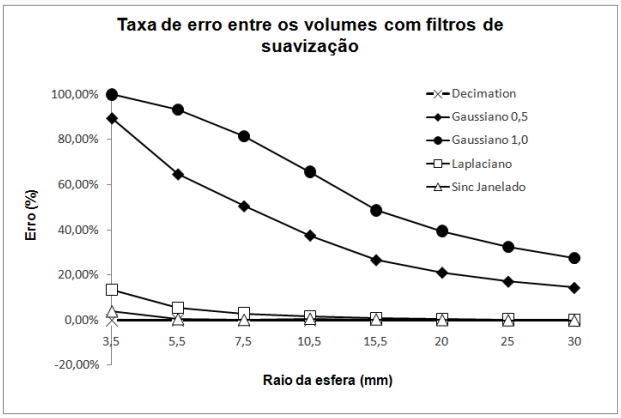


Figura 3. Taxa de erro entre os volumes com filtros de suavização.

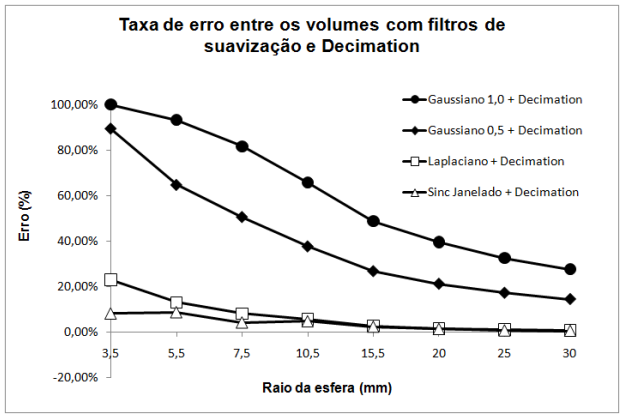


Figura 4. Taxa de erro entre os volumes com filtros de suavização e Decimation.

cálculo do volume fica comprometido. No teste realizado, o sinc janelado apresenta a menor taxa de erro. O filtro laplaciano gera erro semelhante ao sinc janelado em amostras maiores. No entanto, em esferas de raio inferior a 7,5 mm (176,71 mm^3), a diferença em relação ao volume original é maior que 8%. O filtro gaussiano apresenta a maior taxa de erro, variando de 14% a 100%. Na amostra de raio 3,5 mm com fator de raio igual a 0,5, o erro é de 89%. Esse valor apresenta proporção inversa com o tamanho da esfera, de modo que o erro com a amostra de maior raio (30 mm) é de 14%. O filtro de Sinc Janelado apresenta a melhor aparência visual nas esferas. Em contrapartida, seu alto custo computacional, como mostra a Tabela 2, pode se tornar um fator determinante na sua escolha, sendo necessário avaliar o custo-benefício em aplicações candidatas.

No caso clínico, o erro no cálculo de volume mostrou-se maior do que na metodologia de testes utilizando esferas. As principais causas são a presença de estruturas de tamanho pequeno, assim como a sua irregularidade na reconstrução. Como visto nos resultados, as estruturas pequenas são mais suscetíveis à deformação, apresentando diminuição no volume final. Outro fator é a irregularidade das estruturas reconstruídas que faz com que pequenas partes nas bordas da reconstrução sejam perdidas, comprometendo o cálculo volumétrico. O uso do filtro Decimation altera significativamente o volume quando combinado com os filtros Laplaciano e Sinc Janelado, conforme a Tabela 2. No caso do filtro Decimation, os valores de volume aumentaram após a aplicação dos filtros, o que pode ser observado em todos os ca-

sos da Tabela 2. Isso ocorre, provavelmente, devido à sua característica de reformular a malha de triângulos. Assim, algumas extremidades podem ter o seu tamanho original alterado, resultando em um aumento no volume final. Porém, essa característica não foi observada no caso das esferas. Isso nos leva a crer que, devido à natureza irregular do volume clínico em algumas regiões, o filtro Decimation afeta a topologia do objeto reconstruído.

7. Conclusão

Este trabalho realizou um estudo comparativo no cálculo de volume em reconstrução de superfícies utilizando diferentes filtros. Após os testes, foi observado o comportamento gerado por cada filtro nos casos utilizados. O experimento com esferas analisou o impacto dos filtros em superfícies uniformes. No caso clínico, foi visto que a aplicação dos filtros em pequenos fragmentos e em superfícies irregulares influi negativamente no cálculo do volume.

A reconstrução de imagens utilizando Marching Cubes e filtros de suavização é consolidada na área médica. No entanto, o volume gerado é um fator que deve ser levado em conta. Geralmente, tal valor é obtido diretamente dos dados brutos do exame. A coerência desse valor com o volume da reconstrução é importante na avaliação médica, visto que tais informações são utilizadas no diagnóstico de procedimentos cirúrgicos precisos como, por exemplo, a epilepsia. Assim, a utilização de filtros deve ser cuidadosamente avaliada pelo usuário, caso a reconstrução 3D seja um importante requisito na obtenção do diagnóstico.

8. Agradecimentos

Os autores agradecem ao CNPq (projeto 567035/2008-5) e à CAPES pelo apoio financeiro.

Referências

[1] L. Bavoil, S. P. Callahan, P. J. Crossno, J. Freire, C. E. Scheidegger, C. T. Silva, and H. T. Vo. VisTrails: Enabling interactive Multiple-View visualizations. In *VIS '05: IEEE Visualization*, pages 135–142.

[2] M. R. Baxa P, Skala V. An error estimation for isosurfaces. *Proceedings of EDU COMPUGRAPHICS*, pages 202–211, 2007.

[3] L. F. de Oliveira. *Fusão de Imagens 3D para Localização e Quantificação de Zona Epileptogênica, Brasil.* Tese de doutorado, Faculdade de Medicina de Ribeirão Preto - USP, Ribeirão Preto, Ribeirão Preto, Brasil, 2005.

[4] L. F. de Oliveira, P. M. de Azevedo-Marques, L. Wichert-Ana, and A. C. Sakamoto. Brain registration and subtraction - improved localization for spect analysis (b.r.a.s.i.l.): a computer-aided diagnosis in epilepsy tool kit. In *Proceedings of the 2008 ACM symposium on Applied computing*, SAC '08, pages 1390–1394, New York, NY, USA, 2008. ACM.

[5] K. Doi. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31:2007.

[6] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide.* Kitware, Inc. ISBN 1-930934-15-7, http://www.itk.org/ItkSoftwareGuide.pdf, second edition, 2005.

[7] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21(4):163–169, July 1987.

[8] T. J. O'Brien and L. So. Subtraction ictal spect coregistered to mri improves clinical usefulness of spect in localizing the surgical seizure focus. In *Neurology*, volume 50, pages 445–54, 1998.

[9] A. Paiva, R. Seixa, and M. Gattass. *Introdução à Visualização Volumétrica.* White paper, Pontifícia Universidade Católica - PUC-RIO, Rio de Janeiro, Brasil, 1999.

[10] S. Pieper, M. Halle, and R. Kikinis. 3d slicer. pages 632–635. IEEE International Symposium on Biomedical Imaging ISBI 2004, 04 2004.

[11] W. Schroeder, K. M. Martin, and W. E. Lorensen. *The visualization toolkit (2nd ed.): an object-oriented approach to 3D graphics.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1998.

[12] W. J. Schroeder, J. A. Zarge, and W. E. Lorensen. Decimation of triangle meshes. In *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 65–70, New York, NY, USA, 1992. ACM.

[13] E. Sierra-Martínez, R. Cienfuegos-Monroy, and G. Fernández-Sobrino. Osirix, a useful tool for processing tomographic images in patients with facial fracture. *Cirugia y cirujanos*, 77:95–9, 2009 Mar-Apr 2009.

[14] S. W. Smith. *The scientist and engineer's guide to digital signal processing.* California Technical Publishing, San Diego, CA, USA, 1997.

[15] G. Taubin, T. Zhang, and G. H. Golub. Optimal surface smoothing as filter design. In *Proceedings of the 4th European Conference on Computer Vision-Volume I - Volume I*, ECCV '96, pages 283–292, London, UK, 1996. Springer-Verlag.

[16] L. Wichert-Ana, P. de Azevedo-Marques, L. Oliveira, R. Fernandes, T. Velasco, A. Santos, D. Araújo, M. Kato, M. Bianchin, and A. Sakamoto. Ictal technetium-99 m ethyl cysteinate dimer single-photon emission tomographic findings in epileptic patients with polymicrogyria syndromes. *European Journal of Nuclear Medicine and Molecular Imaging*, 35:1159–1170, 2008. 10.1007/s00259-007-0655-3.

Algoritmo Genético aplicado ao estudo de imagens de gel de eletroforese bidimensional

Juliana Helena Tibães[1]
tibaes.juliana@ufpr.br

Vanely de Souza[1]
vanelysouza@ufpr.br

Aline Ariane Schultz
alineschultz@gmail.com

Angélica Inajá Juliani
angelica_juliani@hotmail.com

Letícia Fernandes de Jesus
leticiafdejesus@gmail.com

Michelly Gehlen[1]
cegempac.michelly@gmail.com

Roberto Tadeu Raittz[1]
raittz@ufpr.br

[1]*Universidade Federal do Paraná,*
Programa de Pós-Graduação em Bioinformática
81520-260, Curitiba, PR, Brasil

Abstract

Digital imaging studies of two-dimensional gels originated from electrophoresis analysis allow the identification of proteins related to diverse biological processes. However, these images have noise that may be confused with protein spots. In the present work we report a software able to carry out pre-processing of two-dimensional digital images minimal noise found. This project was divided in five phases: Initial Image Analysis, Detection of protein spots that will be handle, Image partition, Noise elimination, Correction of errors of protein spots and image resize. The system was developed using genetic algorithms and the JGAP library. System's users can save the images and it can be used by other software as Image Master. Furthermore, its user-friendly interface allows researchers to explore all application's features in a very intuituve way. As result, the image processing will clear most of the noisy features, maximizing, therefore, the detection of real protein spots in comparison to pre-existing computational solutions.

Keywords: electrophoresis two-dimensional gel, digital imaging, genetic algorithms

1. Introdução

O desenvolvimento das tecnologias "Ômicas" (genômica, transcriptômica, proteômica e metabolômica) tem o objetivo de dar uma visão geral sobre os sistemas biológicos e tem influenciado a biologia [3].

A proteômica estuda as proteínas expressas nos organismos com o objetivo de documentar a distribuição geral de proteínas nas células, identificar e caracterizar proteínas individuais de interesse e principalmente elucidar as suas associações e funções [4]. Segundo [8], muitos desses estudos envolvem comparações de eletroforese em gel bidimensional para identificar alterações da expressão da proteína entre diferentes amostras. Para isso, é necessário uso de softwares específicos a fim de realizar a análise das imagens providas do processo de eletroforese em gel bidimensional, como o ImageMaster^{TM} 2D Platinum[1], Gellab [6], THYCO [1] e Melanie [2].

A análise da imagem digital do gel bidimensional consiste em primeiro detectar as manchas de proteínas, conhecidas por *spots*, em cada uma das imagens a serem utilizadas no estudo. No entanto, essas imagens também possuem ruídos que dificultam a detecção dos *spots* por serem semelhantes ou muito próximos a esses. Depois, é necessário comparar as imagens a fim de identificar quais *spots* são os mesmos entre elas para, por exemplo, calcular as modificações das expressões gênicas e realizar interpretações estatísticas.

Com relação à etapa de detecção de *spots*, os softwares disponíveis no mercado não são capazes de diferenciar esses *spots* dos ruídos ou rastros de proteínas

1 Disponível em: http://www.gelifesciences.com/aptrix/upp01077.nsf/content/Products?OpenDocument&parentid=976024&moduleid=165660&zone=Elpho

com precisão. Isso faz com que o pesquisador necessite identificá-las manualmente, atrasando em algumas semanas as outras etapas da análise.

Neste artigo, apresentaremos uma ferramenta computacional capaz de realizar o pré-processamento das imagens de géis bidimensionais, com o objetivo de eliminar os ruídos presentes nessas imagens através da técnica de Inteligência Artificial, o Algoritmo Genético.

Os Algoritmos Genéticos (AG) são técnicas inspiradas no princípio Darwiniano de seleção natural e reprodução genética [5]. Nesse contexto, [7] afirma que o algoritmo criado, com essas técnicas, é capaz de resolver problemas de busca complexos os quais não poderiam ser resolvidos com eficiência pelos métodos convencionais.

Este artigo está dividido em: (i) materiais e métodos, a qual descreve quais as ferramentas e como elas foram utilizadas durante o desenvolvimento do software; (ii) resultados e discussão, expondo quais foram as melhoras na imagem bidimensional; (iii) conclusão, apresentando quais as vantagens e limitações.

2. Material e Métodos

2.1. Amostras de imagens digitais de géis bidimensionais

Analisamos as imagens vindas de eletroforese bidimensional provenientes de pesquisa proteômica, correspondentes a géis de agarose. Nos géis, a formação de *spots* é baseada em duas propriedades das proteínas: ponto isoelétrico e peso molecular.

As imagens dos géis contendo esses *spots* foram utilizadas como objetos de estudo e foram submetidas a um pré-processamento. Isso permitiu sua utilização por pesquisadores na obtenção das informações necessárias sobre proteínas expressas correspondentes aos *spots*.

2.2. Desenvolvimento do software

A análise da ferramenta computacional proposta foi realizada seguindo as especificações da padronização UML, com a ferramenta Jude. E para o desenvolvimento foi utilizado o modelo *Extreme Programing* (XP) e utilizando a linguagem de programação JAVA importando a biblioteca Java Genetic Algorithms Package (JGAP) para estruturar o desenvolvimento de Algoritmos Genéticos.

Foi abordada uma metodologia de desenvolvimento de software dividida em quatro etapas: (i) análise inicial da imagem, (ii) detecção dos *spots* na imagem, (iii) eliminação dos ruídos e (iv) tratamento de erro de detecção dos *spots* que serão descritas a seguir.

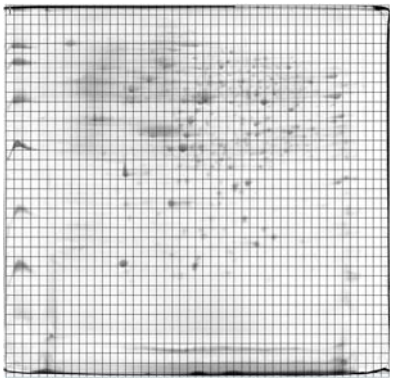


Figura 1. Imagem digital do gel de eletroforese bidimensional representando a segmentação de 50 por 50 pedaços

(i) Análise inicial da imagem

A Figura 1 apresenta um gel bidimensional contendo diversos *spots*. Inicialmente ela é dividida em pedaços de 50 linhas por 50 colunas proporcionais ao seu tamanho, gerando os indivíduos da população. O conjunto de *pixels* (genes) de um indivíduo corresponde ao seu cromossomo.

(ii) Detecção dos *spots* da imagem

Cada um dos indivíduos apresentados anteriormente é classificado para o processo de cruzamento. A Figura 2 mostra o caminho dessa função para atribuir uma nota aos indivíduos gerados e que será usada para seleção desses.

A função de avaliação (*fitness*) percorre a matriz de *pixels* (cromossomos), pega o valor RGB do *pixel* central e compara com cada valor RGB dos *pixels* da matriz. Nesse processo, é contado o número de *pixels* com RGB igual ao central, e o valor obtido é dividido pelo tamanho da imagem, gerando a nota correspondente ao indivíduo. De acordo com esse processo, os *spots* recebem notas mais altas que os ruídos.

Na execução da função, a população inicial é gerada aleatóriamente com tamanho fixado em 2500 indivíduos, e passam pelo processo de evolução 100 vezes e os melhores indivíduos são considerados os *spots*.

(iii) Eliminação dos ruídos

Com o resultado obtido no processo anterior, os *pixels* do indivíduo com valores RGB diferentes do valor RGB correspondente ao *spot* identificado são avaliados por outra função de *fitness*. Nessa função, os valores RGB desses *pixels* são subtraídos de valores de filtros 255 e 144,escolhidos por

Requisitos	Ferramenta computacional proposta	ImageMaster^{TM} 2D Platinum
Facilidade de uso	Apresenta interface intuitiva e vídeo tutorial, o que facilita o aprendizado do uso do software	Interface complexa, o que dificulta o aprendizado do software.
Multiplataforma	O software pode ser executado nas seguintes plataformas: Windows, Linux/UNIX, Solaris. O que possibilita o uso do software para outras pesquisasnas plataformas testadas.	O software só é executado na plataforma Windows.
Detecção automática e manual dos spots	Possui taxa de acerto de 90,3% dos spots, podendo modificar as marcações das localizações manualmente.	Possui taxa de acerto de 72,1% dos spots, podendo modificar as marcações das localizações manualmente.
Viabilidade de utilizar as imagens pré-processadas	O software é capaz de exportar imagens nos formatos TIFF, JPEG, PNG e BITMAP, sendo assim, quase todos os formatos mais utilizados no mercado são contemplados.	O software consegue importar imagens nosformatos: TIFF e MEL. Portanto, as imagens geradas pelo préprocessamento do software proposto podem ser utilizadas.

Tabela 1. Comparação entre os aspectos do software proposto e o software ImageMaster^{TM} 2D Platinum.

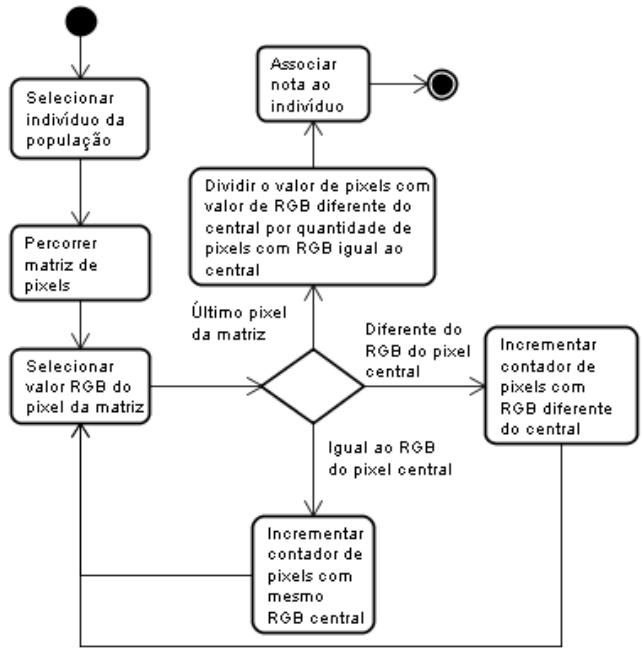


Figura 2. Fluxograma da função para atribuir nota aos indivíduos

gerando uma nota a qual é avaliada. Assim, garante que a nota atribuída seja relativa a variação de RGB da imagem e assim avaliada. Neste AG, a população inicial é de 50 indivíduos, que passa pelo processo de cruzamento 50 vezes e apenas o melhor indivíduo é selecionado para tratar a imagem. Com esse tratamento de exclusão dos ruídos, os *spots* podem ser melhor visualizados. Foi observado que o resultado não apresentou modificações relevantes com uma população superior a 50 indivíduos.

(iv) Tratamento de erro de detecção dos *spots*

O software disponibiliza uma funcionalidade para adicionar ou remover rastros e *spots* que não foram selecionados de maneira correta pelo algoritmo apresentado, o que proporciona ao pesquisador a liberdade de possíveis correções caso necessário.

3. Resultados e Discussão

Os testes do software proposto foram realizados utilizando uma amostra de 25 imagens digitais de géis de eletroforese bidimensional disponibilizadas pelo departamento de Bioquímica da Universidade Federal do Paraná. Seguindo os seguintes passos de utilização do software para processar a imagem:

- Importar;
- Processar;
- Ajuste manual dos *spots*;

- Salvar.

Foram avaliados os seguintes critérios:

- Facilidade de uso;
- Multiplataforma;
- Viabilidade de utilizar as imagens pré-processadas em outros softwares;
- Detecção automática dos *spots*;
- Alteração manual de detecção dos *spots*;

Após esse processo, as imagens foram submetidas ao software ImageMaster^{TM} 2D Platinum, com e sem o pré-processamento. A Tabela 1 mostra aspectos comparados entre o software ImageMaste^{TM} 2D Platinum e o software proposto em relação à detecção dos *spots*.

A multiplataforma possibilita o uso do software para outras pesquisas nos sistemas operacionais testados. As imagens geradas através do pré-processamento podem ser importadas por outros softwares porque a ferramenta computacional proposta é capaz de salvar imagens nos formatos mais utilizados no mercado. Incluindo o software ImagemMaster^{TM} 2D Platinum por esse importar imagens no formato TIFF geradas pelo software proposto.

Observou-se que sem o pré-processamento, o software ImageMaster^{TM} 2D Platinum possui uma taxa média de erro de 27,9%. Utilizando a ferramenta computacional proposta, essa taxa média de erro passa a ser de 9,3%. Isso ocorre porque a imagem pré processada é submetida sem os rastros mais significativos da imagem original, ou seja, quase todos os rastros são eliminados, ficando apenas os próximos aos *spots*. Com isso, ficou comprovada a eficiência da utilização de imagens pré-processadas para a detecção dos *spots*.

4. Conclusão

De acordo com os resultados, a ferramenta computacional apresenta uma alternativa eficiente para melhorar a qualidade das imagens digitais de eletroforese bidimensional. Por possuir interface intuitiva, ser multiplataforma e exportar as imagens pré-processadas em diversos formatos, pode ser estudado posteriormente.

Assim, conclui-se que a utilização do pré-processamento nas imagens digitais de eletroforese bidimensional torna a detecção dos *spots* mais eficiente e o trabalho dos pesquisadores mais eficaz, por não precisarem gastar horas para detectar manualmente cada *spot* reconhecido erroneamente.

5. Agradecimentos

Os autores agradecem aos órgãos financiadores CAPES, CNPq e UFPR pelo incentivo financeiro e estrutural para o sucesso do projeto. Ao Núcleo de Fixação de Nitrogênio e ao departamento de Bioquímica da Universidade Federal do Paraná por cederem as imagens utilizadas na pesquisa e informações biológicas. E aos colegas e professores do Programa de Pós-Graduação em Bioinformática da Universidade Federal do Paraná pelas valiosas discussões e informações.

6. Referências Bibliográficas

Referências

[1] N. Anderson, J. Taylor, A. Scandora, B. Coulter, and N. Anderson. *The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns.* Clin Chem., 1981.

[2] R. Appel, P. Palagi, D. Walther, J. Vargas, J. Sanchez, F. Ravier, C. Pasquali, and D. Hochstrasser. *MelanieII: A third generation software package for analysis of two-dimensional electrophoresis images*, volume 18. Electrophoresis, 1997.

[3] M. Berth, F. Moser, M. Kolbe, and J. Bernhardt. *The state of the art in the analysis of two-dimensional gel electrophoresis images*, volume 76. Appl. Microbiol. Biotechnol., 2007.

[4] Ciero, S. W., and Bellato. *The scientist and engineer's guide to digital signal processing.* California Technical Publishing, San Diego, CA, USA, 2002.

[5] D. Goldberg. *Genetic Algorithms In Search, Optimization And Machine Learning.* Addison-Wesley, Boston, MA, USA, 1989.

[6] P. Lemkin and L. Lipkin. *GELLAB: A computer system for 2D gel electrophoresis analysis. I. Segmentation of spots and system preliminaries.* Comput. Biomed., 1981.

[7] M. Moreira and M. Aguiar. *Um Sistema Classificador para Imagens de Sensoriamento Remoto baseado em Algoritmos Genï¿½ticos.* SBMAC, Santo Amaro, SP, 2005.

[8] P. Palagi, D. Walther, G. Bouchet, S. Voordijk, and R. Appel. *Image Analysis and Quantitation.* Elvier Science, USA, 2006.

Impact of Classifier Prototyping on Classification Systems Optimization

Paulo V. W. Radtke
Setor de Educação Profissional e Tecnológica
Universidade Federal do Paraná
Curitiba, Brazil
paulo.radtke@ufpr.br

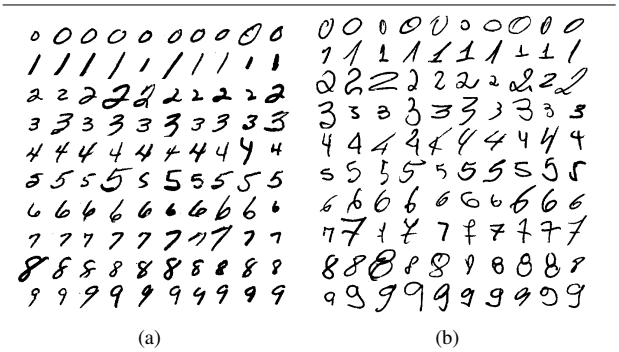
Robert Sabourin, Tony Wong
Laboratoire de vision, d'imagerie et de l'intelligence artificielle
École de technologie supérieure
Montréal, Canada
{robert.sabourin, tony.wong}@etsmtl.ca

Abstract

This paper compares two wrapper approaches to optimize classification systems. The traditional direct approach uses the actual classifier on the optimization process, whereas the second prototypes the target classifier with a computationally inexpensive classifier. Our results indicate that using the actual classifier yields computationally faster classifiers. However, the accuracy in both approaches is comparable and the required optimization time may be prohibitive in certain contexts, justifying the use of a prototyping approach.

1. Introduction

Image-based pattern recognition requires that pixel information be first transformed into an abstract representation (a feature vector) suitable for recognition with classifiers, a process known as *feature extraction*. A relevant classification problem is the *intelligent character recognition* (ICR), most specifically the off-line recognition of isolated handwritten symbols on documents. A methodology to extract features must select the spatial location to apply transformations on the image [6]. The choice takes into account the *domain context*, the type of symbols to classify, and the *domain knowledge*, what was previously done in similar problems. The process is usually performed by a human expert in a trial-and-error process. Associated to this burden, one classification system is tailored to a specific domain. Figure 1 details the difficulties faced by changing the handwriting style. Thus, the same classification system can not be used on another problem with the same reliability, unless the classification system is properly adapted to the new context.



(a) (b)

Figure 1. Handwritten digits: (a) NIST SD-19 and (b) Brazilian checks.

This context mandates a semi-automated approach that uses the expert's domain knowledge to optimize the classification system. To minimize the human intervention in defining and adapting classification systems, this problem is modeled as an optimization problem, using the expert's domain knowledge and information from the domain context – actual images. The approach optimizes classification systems as in Fig. 2. The *Intelligent Feature Extraction* (IFE) methodology extracts feature sets in RS_{IFE}, where the best performing classifier SI is selected. The IFE is further detailed in Section 2.

The optimization algorithm employed also plays an im-

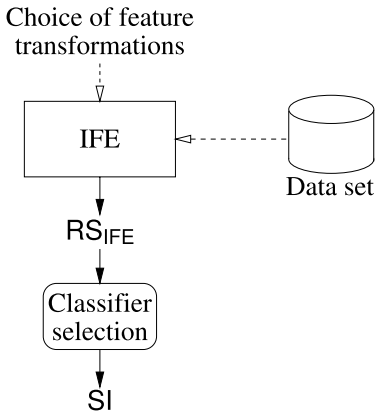


Figure 2. Classification system optimization approach.

portant role, not only related to obtained feature set quality, but to the required processing time as well. Experiments were performed in [9] using *multi-objective genetic algorithms* (MOGAs)[1] and in [10] with the *Record-to-Record Travel* (RRT) algorithm [8]. Both algorithms successfully optimized the problem, producing solutions that outperformed the human expert approach. We used a prototyping wrapper approach to optimize a *multi-layer perceptron* (MLP) neural network, owing to the long processing time required to train the MLP classifiers. The prototyping consisted in using the *projection distance* (PD) classifier [3] to obtain RS_{IFE} and select SI, which was then trained using an MLP classifier. Our hypothesis is that SI could be improved if we used the target classifier (MLP) in the optimization process, instead of prototyping feature sets with a different classifier.

The new contribution is to investigate this hypothesis of whether prototyping impacts heavily the solutions obtained, by using the RRT algorithm to optimize the IFE directly with MLP classifiers. The RRT algorithm is detailed in Section 3 and the experiments performed are indicated in Section 4. Results obtained are detailed in Section 5 and discussed in Section 6.

2. Intelligent Feature Extraction

The goal of IFE is to help the human expert define representations in the context of isolated handwritten symbols, using a wrapper approach to optimize solutions. IFE models handwritten symbols as features extracted from specific *foci* of attention on images using *zoning*. Two operators are used to generate representations with IFE: a *zoning operator* to define *foci* of attention over images, and a *feature extraction* operator to apply transformations in zones. The choice of transformations for the feature extraction operator constitutes the human expert domain knowledge. The domain context is introduced as actual observations in the *optimization* data set used to evaluate and compare solutions. Hence, the zoning operator is optimized by the IFE to the domain context and domain knowledge.

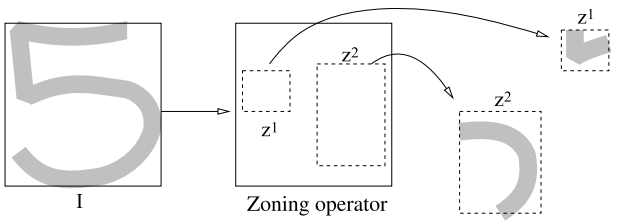


Figure 3. IFE structure.

The IFE structure is illustrated in Fig. 3. The zoning operator defines the zoning strategy $Z = \{z^1, \ldots, z^n\}$, where $z^i, 1 \leq i \leq n$ is a zone in the image I and n the total number of zones. Pixels inside the zones in Z are transformed by the feature extraction operator in the representation $F = \{f^1, \ldots, f^n\}$, where $f^i, 1 \leq i \leq n$ is the partial feature vector extracted from z^i. At the end of the optimization process, the optimization algorithm has explored the representation set $RS_{IFE} = \{F^1, \ldots, F^p\}$. For MOGAs, RS_{IFE} is the optimal set at the last generation, and for the RRT algorithm, it is the explored solution set. Whereas the later may seem larger at first, their sizes are actually comparable.

The result set RS_{IFE} is used to train the classifier set $K = \{K^1, \ldots, K^p\}$, where K^i is the classifier trained with representation F^i. The first hypothesis is to select the most accurate classifier $SI, SI \in K$ for a single classifier system. The second hypothesis, which is not subject of this work, is to use K to optimize an *ensemble of classifiers* (EoC) [5, 4] for higher accuracy. The remainder of this section discusses the IFE operators chosen for experimentation with isolated handwritten digits and the candidate solution evaluation.

2.1. Zoning Operator

To compare performance to the traditional human approach, a *baseline* representation with a high degree of accuracy on handwritten digits with a *multi-layer Perceptron* (MLP) classifier [7] is considered. This baseline representation was defined on a traditional trial and error basis. Its zoning strategy, detailed in Fig. 4.b, is defined as a set of three image dividers, producing 6 zones. The *divider zoning operator* expands the baseline zoning concept into a set of 5 horizontal and 5 vertical dividers that can be either *ac-*

tive or *inactive*, producing zoning strategies with 1 to 36 zones. Figure 4.a details the operator template, encoded by a 10-bit binary string. Each bit is associated with a divider's state (1 for active, 0 for inactive).

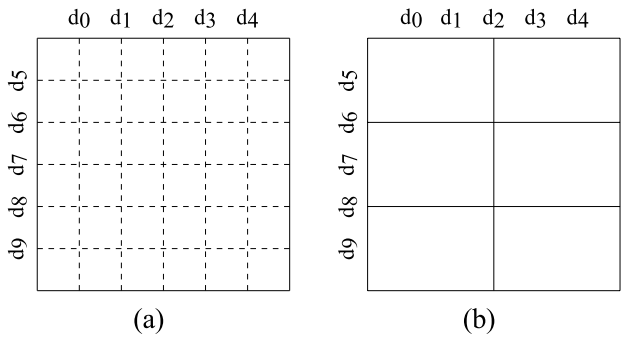


Figure 4. Divider zoning operator (a). The baseline representation in (b) is obtained by setting only d_2, d_6 and d_8 as active.

2.2. Feature Extraction Operator

The classification system in [7] used and detailed a mixture of concavities, contour directions and black pixel surface transformations, extracting 22 features per zone (13 for concavities, 8 for contour directions and 1 for surface). To allow a direct comparison between IFE and the baseline representation, the same feature transformations (the domain knowledge) are used to assess the IFE.

Black pixel surface indicates the percentage of image black pixels that are inside the zone. The contour direction transformation is performed with a histogram of contour directions. For each zone in the image, contour line segments are labeled clockwise and counted regarding the 8-Freeman directions in Fig. 5.b, producing the 8 features vector in Fig. 5.c. The example in Fig. 5 details the contour count for the highlighted zone in the image. The first line is the direction count in this zone, whereas the second line is the associated Freeman direction.

Heutte *et al.* discussed in [2] the concept of concavities transformation. For every white pixel in the image in Fig. 6.a, we search in 4-Freeman directions (Fig. 6.b) for black pixels and determine the directions we reach black pixels and the directions we do not. If we find black pixels in all directions, we search in four auxiliary directions (Fig. 6.c) to confirm that the white pixel is inside a closed contour or not. Pixels that reach only one black pixel are ignored as they are not inside a concavity. For each zone the transformation produces the 13 positions feature vector in Fig.

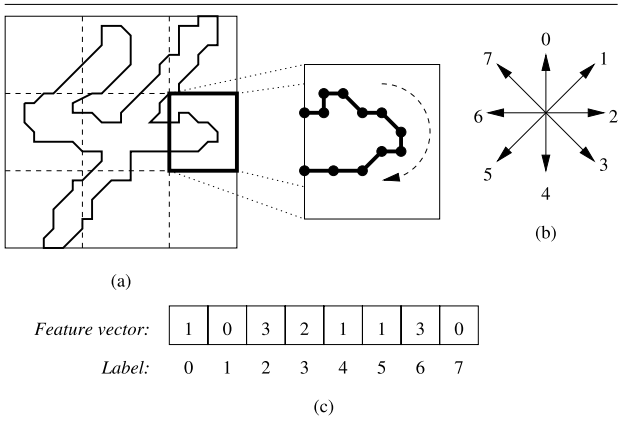


Figure 5. Contour feature extraction process.

6.d, where the first line labels the number of directions that black pixels can be reached, the second line labels the directions we find white pixels (for pixels on an open contour) and the last indicates the pixel count associated to that feature. Two examples in Fig. 6.a illustrates the process, pixel x_1 reaches four black pixels but fails to find black pixels in direction S_1, and pixel x_2 reaches three black pixels and misses direction 1.

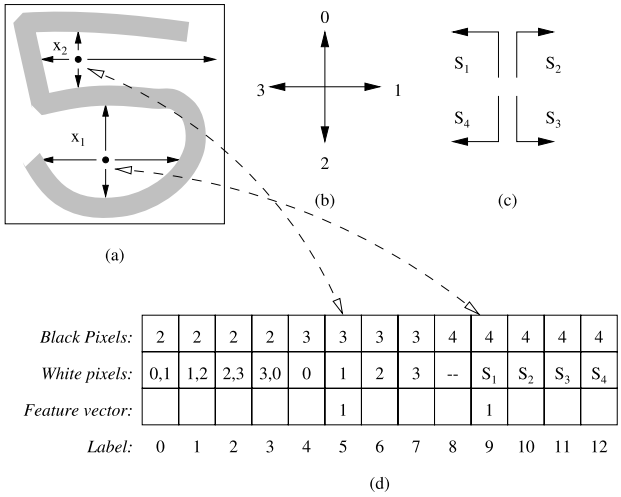


Figure 6. Concavities feature extraction process.

2.3. Candidate Solution Evaluation

Candidate solutions are evaluated with respect to their classification accuracy. Thus, the objective is to minimize the classification error rate on the *optimization* data set (the domain context). To compare wrapper methods, candidate solutions are optimized with both the PD and MLP classifiers. The PD classifier is used as an inexpensive prototyping classifier, and obtained feature sets are used to train the SI MLP classifier. The optimization process is also performed directly with the MLP classifier. The goal is to compare classifier performance when using prototyped feature sets against feature sets obtained directly with the target classifier.

3. Optimization Algorithm

The *Record-to-Record Travel* (RRT) algorithm [8], is an annealing based heuristic. It is said to be a local search algorithm, searching for new solutions in the vicinity of the current solution. The RRT algorithm improves an initial solution i by searching in its neighborhood for better solutions based on their evaluation (classification error rate). The RRT algorithm, detailed in Algorithm 1, produces after a number of iterations the record solution r. The algorithm is similar to a hill climbing approach, but avoids local optimum solutions by allowing the search towards non-optimal solutions with a fixed deviation D. Earlier experiments indicated that the RRT algorithm over-fitted solutions during the optimization process. The global validation strategy discussed in [11] is used to avoid this effect, and Algorithm 1 includes support for this strategy.

Given the initial solution i, the algorithm will copy it to the record solution r and store its evaluation value in $RECORD$. It also copies i as the current solution p. Next it will repeat the following process during a number of iterations, until the current solution is worse than the record solution plus the allowed deviation. First it will find the set P, solutions neighbor to p, and select the best neighbor $p', p' \in P$. To avoid cyclic optimization, solutions already evaluated are not considered for p'. If evaluating p' yields results within the allowed deviation, it is copied as p for the next iteration. Solution p' replaces the record solution r only if it yields better results. If p' is worse than the allowed deviation, the optimization process stops. The explored solution set S is responsible to store solutions tested by the RRT algorithm for the global validation strategy. At each iteration, the algorithm inserts into S the solutions in the neighbor set P. At the end of the optimization process, solutions in S are validated and the most accurate solution is selected. For the IFE process, S is the result set RS_{IFE} used to create the classifier set K.

Data: Initial solution i
Data: Deviation D
Result: Record solution r
Result: Explored solution set S
$r = i$;
$RECORD = eval(r)$;
$p = i$;
$S = \emptyset$;
repeat
  Create the solution set P, neighbor to p;
  $S = S \cup P$;
  Select the best solution $p' \in P$ such as that p' has not yet been evaluated;
  if $eval(p') < RECORD + RECORD \times D$ **then**
    $p = p'$;
    if $eval(p') < RECORD$ **then**
      $RECORD = eval(p')$;
      $r = p'$;
    end
  end
until $eval(p) <= RECORD + RECORD \times D$;

Algorithm 1: Modified record to record travel (RRT) algorithm used to optimize classification systems with global validation.

Neighbors to solution X^i are created by swapping bits in the binary string with their complement. For a binary string E with p bits, a set of p neighbors is created by complementing each bit $i, 1 \leq i \leq p$ on solution E^i. For the IFE, solution in Fig. 7.a has solutions in Figs. 7.b and 7.c as two possible neighbors.

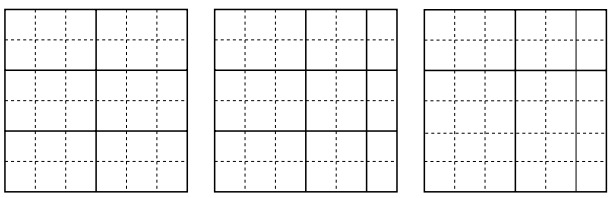


Figure 7. Zoning strategy (a) and two neighbors (b and c) using the proposed divider zoning operator.

4. Experimental Protocol

The tests are performed as in Fig. 2. The IFE methodology is solved to obtain the representation set RS_{IFE}, which is used to train the classifier set K and the most accurate

classifier $SI, SI \in K$ is selected. When using the prototyping approach, the RRT algorithm produces RS_{IFE} using the PD classifier, and only when training the selected SI that the MLP classifier is used. The direct wrapper approach uses the MLP during the optimization process. To select the most accurate solution in the result set S, we use the global validation approach detailed in [11]. Solutions obtained are compared to the baseline representation defined in [7] by an human expert. The RRT algorithm is deterministic, thus no replications are required to evaluate mean values.

The data sets in Table 1 are used in the experiments – isolated handwritten digits from NIST SD-19. MLP classifier training is performed with the *training* data set, whereas PD classifier training is performed with the $training'$ data set. The *validation* data set is used to adjust the classifier parameters (PD hyper planes, MLP weights). The optimization process is performed with the *optimization* data set, and the *selection* data set is used with the global validation strategy to select solutions. The $test_a$ and $test_b$ databases are used to compare solutions, where the later is known to require a robust classifier for higher accuracies. The number of MLP hidden layers is selected as the most performing in val in one of the five factors of feature set cardinality in $hn = \{0.4, 0.45, 0.50, 0.55, 0.6\}$.

Data set	Size	Origin	Sample range
$training'$	50000	hsf_0123	1 to 50000
training	150000	hsf_0123	1 to 150000
validation	15000	hsf_0123	150001 to 165000
optimization	15000	hsf_0123	165001 to 180000
selection	15000	hsf_0123	180001 to 195000
$test_a$	60089	hsf_7	1 to 60089
$test_b$	58646	hsf_4	1 to 58646

Table 1. Handwritten digits data sets extracted from NIST SD-19.

Both experiments have the IFE initial solution associated to an empty string (all bits set to 0). Thus, there are no active dividers in the initial IFE solution. The deviation D is set empirically to $D = 5\%$. The RRT is a deterministic algorithm, hence a single run is performed with both processes. The baseline representation is compared directly with solutions SI_p, obtained with the prototyping approach (PD classifier), and SI, obtained directly with the MLP classifier, regarding both accuracy, feature set cardinality and IFE optimization required processing time.

5. Results

Experimental results are detailed in Table 2, where Z is the solution zone number, $|S|$ is the solution cardinality (feature number), hn is the MLP hidden layer size, e_{test_a} and e_{test_b} are classification error rates on $test_a$ and $test_b$. Solution SI_p is obtained through classifier prototyping with PD during the optimization process, whereas SI is obtained using the MLP during the optimization process. The baseline representation trained with an MLP classifier is included for comparison purposes.

Solution	Z	$\|S\|$	hn	e_{test_a}	e_{test_b}
Baseline	6	132	60	0.91%	2.89 %
SI_p	15	330	132	0.82%	2.51 %
SI	10	220	132	0.80%	2.48 %

Table 2. Digits optimization results.

Regardless the wrapper approach used during the optimization process, the IFE outperformed the baseline representation defined by the human expert. When comparing SI and SI_p accuracies, we find that both solutions are comparable. Whereas SI is numerically better than SI_p, it represents only a 2.43% error rate decrease on $test_a$ (0.02% difference), and an 1.2% error rate decrease on $test_b$ (0.03% difference). Thus, we observe no significant accuracy improvement on SI when compared to SI_p. On the other hand, SI has a much smaller feature set size $|SI| = 220$, if compared to $|SI_p| = 330$, a 33% feature set size decrease. Classification system computational time is directly related to the feature set size it uses, thus, using SI is preferable, instead of SI_p.

The obtained zoning strategies are shown in Fig. 8, along with the baseline solution. We observe that the baseline solution, defined by the human expert, shares the same basic structure as SI and SI_p. All divider bits that are set on the baseline solution are also set on SI and SI_p. We also observe that most divider bits are similar on SI and SI_p, and that differences observed demonstrate that different classifiers require feature sets better adapted for higher performance.

As different classifiers were used on both tests, the number of different solutions evaluated is also different. To optimize SI_p using the PD classifier, a total of 76 different candidate solutions was tested. To optimize SI using the MLP classifier, the RRT algorithm evaluated 53 different candidate solutions. Finally, considering the required processing time, the prototyping approach is faster. On a Core2DUO 3GHz processor with 4GB memory, optimizing SI directly with the MLP classifier took 4 days using both processor cores (one MLP training process per core), whereas opti-

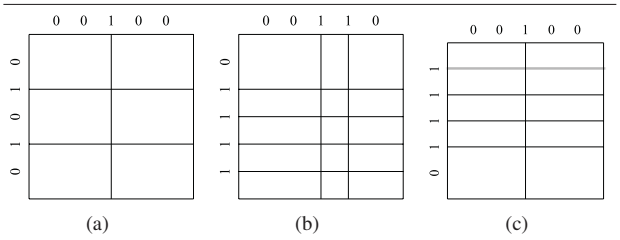


Figure 8. Baseline zoning strategy (a), SI_p (b) and SI (c).

mizing SI_p took 10 hours using only one processor core, which can be further optimized if more processor cores are used. For comparison purposes, using a MOGA to optimize SI_p with the IFE required the evaluation of 450 different candidate solutions. Thus, optimizing IFE using a MOGA and the direct wrapper approach will require further optimizations so it can be processed in reasonable time.

6. Discussion

We tested two different wrapper approaches to optimize classification systems. One direct approach, which uses the target classifier, and a prototyping approach, which uses a different classifier to prototype feature sets to reduce processing time. Whereas both approaches produced solutions that have comparable accuracy, we observed that their cardinality was different, favoring the direct approach to use the target classifier. We also observed that the optimization process took about 10 times longer when using the direct approach, in comparison to the prototyping approach. As the classification system computational cost is related to the feature set used, we can say that it is preferable to optimize classification systems with the direct approach when the time required to optimize the classification system is not an issue, or the training data sets are smaller. Otherwise, the prototyping approach will provide a good trade-off between classification accuracy and required optimization process time.

Current processors are the bottleneck when using the direct approach (MLP classifier). To avoid long optimization times, we should investigate GPU based parallelism to accelerate the MLP and PD training procedures, or to use a cluster to increase the computational power for MLPs. Future works will also investigate the use of RS_{IFE} obtained with the direct approach to optimize an EoC in order to obtain a classification system with higher accuracy.

Using the IFE with a MOGA as the optimization algorithm might also be possible after accelerating the MLP training stage. This test would be interesting to confirm the RRT as the computational efficient choice for the IFE, as in [10] we observed that the IFE produced comparable results with both a MOGA and the RRT algorithm.

References

[1] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, LTD, Baffins Lane, Chichester, West Sussex, PO19 1UD ,England, 2001.

[2] L. Heutte, T. Paquet, J. Moreau, Y. Lecourtier, and C. Olivier. A structural/statistical feature based vector for handwritten character recognition. *Pattern Recognition Letters*, 19(7):629–641, 1998.

[3] F. Kimura, S. Inoue, T. Wakabayashi, S. Tsuruoka, and Y. Miyake. Handwritten Numeral Recognition using Autoassociative Neural Networks. In *Proceedings of the International Conference on Pattern Recognition*, pages 152–155, 1998.

[4] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.

[5] L. I. Kuncheva and L. C. Jain. Design classifier fusion systems by genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 4(4):327–336, 2000.

[6] Z.-C. Li and C. Y. Suen. The partition-combination method for recognition of handwritten characters. *Pattern Recognition Letters*, 21(8):701–720, 2000.

[7] L. S. Oliveira, R. Sabourin, F. Bortolozzi, and C. Y. Suen. Automatic Recognition of Handwritten Numerical Strings: A Recognition and Verification Strategy. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(11):1438–1454, 2002.

[8] J. W. Pepper, B. L. Golden, and E. A. Wasil. Solving the traveling salesman problem with annealing-based heuristics: A computational study. *IEEE Trans. on Systems, Mand and Cybernetics – Part A: Systems and Humans*, 32(1):72–77, 2002.

[9] P. V. W. Radtke, T. Wong, and R. Sabourin. Classification system optimization with multi-objective genetic algorithms. In *Proceedings of the 10th International Workshop on Frontiers in Handwriten Recognition (IWFHR 2006)*, pages 331–336. IAPR, 2006.

[10] P. V. W. Radtke, T. Wong, and R. Sabourin. Using the rrt algorithm to optimize classification systems for handwritten digits and letters. In *Proceedings of The 23rd Annual ACM Symposium on Applied Computing - SAC 2008*, pages 1748–1752. ACM, 2008.

[11] P. V. W. Radtke, T. Wong, and R. Sabourin. Solution over-fit control in evolutionary multiobjective optimization of pattern classification systems. *IJPRAI*, 23(6):1107–1127, 2009.

Modelagem Contextual De Padrões Aplicada Ao Monitoramento De Eventos Em Vídeo

Fabio Borges dos Santos
Pontifícia Universidade Católica do Paraná
1155, Imaculada Conceição
Curitiba, Brasil
fabiozesge@gmail.com

Edson José Rodrigues Justino
Pontifícia Universidade Católica do Paraná
1155, Imaculada Conceição
Curitiba, Brasil
justino@ppgia.pucpr.br

Alceu de Souza Britto Jr
Pontifícia Universidade Católica do Paraná
1155, Imaculada Conceição
Curitiba, Brasil
alceu@ppgia.pucpr.br

Resumo

This paper proposes a method for the detection of conventional and non-conventional events in video sequences. The proposed method is to analyze video streams, detect the moving, tracking and extract features to further classification of motion. For detection of moving is used the temporal difference. For tracking objects are extracted characteristics make as speed, position, etc. With these features, a vector is created to provides the link between the detected objects in consecutive frames. For classification of events are also the features extracted from the center of mass of the object to the extremities and distance traveled by the object. These features are presented to a SVM classifier trained to classify the movement as conventional or non-conventional. The experimental results using three databases (Parking Lot, CAVIAR and BEHAVE) show that the recognition rate was between 69.72% and 75.07%.

1. Introdução

A crescente necessidade de utilização de sistemas de monitoramento, incentiva o crescimento de um novo campo de pesquisa em computação, o de análise de eventos em vídeo. Este campo de pesquisa dá subsídios à criação de ferramentas eficazes no combate ao crime, na detecção de acidentes domésticos, de infrações de trânsito e outros eventos indesejáveis que podem ser monitorados por um sistema inteligente com capacidade de identificar os eventos não condizentes com o ambiente [17, 16, 6].

A definição de quais eventos são convencionais ou não-convencionais, depende do ambiente onde eles ocorrem, ou seja, o contexto. São muitas as possibilidades de movimentos existentes, identificados, estes movimentos são agrupados em duas meta-classes de movimento. Uma meta-classe é a união de duas ou mais classes com o objetivo de reduzir a complexidade do processo de reconhecimento[18].

O objetivo deste artigo é propor um método para a detecção de eventos convencionais e não-convencionais de vídeo através da modelagem contextual de padrões. Desta forma nossos desafios são encontrar uma combinação de técnicas de detecção de movimento que seja simples mas eficaz, fazer o correto rastreamento dos objetos detectados e encontrar dentre as características extraídas dos objetos o conjunto com maior capacidade de discriminação dos eventos utilizando um classificador SVM (*Suport Vector Machine*).

A primeira etapa, consiste na detecção dos movimentos na cena. As técnicas mais utilizadas para a detecção do movimento são: subtração de fundo, diferença temporal e fluxo óptico [9]. Estas técnicas podem ser utilizadas separadamente ou combinadas[2]. É possível combinar as técnicas e utilizar Gaussianas[22] para detectar as regiões de movimento e a subtração de fundo local nas regiões onde foi detectado o movimento para refinar os resultados [7].

O rastreamento dos objetos em quadros consecutivos utiliza características como pontos, linhas ou formas geométricas [9]. As técnicas de rastreamento são divididas em quatro grupos: baseado no contorno; baseado em regiões; baseado em características; e baseado em modelos[13, 14].

A análise do comportamento e classificação baseia-se em

duas técnicas: comparação de modelos ou espaço de estados [1, 20]. Para modelar o movimento realizado pelos objetos na cena são extraídas características como o centróide do objeto, alterações na velocidade, direção, etc. Estas características irão compor os protótipos de trajetórias de movimento. Estes protótipos são armazenados. Quando apresentado um objeto a ser classificado é utilizado o algoritmo *Altruistic Vector Quantatization* (AVQ) para medir a distância entre as trajetórias [17]. Em outra abordagem, com características de velocidade e posição, são traçadas curvas que representam os modelos de atividades. Estas curvas são comparadas com os modelos obtidos na fase de treinamento, por um classificador SVM [19].

Este trabalho apresenta um método de detecção e classificação de eventos de vídeo que visa auxiliar os sitemas de monitoramento já existentes. A principal característica deste é a construção contextual de um modelo de movimentos. Das sequências de vídeo são extraídas características que modelam os diferentes tipos de movimento. Estes tipos de movimentos são agrupados de acordo com o ambiente analisado em meta-classes de movimentos: Convencionais e Não-convencionais. Para a classificação final é utilizado o SVM. As etapas de detecção de movimentos, rastreamento dos objetos, extração de características e classificação dos eventos visam a utilização de técnicas de baixo custo computacional, para que o método possa ser utilizado em tempo real.

2. Visão Geral do Método

O método aqui proposto tem como entrada uma cena de vídeo a ser analisada, desta cena são extraídos os quadros que compõem a sequência como um todo. Cada quadro é analisado e comparado com os quadros anteriores, para que seja possível detectar as regiões de movimento. Estas regiões representam os objetos em movimento presentes na cena, estes objetos são analisados com base em suas dimensões e velocidade, para determinar se são de interesse ou não. Desta forma apenas objetos de nosso interesse são rastreados, em nosso estudo focamos nosso objetivo em analisar o movimento de pessoas.

Durante o tempo de permanência do objeto de interesse em cena, dele são extraídas características referentes ao próprio objeto e ao movimento por ele realizado, que compõem o vetor de características que representa o movimento realizado pelo objeto, este vetor é apresentado a um classificador SVM devidamente treinado, para que o movimento realizado seja classificado como convencional ou não-convencional de acordo com as meta-classes de movimentos modelados com base no contexto apresentado. A Figura 1 apresenta a visão geral deste método bem como as etapas seguidas pra chegar ao resultado final.

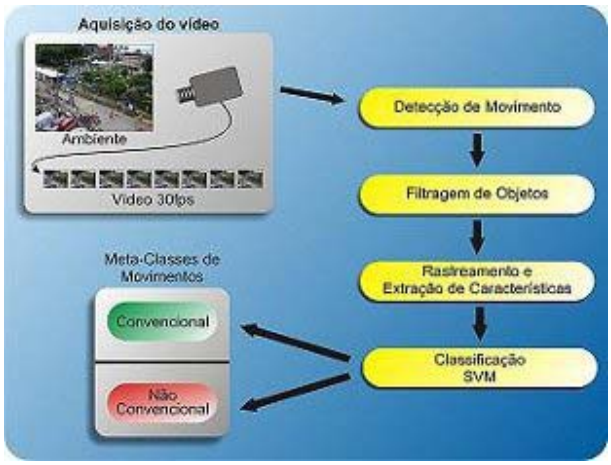


Figura 1. Etapas do modelo contextual de padrões aplicado ao monitoramento de eventos em vídeo.

Utilizou-se o classificador SVM para a classificação dos eventos em vídeos. Onde este classificador é treinado com base no ambiente analisado. Nos mais diversos ambientes encontrou-se diversos tipos de movimentos como pessoas caminhando, caminhando em zig-zag, correndo, praticando exercícios ou em casos extremos brigando. De acordo com os ambientes estes tipos de movimentos são agrupados, onde o critério de agrupamento é a normalidade do movimento no ambiente. São formados dois grupos chamados meta-classes. Para elaboração deste artigo, consideramos como um tipo de movimento convencional apenas pessoas caminhando. O restante dos movimentos, como pessoas correndo, praticando exercícios, brigando ou andando em zig-zag foram considerados como eventos não-convencionais. O classificador SVM utilizado é treinado previamente com exemplos de movimentos convencionais e não convencionais devidamente rotulados. Um objeto em movimento provê muitos vetores de características que detalham o movimento realizado, em um primeiro momento cada vetor extraído foi analisado de maneira isolada. Para suavizar possíveis divergências na classificação dos vetores de um mesmo objeto foi utilizada uma janela temporal, onde os vetores que estão dentro da janela são classificados e é feita uma votação para chegar ao resultado final.

2.1. Deteção de Movimento

A detecção do movimento é a primeira etapa deste processo de análise e classificação de eventos, é nela, que os objetos em movimento na cena são detectados. Desta etapa em diante a análise fica restrita a estes objetos. O primeiro

passo é a aplicação da técnica de diferença temporal, onde é utilizado um limiar de distância entre os quadros. Este limiar determina com qual quadro anterior o quadro atual será comparado [13]. Depois de aplicada a diferença temporal, com o objetivo de refinar os resultados, são aplicadas operações morfológicas de dilatação e erosão, desta maneira pequenos ruídos oriundos de variações de iluminação ou mesmo dos algoritmos de compressão de vídeo são eliminados e os pixels pertencentes a um mesmo objeto são agrupados [4]. As técnicas selecionadas até aqui são simples, porém apresentam bons resultados se comparadas a técnicas que utilizam as Gaussianas [22]. Após detectados os objetos é aplicado um filtro para identificar os objetos de interesse, este filtro utiliza características de velocidade e dimensões do objeto em movimento. E tem por objetivo descartar objetos que não se enquadrem nos valores parametrizados [13].

2.2. Rastreamento de Objetos

O rastreamento dos objetos consiste em acompanhá-lo durante a sua permanência em cena. Para isso é utilizada a fórmula de custo [11, 7]. Temos um objeto O^i em um quadro F^n onde O^i representa um objeto rastreado. No quadro $F^{(n+1)}$ são identificadas as regiões de movimento, é necessário saber qual região representa o objeto O^i no quadro atual. É utilizada a seguinte função de custo:

$$Custo = (wP * dP) + (wS * dS) + (wD * dD) + dT \quad (1)$$

Onde wP, wS e wD são pesos atribuídos a cada parâmetro e somados não devem ultrapassar 1 (um). dP é distância euclidiana entre os centros do objeto O^i e a região de movimento analisada. dS é a diferença de tamanho entre o objeto O^i e a região de movimento. dD é a diferença de direção da região de movimento e a posição estimada pelo algoritmo Lucas-Kanade no quadro atual e da diferença entre os centros da região de movimento e do objeto. E dT é a diferença da permanência máxima de um objeto em cena e o tempo de permanência do objeto rastreado. A distância entre os centros dos objetos é medida pela seguinte equação:

$$dP = |R_c^j - O_c^i| \quad (2)$$

onde R_c^j é a região de movimento que está sendo analisada e O_c^i é a última posição conhecida do objeto.

$$dS = |R_r^j - O_r^i| / (R_r^j - O_r^i) \quad (3)$$

onde R_r^j e O_r^i representam o tamanho dos objetos, dos quais estamos extraindo a diferença de tamanho.

$$dD = |\arctan(O_s^i - O_c^i) - \arctan(R_c^j - O_c^i)| \quad (4)$$

é a diferença do ângulo formado pela reta traçada da posição do objeto no quadro anterior até posição estimada pelo algoritmo Lucas-Kanade [15] e a reta traçada também da posição do objeto no quadro anterior até a posição da região de movimento analisada, estes ângulos devem ser entre 0 e 2π.

$$dT = |TTL_{max} - O_{TTL}^i| \quad (5)$$

onde TTL_max é o tempo máximo de permanência de um objeto em cena, em quadros. E O_{TTL}^i é o tempo de permanência do objeto analisado. Se o objeto é encontrado no quadro atual, este valor é ajustado em TTL_{max}, do contrário o valor é decrementado até que chegue a 0 e o objeto seja eliminado do rastreamento. A sugestão do valor do TTL_{max} é de 3 vezes o número de quadros por segundo, tal valor foi obtido através de experimentos realizados.

Quando detectada a região de movimento que apresenta o menor custo os valores referentes ao objeto passam a ser os valores da região de movimento, ou seja, a região de movimento passa a ser considerada como objeto O^i.

3. Extração de Características

Os objetos detectados na cena são rastreados e analisados com base nas características extraídas. Tais informações são utilizadas para classificá-los como convencionais ou não-convencionais.

As características extraídas formam um vetor, composto por sete características. Este vetor é representado por V_i.

$$V_i = [Vel, Pos, Desl, Tam, Dir, DistEx, Dist] \quad (6)$$

As sete características são representadas por:

$$Vel = \sqrt{(O_{c_{t-1}}^i - O_{c_t}^i)^2} / Q \quad (7)$$

Sendo que Vel representa a velocidade do objeto e é obtida pela distância percorrida pelo centro do objeto em quadros dividida pelo tempo Q. Com a característica de velocidade acreditamos ser possível diferenciar objetos com movimentos rápidos ou lentos em cena [19]. A posição do objeto em cena é representada por Pos e compreende as coordenadas x e y do centro do objeto [17].

$$Desl = O_{c_{t-1}}^i - O_{c_t}^i \quad (8)$$

Onde $Desl$ representa o deslocamento do objeto, seu valor é obtido pela diferença da posição atual do objeto e sua posição no quadro anterior, considerando os eixos xey. Esta característica permite distinguir qual o sentido do movimento realizado pelo objeto [7].

$$Tam = |O_{r_{t-1}}^i - O_{r_t}^i| \quad (9)$$

Tam é a variação do tamanho do objeto em quadros consecutivos. Em determinados casos esta característica pode indicar, por exemplo, se o objeto se movimenta de maneira uniforme, em casos de cenas onde o objeto analisado é uma

pessoa correndo ocorre mais variações. Para o caso de uma pessoa caminhando esta variação é menor [7].

$$Dir = |\arctan(O^i_{c_{t-2}} - O^i_{c_{t-1}}) - \arctan(O^i_{c_{t-1}} - O^i_{c_t})| \quad (10)$$

A direção do movimento é representada por Dir, apresenta a variação na direção do movimento realizado pelo objeto. Esta característica apresenta as variações na trajetória do objeto em quadros consecutivos, esperamos distinguir entre movimentos retilíneos e movimentos sinuosos (zig-zag) [17].

$$Dist = |(O^i_{c_{t-1}} - O^i_{c_t})| \quad (11)$$

$Dist$ representa a distância percorrida pelo objeto em cena. Esta característica está diretamente ligada a característica que se refere a velocidade do objeto [12]. E $DistEx$ que representa as distâncias do centro de massa do objeto até as extremidades superior esquerda e inferior direita. Esta característica foi escolhida para indicar objetos onde as pessoas realizem movimentos dos membros, por exemplo, acenar [25].

$$DistEx = |O^i_c - O^i_e| \quad (12)$$

A extração destas características é feita em um intervalo de tempo Q pré-configurado, onde em nossos estudos $Q = 3$.

4. Seleção de características

Um fator relevante na construção de um classificador é a seleção das características mais discriminantes. É comum encontrar situações onde a inclusão de novas características ao modelo não resulta em ganhos no desempenho do classificador. A escolha das características é de suma importância. O melhor conjunto de características deixa as classes bem definidas e torna a etapa de classificação trivial. Na prática o processo de seleção de características consiste em uma função de otimização multi-critérios, onde busca-se minimizar o número de características e maximizar o desempenho do classificador [5, 3, 23].

Afim de encontrar o melhor conjunto de características, ou seja, o conjunto com maior capacidade de discriminação entre as classes, utilizou-se o método exaustivo sobre o conjunto de dados de treinamento, este método foi utilizado por termos um baixo número de características.

5. Classificação de Eventos

Os vetores de características extraídos são armazenados em duas bases de dados, uma para treinamento e outra para teste. Estes vetores são rotulados de acordo com o tipo de movimento que realizam em convencionais ou não-convencionais. A primeira base é utilizada para o treinamento do classificador. A segunda base de dados é utilizada para avaliar a desempenho do classificador diante de exemplos desconhecidos.

Como estratégia de aprendizagem de máquina foi utilizado o classificador SVM. O SVM é um classificador que tem chamado a atenção nas áreas de aprendizagem de máquina e reconhecimento de padrões, por possuir a capacidade de generalização [24]. Como parâmetros para C e γ utilizados pelo classificador SVM foram utilizados os valores padrão sugeridos pela LIBSVM. Foi utilizado o método exaustivo de seleção de características para fazer a busca pelo melhor conjunto de características. No processo de treinamento do classificador a base de dados de treinamento é dividida em cinco grupos de igual tamanho, quatro destes grupos foram utilizados no treinamento do classificador e o quinto grupo foi utilizado para testar o classificador treinado, este processo é denominado validação cruzada [8].

6. Resultados Experimentais

Em nossos experimentos utilizamos diversos exemplos de movimentos. Para comprovar que a utilização das meta-classes gera ganhos no desempenho do classificador, foram realizados experimentos com onde os tipos de movimentos foram classificados separadamente. Em seguida estes exemplos de movimentos foram agrupados de acordo com o contexto analisado, em dois grupos, as duas meta-classes de movimentos, convencionais e não-convencionais. Dentre os movimentos realizados temos pessoas correndo, caminhando, fazendo exercícios, andando em zig-zag e brigando. Considerou-se como eventos convencionais movimentos de pessoas caminhando em linha reta, os demais tipos de movimentos foram considerados como não-convencionais.

Os vídeos utilizados nos experimentos foram selecionados de maneira aleatória de três bases de dados diferentes: CAVIAR; BEHAVE e *Parking Lot*. A base de dados *Parking Lot* é composta por vídeos de uma câmera de segurança instalada em um dos prédios da universidade. Os vídeos apresentam um ambiente externo e desprovido de qualquer controle de iluminação. Os exemplos de movimentos das três bases foram avaliados em conjunto.

Para cada objeto de nosso interesse, foram extraídos diversos vetores de características. Na primeira etapa dos experimentos cada vetor foi analisado separadamente, sem a utilização de janela temporal e votação. Foram extraídos 4934 vetores de eventos considerados como convencionais, pessoas andando. E 2837 vetores de eventos considerados não-convencionais, pessoas correndo, fazendo exercícios e brigando. Estes vetores foram divididos em 2971 convencionais e 1909 não-convencionais para a base de dados de treinamento e 1962 convencionais e 928 não-convencionais para a base de dados de teste. Na primeira etapa onde foi analisado cada tipo de movimento como uma classe a parte, foi treinado o classificador SVM multi-classe. Na segunda

Movimentos	Convencional	Não-conviencional
Convencional	1341(68,35%)	621(31,65%)
Não-Convencional	278(29,96%)	650(70,04%)

Tabela 1. Matriz de confusão entre as meta-classes, com a utilização de todas as características e duas meta-classes de movimentos.

Movimentos	Convencional	Não-conviencional
Convencional	1731(88,23%)	231(11,77%)
Não-Convencional	530(57,11%)	398(42,89%)

Tabela 2. Matriz de confusão entre as meta-classes com a utilização das características de $Desl_y$ e $DistExt$

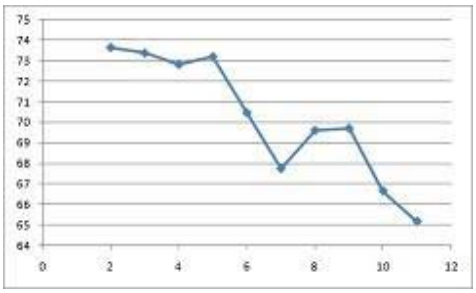


Figura 2. Taxas de acerto para os intervalos de características.

etapa as classes de movimentos forma agrupadas nas duas meta-classes de movimentos. E na ultima etapa dos experimentos foi aplicada a janela temporal para avaliar a variação das taxas de classificação.

Um dos objetivos deste trabalho é apresentar a performance do classificador SVM na classificação de eventos em cenas de vídeo, bem como encontrar o melhor conjunto de características para este tipo de problema. O classificador SVM apresentou taxas de reconhecimento total de 67,78% quando analisadas as classes de movimento separadamente. Ao agrupá-las nas duas meta-classes de movimentos convencionais e não-convencionais, esta taxa de reconhecimento passou para 68,87%, considerando a utilização de todas as características do vetor e sem utilização da janela temporal. Na tabela 01 é apresentada a matriz de confusão gerada pela classificação dos eventos quando utilizadas as duas meta-classes. Pode-se observar que taxa de reconhecimento para os eventos convencionais ficou em 68,35% e para os eventos não-conviencionais ficou em 70,04%.

Utilizando a janela temporal a taxa de reconhecimento ficou em 69,72%, o tamanho da janela foi 9, onde a cada 9 vetores extraídos para detalhar o movimento do objeto é feita a classificação e a votação para definir o resultado final da classificação.

Aplicada a seleção de características o melhor resultado sem a janela temporal foi obtido com as características de $Desl$, considerando o eixo y e $DistExt$, analisando cada evento separadamente, a taxa de reconhecimento foi em 72,21 %. Ao utilizar as duas meta-classes a taxa passou a 73,64%. Aplicada a janela temporal a taxa passou a ser de 75,07%. A tabela 02 apresenta a matriz de confusão gerada pela utilização do melhor conjunto de características encontrado. É possível observar que para os eventos convencionais a taxa de reconhecimento passou para 88,23%, porém para os eventos não-convencionais a taxa passou a ser muito inferior, 42,89%. Desta maneira a eficiência do método para a detecção de eventos não-convencionais fica comprometida.

Trabalhos semelhantes com a utilização de um classificador baseado em instâncias (K-NN), relatam que para um dado ambiente, as caracaterísticas de velocidade e variação de direção do movimento apresentaram uma taxa de 77,20% de deteção de eventos não-convencionais quando utilizada uma janela temporal [7].

Na Figura 2 são apresentados os melhores resultados obtidos para os diferentes conjuntos de características sem a utilização da janela temporal. Pode-se observar que com apenas duas características o resultado obtido foi de 73,64% considerando as duas meta-classes de movimentos.

7. Conclusões

Com o estudo e desenvolvimento do método de modelagem contextual de padrões aplicado ao monitoramento de eventos em vídeo, conclui-se que, o método exaustivo de seleção de características para o classificador SVM, com a utilização dos valores padrão sugeridos pela LIBSVM não garante que o conjunto de características encontrado é definitivamente o melhor, pois com a variação dos parâmetros de C e γ, ou mesmo com a utilização de outras funções de kernel, as taxas de reconhecimento podem variar.

A utilização de meta-classes de movimento mostra-se como uma alternativa para a redução da complexidade de classificação. Considerando que o classificador torna-se mais simples e existe um ligeiro aumento na taxa de reconhecimento, bem como a utilização de uma janela temporal, onde mais de um vetor de características do mesmo objeto seja analisado para a tomada de decisão, elimina pequenos erros na classificação, isto pode ser observado pelo aumento nas taxas de reconhecimento quado apli-

cada a janela temporal.

Como ponto negativo, o método apresenta a busca exaustiva pelo melhor conjunto de características, pois sempre que incluída uma característica no vetor, o número de tentativas para encontrar o melhor conjunto de características aumenta exponencialmente. Uma alternativa para agilizar o processo de seleção de características seria a não utilização da abordagem *wrapper*, pelo fato dela possuir alto custo computacional, ao invés disto utilizar o método *filter*, já que esta abordagem apresenta independência com relação ao classificador a ser utilizado.

A selecção de características quando consideradas a taxa de reconhecimento geral mostra que o conjunto de características pode ser considerado bom para detectar um dos tipos de eventos. Nos resultados apresentados, é possível observar que o melhor conjunto de características encontrado possui 88,23% de capacidade de reconhecimento de eventos convencionais, em contra partida apresenta baixa performance no reconhecimento dos eventos não-convencionais (42,89%). Desta maneira conclui-se também que é necessário evidenciar o que se espera do método como um todo, maior precisão na deteçao dos eventos convencionais ou não-convencionais, pois o melhor conjunto de características para detecção dos eventos convencionais não necessariamente é o melhor para a deteção dos eventos convencionais.

Contudo os resultados mostraram-se satisfatórios na utilização do classificador SVM desprovido de quais quer parâmetros de configuração além do padrão sugerido pela LIBSVM.

Referências

[1] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. *Computer Vision and Image Understanding: CVIU*, 73(3):428–440, 1999.

[2] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson. A system for video surveillance and monitoring. In *Tech. Report CMU-RI-TR-00-12*. Robotics Institute, Carnegie Mellon University, 2002.

[3] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons, Chichester, 2001.

[4] Y. Dedeoglu. Moving object detection, tracking and classification for smart video surveillance, 2004.

[5] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, 2001.

[6] K. Hara, T. Omori, and R. Ueno. Detection of unusual human behavior in intelligent house. *IEEE Workshop on Neural Networks for Signal Processing*, pages 697–706, 2002.

[7] A. G. Hochuli, A. de Souza Britto Jr., and A. L. Koerich. Detection of non-conventional events on video scenes. In *SMC*, pages 302–307. IEEE, 2007.

[8] C.-W. Hsu, C.-C. Chang, , and C.-J. Lin. A practical guide to support vector classification, 2003.

[9] W. Hu, T. Tan, L. Wang, and S. J. Maybank. A survey on visual surveillance of object motion and behaviors. *SMC-C*, 34(3):334–352, Aug. 2004.

[10] D. Koller, J. Weber, and J. Malik. Towards realtime visual based tracking in cluttered traffic scenes. 1994.

[11] L. J. Latecki and R. Miezianko. Object tracking with dynamic template update and occlusion detection. In *18th Intl Conf on Pattern Recognition*, pages 556–560, Washington, USA, 2006.

[12] K. K. Lee, M. Yu, and Y. Xu. Modeling of human walking trajectories for surveillance. In *Proceedings of the 2003 IEEE/RSJ, International Conference on Intelligent Robots and Systems*, 2003.

[13] B. Lei and L.-Q. Xu. From pixels to objects and trajectories: A generic real-time outdoor video surveillance system. *IEEE International Symposium on Imaging for Crime Detection and Prevention*, pages 117–122, 2005.

[14] W. Lok and K. L. Chan. Model-based human motion analysis in monocular video. *ICASSP2005*, 2:697–700, 2005.

[15] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *7th Intl Joint Conf Artificial Intelligence*, pages 674–679, Vancouver, Canada, 1981.

[16] S. Lühr, H. H. Bui, S. Venkatesh, and G. A. W. West. Recognition of human activity through hierarchical stochastic learning. In *PerCom*, pages 416–. IEEE Computer Society, 2003.

[17] A. Mecocci, M. Pannozzo, and A. Fumarola. Automatic detection of anomalous behavioral events for advanced real-time video surveillance. In *CIMSA'03*, pages 187–192. IEEE Computer Society, 2003.

[18] M. Morita, R. Sabourin, F. Bortolozzi, and C. Suen. Segmentation and recognition of handwritten dates. In *Proc. 8 Frontiers of Handwriting Recognition (IWFHR)*, pages 105–110, Aug. 2002.

[19] W. Niu, J. Long, D. Han, and Y.-F. Wang. Human activity detection and recognition for video surveillance. In *ICME*, pages 719–722. IEEE, 2004.

[20] R. Polana and R. Nelson. Low level recognition of human motion (or how to get your man without finding his body parts, 1994.

[21] C. Ridder, O. Munkelt, and H. Kirchner. Adaptive background estimation and foreground detection using kalman filtering. *ICAM*, pages 193–199, 1995.

[22] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):747–757, 2000.

[23] G. Trunk. *A problem of dimensionality: a simple example*. IEEE Trans. Pattern Anal. Mach. Intell. v1 i3. 306-307.

[24] V. Vapnik. *The nature of statistical learning theory*. Springer, New York, 1995.

[25] X. Wu, Y. Ou, H. Qian, and Y. Xu. A detection system for human abnormal behavior. In *IROS 2005: IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1204 – 1208. IEEE Computer Society, 2005.

Recuperação de imagens por conteúdo utilizando Lógica *Fuzzy* - um estudo de caso sobre imagens faciais

Leila Cristina Carneiro Bergamasco; Fátima L.S. Nunes
Universidade de São Paulo – Escola de Artes, Ciências e Humanidades
leila.cristina@usp.br; fatima.nunes@usp.br

Resumo

A técnica de CBIR (Content-Based Image Retrieval) consiste em recuperar imagens de um banco de dados com base na similaridade entre uma imagem modelo e as imagens armazenadas. A lógica Fuzzy tem por objetivo classificar essa similaridade considerando graus de pertinência, sendo que quanto maior o grau, maior a similaridade. O presente trabalho propõe a implementação e a análise da recuperação de imagens por conteúdo utilizando a técnica de Inteligência Artificial de lógica Fuzzy. A integração entre CBIR e lógica Fuzzy ocorrerá por meio da técnica de realimentação por relevância, na qual o usuário classifica cada imagem resultante como relevante ou não a sua busca, de modo que o algoritmo 'aprenda' e adeque seus graus de pertinência de acordo com as especificidades do usuário.

Abstract

The CBIR (Content-Based Image Retrieval) technique consists in retrieve images from a database based on the similarity among stored images in a database and another image used as a model. Fuzzy Logic, aims at classify this similarity on relevance degrees, as higher the level, greater the similarity. This paper proposes the implementation and analysis of image retrieval by content using the technique of Fuzzy Logic. The integration of Fuzzy Logic and CBIR occurs through by relevance feedback technique, which the user rates each resulting image as relevant or not, in such way the algorithm 'learns' and compute its relevance degree according to the specific user.

1. Introdução

Soluções automatizadas estão cada vez mais presentes no cotidiano das pessoas com o objetivo de facilitar e garantir resultados com grande qualidade.

O reconhecimento facial é uma dessas soluções e está sendo empregado com maior frequência na sociedade, principalmente nas áreas de segurança e investigação. Por esse motivo, a otimização desse processo é estudado e vem aumentando a sua significância na medida em que se mostra eficaz nas rotinas aplicadas [1].

Para o processamento e busca das imagens é possível utilizar CBIR (*Content Based Image Retrieval*), que por meio de extratores (algoritmos que extraem características das imagens) e funções de similaridade, comparam as imagens de uma base de dados a uma imagem modelo, retornando as mais semelhantes ao modelo fornecido.

Para aumentar a qualidade das respostas do sistema, é possível utilizar a lógica *Fuzzy*, uma técnica amplamente divulgada em Inteligência Artificial, que tem por objetivo diminuir o *gap* semântico entre o usuário e a lógica matemática do computador [2].

A proposta desse artigo é discutir os resultados e eficácia da aplicação da lógica *Fuzzy* em um sistema CBIR no contexto da rotina de reconhecimento facial.

Na seção 2 são apresentados os trabalhos que serviram como base para o presente estudo, nas seções 3 e 4 são definidos os conceitos de CBIR, realimentação por relevância e lógica *Fuzzy* e a metodologia empregada para a conclusão do trabalho, respectivamente, e nas seções 4 e 5 os resultados obtidos são mostrados e discutidos.

2. Trabalhos correlatos

Para esse estudo de caso foram analisadas técnicas principalmente de realimentação por relevância e CBIR para entender qual o estado de conhecimento de ambos os assuntos. Em [3], Torres e Falcão fazem uma revisão das principais técnicas de CBIR de cor, textura e forma, além de elucidar sobre algumas funções de similaridade utilizadas atualmente e sob quais aspectos esses quesitos são funcionais ou não.

No trabalho de Marques [4] foi empregada a técnica de realimentação por relevância por meio de duas técnicas: *Relevance Feedback Projection* que separa as características em faixas de relevância e gera um novo valor calculado a partir de projeções sobre a faixa de objetos relevantes, e a técnica *Multiple Point Projection* que combina a técnica anterior ao uso de múltiplos centros de consulta.

Em [5] foi utilizada a técnica de Programação Genética, conceito extraído da Inteligência Artificial para executar a realimentação por relevância nos vetores de características extraídos da imagem.

O principal objetivo desse trabalho é propor uma nova abordagem para a realimentação por relevância utilizando a lógica *Fuzzy* a fim de tornar a interação entre usuário e o sistema CBIR mais intuitivo.

3. Conceitos

A técnica de CBIR, conhecida como reconhecimento de imagens baseada em conteúdo, tem por objetivo, a partir de uma imagem modelo, apresentar ao usuário as imagens mais relevantes que constam na base de dados.

A busca pelas imagens mais relevantes ocorre por meio das funções de similaridade que são aplicadas nas imagens, gerando os vetores de características de cada uma delas. Esse vetor é composto de dados matemáticos extraídos das imagens usando-se extratores. A partir do resultado proposto, o usuário pode, por meio do sistema de CBIR, definir quais as imagens são mais relevantes no resultado da busca e quais imagens não foram relevantes, na sua concepção [3].

Os extratores são algoritmos que extraem características matemáticas que, quando transformadas em números, representam o conteúdo da imagem. Os extratores podem ser globais, recuperando características da imagem inteira, ou locais, agindo em regiões específicas da imagem. As características extraídas formam um vetor de características que define matematicamente a imagem [6]. Os extratores podem representar características diversas, sendo bastante aplicados para representar aspectos de cor, forma e textura.

As funções de similaridade buscam demonstrar matematicamente o quanto uma imagem é similar a outra. Esse processo envolve a escolha de uma ou mais funções de distâncias métricas e sua aplicação nos vetores de características das imagens obtidos por meio dos extratores.

A técnica de CBIR permite qualificar as respostas dadas como relevantes ou não, fazendo com que o algoritmo 'aprenda' quais propriedades visuais são melhores para determinado contexto.

Esse conceito chamado por realimentação por relevância tem dois objetivos principais: reduzir o *gap* semântico entre a análise visual de alto nível feita pelo usuário e a análise de baixo nível feita por meio dos extratores e funções de similaridade; e reduzir a subjetividade da percepção humana, uma vez que diferentes usuários podem ter diferentes percepções de relevância para a mesma imagem [4].

A lógica *Fuzzy* é uma proposta que procura atender o aspecto vago deixado pela teoria clássica dos conjuntos, em relação a informações com dados imprecisos [6].

Enquanto na teoria clássica, aqui chamada por conjuntos *'crisp'*, um elemento pertence ou não a um determinado conjunto, ou seja, tem valores 0 ou 1 de pertinência, nos conjuntos *Fuzzy* um determinado elemento pode pertencer com grau em um intervalo de 0 a 1.

Um sistema *Fuzzy*, é composto por três processos base*: fuzzyficação*, inferência e *defuzzyficação.*

Tomando por base, um exemplo de um sistema de controle de temperatura, demonstrado graficamente na Figura 1, onde o eixo x representa a temperatura em graus Celsius e o eixo y, a pertinência de cada grau a cada uma das variáveis linguísticas {muito baixa, baixa, agradável, alta}, será explicada a função de cada etapa de um sistema *Fuzzy*.

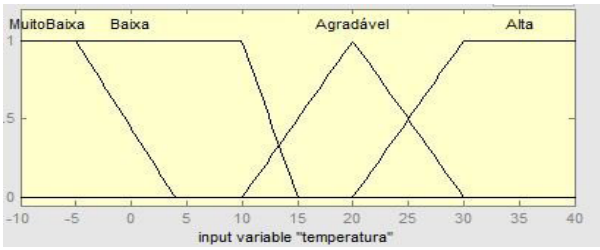


Figura 1: Sistema *Fuzzy* de controle de temperatura

O processo de *fuzzyficação* consiste em transformar os valores de entrada, que são valores *crisp*, em valores *Fuzzy* [7]. No exemplo anterior, dada uma temperatura de 25 graus, é verificado o grau de pertinência em cada variável lingüística e dessa forma não é mais analisada a informação pelo seu valor matemático, mas sim como parte integrante de um dos conjuntos de variáveis. Assim 25 graus pode ser uma "temperatura agradável" e uma "temperatura alta" com grau de pertinência menor.

Na inferência *Fuzzy*, são aplicadas as regras definidas do tipo "Se <premissa> então <conclusão>". Esse processo é o mais importante de um sistema *Fuzzy* devido ao fato que nele a lógica *Fuzzy* é amplamente utilizada por meio dos operadores de implicação. Para cada dado de entrada é feita uma análise em cada regra definida e as conclusões obtidas individualmente são agrupadas para obtenção de um resultado global.

No processo de *defuzzyficação* é implementado um algoritmo para extração de um valor *crisp* a partir de um conjunto de dados nebulosos. Alguns exemplos são: maior dos máximos, menor dos mínimos, média dos máximos e centróide.

4. Metodologia

Para atender aos objetivos do presente trabalho e baseado na revisão bibliográfica feita, foi visto que o desenvolvimento de um sistema de reconhecimento facial seria a forma mais adequada para se avaliar a eficácia do processo, dessa forma, três módulos foram desenvolvidos:

módulo CBIR, módulo Fuzzy e o módulo de Realimentação por Relevância.

A primeira etapa do módulo CBIR consistiu em escolher uma base de dados para que fosse possível extrair mais facilmente características e testar a eficiência do sistema. Foi escolhido um cenário controlado, com fotografias de rostos de homens e mulheres de diferentes raças na fase adulta. A base de dados é da Universidade de Essex [7].

No sistema foram implementados 17 extratores, contemplando as categorias de cor, forma e textura. Na Tabela 1 é apresentado a finalidade de cada um deles. Para enriquecer a pesquisa os extratores de cor foram calculados por meio de histogramas, que dão uma visão global da imagem, e também utilizando os *pixels* com os seus valores individuais.

Tabela 1**:** Extratores implementados

Tipo	**Extrator**	**Finalidade**
Cor	Média do Histograma em Nivel de Cinza	Obter informação sobre a tonalidade da imagem (mais clara ou mais escura)
	Média do Histograma em RGB	Obter informação relativa ao tom da imagem (mais clara ou mais escura) em cada canal de cor.
	Desvio Padrão do Histograma em Nivel de Cinza	Obter informação do índice de variação de cores em uma imagem
	Desvio Padrão do Histograma em RGB	Obter informação do índice de variação de cores em uma imagem, em cada canal de cor.
	Obliquidade do Histograma em Nível de Cinza	Obter o valor da variação das cores
	Obliquidade do Histograma em RGB	Obter o valor da variação das cores em cada canal de cor.
	Média dos Pixels em Nivel de Cinza	Obter informação sobre a tonalidade da imagem (mais clara ou mais escura), analisando os valores dos *pixels* individualmente.
	Média dos Pixels em RGB	Obter informação sobre a tonalidade da imagem (mais clara ou mais escura) em cada canal de cor, analisando os valores dos *pixels* individualmente.
	Desvio Padrão dos Pixels em Nivel de Cinza	Obter informação do índice de variação de cores em uma imagem, analisando os *pixels* individualmente.
	Desvio Padrão dos Pixels em RGB	Obter informação do índice de variação de cores em uma imagem, em cada canal de cor, analisando os *pixels* individualmente.
	Obliquidade dos Pixels em Nível de Cinza	Obter o valor da variação das cores, analisando os *pixels* individualmente.
	Obliquidade dos Pixels em RGB	Obter o valor da variação das cores, analisando os *pixels* individualmente, em cada canal de cor e analisando os *pixels* individualmente.
Textura	Contraste	Determinar a variação dos *pixels,* que indica o seu contraste
	Entropia	Determinar o índice de dispersão dos *pixels.*

Forma	Escala	Por meio do índice de escala, comparar duas imagens e ver a diferença entre os tamanho dos rostos
	Rotação	Encontrar os eixos de inércia da imagem para comparação.
	Translação	Encontrar os prolongamentos e orientações da forma do rosto.

A função de similaridade implementada foi a distância Euclidiana e para as imagens de testes foram escolhidos três imagens faciais, denominadas R1, R2 e R3 respectivamente, conforme mostrado na Figura 2.

Figura 2: Imagens faciais utilizadas para o teste[7].

No módulo *Fuzzy* foram definidas as variáveis para o processo de *fuzzyficação, a*s regras para a inferência *Fuzzy* e escolhido o método centróide para a extração do valor *crisp* do sistema, esse valor corresponde ao peso a ser ajustado de cada imagem.

Cada imagem possui um peso que é ajustado a cada iteração de treinamento, de acordo com a resposta do sistema Fuzzy. O valor do peso encontrado é, então, multiplicado pela distância Euclidiana da respectiva imagem, o que implica que pesos maiores causarão um aumento substancial na distância da imagem. Esta ação levará consequentemente, em uma próxima iteração, a não retornar a imagem como sendo uma das imagens candidatas. Dessa forma, o algoritmo diminui a margem de erros causada pelo uso dos extratores puramente matemáticos e faz com que a busca por imagens candidatas fique mais refinada.

Para implementar o módulo de Realimentação por Relevância foi construída uma interface que permite ao usuário escolher a imagem modelo e, em seguida as cinco mais parecidas com ela são apresentadas para avaliação.

A realimentação por relevância acontece quando, a partir da nota atribuída pelo usuário, o algoritmo identifica quais imagens pode descartar e quais deve usar como referência em uma próxima busca. A interação com o usuário termina somente quando ele julga que todas as imagens candidatas atendem a sua solicitação. Na Figura 3 é apresentado o fluxograma do funcionamento do sistema de reconhecimento facial.

Figura 3: Fluxograma do funcionamento do sistema

Para cada um dos rostos escolhidos para o teste há outras quatro imagens similares, apenas alterando a expressão facial. Elas estão em uma base de dados com outras 40 imagens de outras pessoas. O objetivo do teste foi verificar se através da realimentação por lógica *Fuzzy*, o sistema é capaz de selecionar as quatro imagens corretamente para cada rosto buscado.

5. Resultados

Na Figura 4 é possível ver que na primeira interação, nem todas as imagens retornadas são as similares a imagem-objetivo, denominada R1.

Figura 4: Primeira interação da imagem R1

Tendo em vista que com a função Euclidiana é possível ter apenas a visão global da imagem, é de se concluir que

alguns extratores podem ter prejudicado no cálculo da função de similaridade, levando a um valor errado. Isso pode ser constatado na Tabela 2, nos extratores em negrito, que mostram a distância Euclidiana entre o rosto representado na imagem R1 e as cinco primeiras imagens que foram retornadas na primeira interação.

Tabela 2: Distância Euclidiana entre R1 e imagens candidatas

	Imagens - Candidatas				
Extratores	**1**	**2**	**3**	**4**	**5**
Média do Histograma em Nivel de Cinza	0,00	0,071	0,006	0,039	0,036
Média do Histograma em RGB	0,00	0,578	0,003	6,417	3,533
Desvio Padrão do Histograma em Nivel de Cinza	0,00	2,372	2,593	0,864	2,372
Desvio Padrão do Histograma em RGB	0,00	0,130	0,002	0,030	0,044
Obliquidade do Histograma em Nível de Cinza	**0,00**	**38,583**	**12,505**	**78,023**	**118,81**
Obliquidade do Histograma em RGB	**0,00**	**0,000**	**0,000**	**142,88**	**37,992**
Média dos Pixels em Nivel de Cinza	0,00	0,071	0,006	0,039	0,036
Média dos Pixels em RGB	0,00	28,930	45,328	40,039	178,834
Desvio Padrão dos Pixels em Nivel de Cinza	0,00	17,961	35,033	40,997	17,961
Desvio Padrão dos Pixels em RGB	0,00	0,130	0,002	0,030	0,044
Obliquidade dos Pixels em Nível de Cinza	**0,00**	**36,871**	**188,85**	**54,031**	**58,956**
Obliquidade dos Pixels em RGB	**0,00**	**498,90**	**467,52**	**140,08**	**150,48**
Contraste	0,00	0,007	0,011	0,016	0,007
Entropia	0,00	0,005	0,005	0,008	0,005
Escala	0,00	0,013	0,052	0,060	0,062
Rotação	0,00	0,016	0,063	0,074	0,077
Translação	0,00	0,006	0,104	0,083	0,044

Aplicando a realimentação por relevância por meio de notas variando de 0 a 10 e o processamento de ajustes de peso por meio da lógica *Fuzzy*, o usuário consegue minimizar o erro causado pelos extratores, refinando a sua busca.

Isso pode ser constatado por meio do gráfico mostrado pela Figura 5, onde para a imagem R1, houve a necessidade de sete interações para a busca das outras cinco imagens que eram similares à ela. Para a imagem

R2 foram feitas 10 interações e para a imagem R4, seis avaliações.

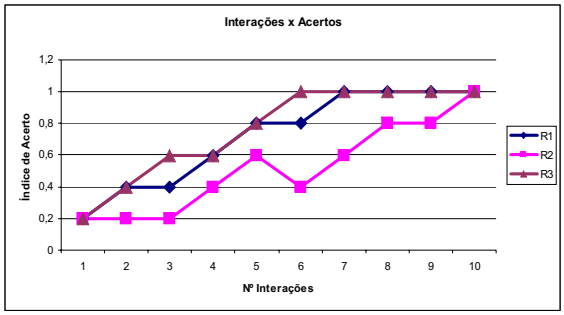
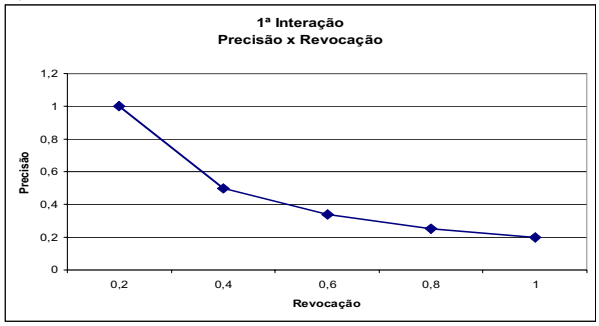


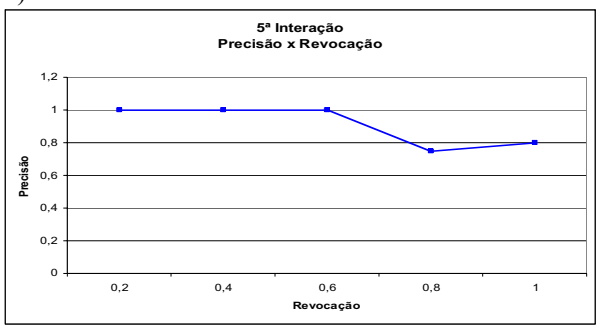
Figura 5: Interações x Acertos

Conforme o número de interações com o usuário foi aumentando, a busca se tornou mais refinada e começou a retornar resultados melhores, fazendo com que a área dos gráficos de precisão e revocação também crescessem, indicando sucesso na busca. A evolução dos resultados da imagem R1 estão apresentados na Figura 6, que mostra três momentos do sistema, o momento A indica a primeira interação com o usuário, o momento B a quinta interação, e o momento C a sétima e última interação.

A)



B)



C)

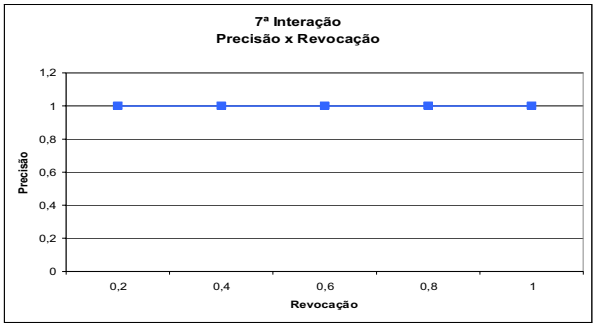


Figura 6: Gráfico de precisão e revocação para a imagem R1, representando diferentes momentos do sistema. Gráfico A) 1ª interação; B) 5ª interação; C) 7ª interação.

Na Figura 7 é possível analisar através da interface do sistema, a evolução que houve no decorrer das interações utilizando um sistema de lógica *Fuzzy*.

Figura 7: Resultado da sexta interação com R1.

6. Conclusões

Faz-se necessário discutir alguns pontos observados durante a implementação do sistema, o primeiro deles é sobre os extratores e funções de similaridade. Como visto anteriormente, houve extratores que prejudicaram a distância Euclidiana com valores que se diferenciavam em grande escala dos demais. No sistema foram utilizados extratores de todas as categorias (cor, textura e forma), porém há alguns estudos que apontam para a relação da eficiência de determinados extratores no contexto em que são inseridos. Nesse sentido, um aspecto a ser aprofundado é a escolha do melhor conjunto de extratores de forma a diminuir a quantidade de iterações necessárias ao processo.

Outro ponto em destaque foi a função de similaridade usada. A distância Euclidiana retorna um valor global, reflexo do conjunto de todos os extratores, isso também ajudou a trazer resultados menos coerentes ao usuário. Uma vez que o extrator estava retornando resultados ruins, o valor da distância Euclidiana também piorava. Uma ideia que pode ser explorada é como a lógica Fuzzy poderia realimentar os valores de cada um dos extratores e não da distância global obtida.

O uso da técnica de realimentação por relevância em conjunto com a lógica *Fuzzy* produziu resultados bastantes satisfatórios, e mostrou que a utilização de métodos que aproximam o usuário do processo de busca, diminuindo o *gap* semântico que ainda existe entre a percepção do usuário e a lógica matemática da computação é extremamente importante e que é possível que ambas trabalhem juntas e produzam resultados cada vez mais significativos.

7. Agradecimentos

Ao Projeto INCT – MACC (Institutos Nacionais de Ciência e Tecnologia – Medicina Assistida por Computação Científica), pelo apoio financeiro.

8. Referências Bibliográficas

[1] MÁXIMO, A. *A importância do mapeamento da criminalidade utilizando-se tecnologia de sistema de informação geográfica para auxiliar a segurança pública no combate à violência*. 2004. 101f. Dissertação (Mestrado em Engenharia de Produção e Sistemas) – PPGEP, Universidade Federal de Santa Catarina, Florianópolis. cap. 1, p. 11-13.

[2] PEDRYCZ, W.; GOMIDE F. *An introduction of Fuzzy Sets: Analysis and Design.* Mit Press, 1998.

[3] TORRES, R. da S.; FALCÃO, A. X. *Recuperação de Imagens Baseadas em Conteúdo*. In: Workshop de Visão Computacional, 4. 2008, Bauru. São Paulo: Universidade Estadual de São Paulo - UNESP, 2008.

[4] MARQUES, J. *Realimentação de relevância para recuperação por conteúdo de imagens médicas visando diminuir a descontinuidade semântica*. In: Simpósio de Teses e Dissertações, 11, 2008. São Paulo. São Paulo: Instituto de Ciências Matemáticas e Computação, Universidade de São Paulo, 2008.

[5] SILVA, A. T. *Recuperação de Imagens por Realimentação de Relevância*. In: EADCA - Encontro dos Alunos e Docentes do Departamento de Engenharia de Computação e Automação Industrial, 2, 2009, Campinas. Anais.. São Paulo: Universidade de Campinas – UNICAMP, 2009.

[6] SANDRI, S.; CORREA, C. *Lógica Nebulo*sa. In: Escola de Redes Neurais, 5. 1999, São Paulo. São Paulo: Instituto de Tecnologia e Aeronáutica – ITA, 1999.

[7] ESSEX UNIVERSITY. *Collection of Facial Images*. Inglaterra, 2008. Disponível em: http://www.face-rec.org. Acessada em: Março, 2011.

Traçado e Análise Cefalométrica: Uma Solução Computacional

Lucas Lima Batista[1], Roberto Lago[1], Claudio Eduardo Goes[2], Mauricio Cunha Escarpinati[3], Michele Fúlvia Angelo[1]

[1]*Departameto de Tecnologia – Universidade Estadual de Feira de Santana (UEFS), Av. Transnordestina, S/N, Novo Horizonte, CEP: 4436-900, Feira de Santana, BA, BR*

[2]*Departameto de Ciências Exatas – Universidade Estadual de Feira de Santana (UEFS), Av. Transnordestina, S/N, Novo Horizonte, CEP: 4436-900, Feira de Santana, BA, BR*

[3]*Faculdade de Computação - Universidade Federal de Uberlândia, Uberlândia, MG, BR*
Email: {lucaslbatista, betoecomp, cegoes}@gmail.com, mauricio@facom.ufu.br, mfangelo@ecomp.uefs.br

Abstract.

This paper presents the development of routines for the marking of cephalometric landmarks and the anatomical drawing on cephalometric images, incorporated in the computational tool in development. This computational tool has been designed to assist the expert in the dental radiology area showing the results of the cephalometric analysis faster and with fewer errors caused by the completion of these analyzes manually.

1. Introdução

A cefalometria radiográfica corresponde às mensurações da imagem radiográfica da cabeça (ossos, dentes e tecido mole). As aplicações da cefalometria na Ortodontia são diversas, entre elas: estudo do crescimento e desenvolvimento craniofacial do paciente; diagnóstico radiográfico de possíveis patologias instaladas; detecção de alterações nas várias regiões do crânio, permitindo a avaliação do local exato da anormalidade morfológica, estrutura dentária, esquelética e/ou tegumentar; entre outros [1].

A análise e o diagnóstico, a partir da cefalometria, são realizados através dos traçados cefalométricos ou análises cefalométricas. As análises cefalométricas são compostas por um conjunto de pontos cefalométricos que, por sua vez, unidos a outros, específicos das estruturas ósseas, dentárias e tegumentar por meio do desenho de linhas permitem a realização das mensurações (ângulos e distâncias) ditadas pelas diferentes análises cefalométricas [1].

O traçado cefalométrico, mesmo com os constantes avanços tecnológicos, em muitos locais ainda é feito de forma totalmente manual pelo ortodontista. Isto se deve principalmente ao alto custo para aquisição e uso dos softwares que existem no mercado. No Centro Integrado de Odontologia Prof. José Sobreira Filho da Universidade Estadual de Feira de Santana – Cion/UEFS, serviços gratuitos de atendimento odontopediátrico são prestados à comunidade, e os especialistas utilizam métodos manuais tanto para marcação dos pontos cefalométricos, como para a realização dos traçados, o que demanda um considerável tempo de trabalho dos especialistas ao realizar tais tarefas, além da possibilidade de erros de medida por instrumentos manuais (régua, compasso, lápis) [2].

Diante dos problemas apresentados, uma ferramenta computacional está sendo desenvolvida com a finalidade de auxiliar especialistas da área de radiologia odontológica ao apresentar o resultado de análises cefalométricas de forma mais rápida e com menos erros causados pela realização destas análises manualmente. Assim, o objetivo deste trabalho é apresentar o desenvolvimento de rotinas para a marcação dos pontos cefalométricos e a realização do desenho anatômico em imagens cefalométricas, incorporadas à ferramenta em desenvolvimento.

2. Metodologia

Para o desenvolvimento das rotinas para a marcação e edição manual dos pontos cefalométricos foi necessário conhecer como esse processo é feito de forma totalmente manual, com a utilização de papel vegetal sobre a imagem radiográfica, régua e lápis. Para isso, um profissional da área de ortodontia mostrou as etapas para realização do traçado cefalométrico, incluindo uma aula

prática, o que proporcionou uma base suficiente para que as rotinas pudessem ser projetadas e desenvolvidas.

Após conhecer o processo feito manualmente, foi decidido que além da função de marcação dos pontos, seria necessária outra função para a inserção e edição das estruturas anatômicas na imagem. Essas estruturas têm o papel de auxiliar o especialista a encontrar a exata localização dos pontos cefalométricos na radiografia.

Primeiramente foi projetada e implementada a interface gráfica para abrir e visualizar a imagem radiográfica, fornecendo ao usuário uma visão ampla da imagem em que ele irá trabalhar. A seguir são apresentadas, separadamente, as rotinas implementadas para a marcação e edição do desenho anatômico e as rotinas referentes a marcação dos pontos cefalométricos.

2.1 Marcação e Edição do Desenho Anatômico

A marcação e edição do desenho anatômico foi dividida em duas fases. A primeira foi a obtenção de um desenho anatômico base, desenhado por um especialista. O desenho anatômico foi capturado através de um software onde o especialista desenhou estrutura por estrutura (Figura 1) e este desenho é utilizado como um desenho anatômico de referência, ou seja, sempre que uma nova imagem é aberta, este desenho de referência é carregado e cabe ao especialista, para cada caso, ajustá-lo à radiografia do paciente.

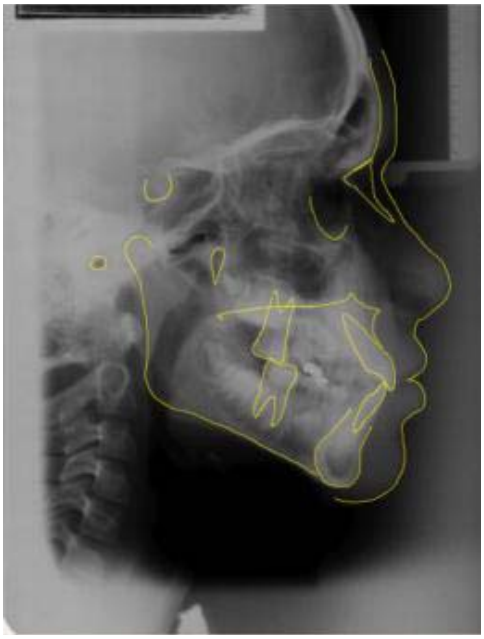


Figura 1: Desenho anatômico.

A segunda fase foi o desenvolvimento de rotinas para o ajuste do desenho anatômico, para as quais foram implementadas três transformações geométricas: rotação, translação e escala.

Transladar significa movimentar o objeto, ou seja, adicionar quantidades às suas coordenadas no plano (x, y). Assim, cada ponto pode ser movido por T_x unidades em relação ao eixo e T_y unidades em relação ao eixo y [3]. Portanto, pode-se representar a nova posição de um ponto transladado por:

$$[x' \; y'] = [x \; y] + [T_x \; T_y]$$

Escalonar um objeto é mudar suas dimensões. Para fazer com que um objeto mude de escala deve-se multiplicar os valores de suas coordenadas por um fator de escala [3]. Supondo que S_x e S_y sejam os fatores de escala pode-se representar essa operação por:

$$[x' \; y'] = [x \; y] \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix}$$

Rotacionar um objeto é girá-lo em torno de algum referencial. Se um ponto de coordenada (x, y) que está a uma distância $r = \sqrt{x^2 + y^2}$ da origem do sistema de coordenadas, for rotacionado de um ângulo θ em torno da origem, suas novas coordenadas podem ser definidas como:

$$x' = r.\cos(\theta + \phi)$$
$$y' = r.\mathrm{sen}(\theta + \phi)$$

$$x' = \mathrm{r}.\cos(\phi).\cos(\theta) - \mathrm{r}.\mathrm{sen}(\phi).\mathrm{sen}(\theta)$$
$$y' = \mathrm{r}.\mathrm{sen}(\phi).\cos(\theta) + \mathrm{r}.\cos(\phi).\mathrm{sen}(\theta)$$

Ou seja:

$$x' = x.\cos(\theta) - \mathrm{y}.\mathrm{sen}(\theta)$$
$$y' = y.\cos(\theta) + \mathrm{x}.\mathrm{sen}(\theta)$$

Essa operação também pode ser representada da seguinte forma:

$$[x' \; y'] = [x \; y] \begin{bmatrix} cos(\theta) & sen(\theta) \\ -sen(\theta) & cos(\theta) \end{bmatrix}$$

É importante salientar que no caso do objeto não estar definido na origem do sistema de coordenadas, a multiplicação das coordenadas do objeto por essa matriz também resultará em uma translação. Para rotacionar o objeto em relação a um ponto pivô deve-se transladar

todos os pontos do objeto em relação ao ponto pivô, depois aplicar a rotação e por fim aplicar a translação inversa para reposicionar o objeto. Esse mesmo procedimento pode ser feito para combinar as outras transformações em torno de um ponto [3].

Como cada estrutura é formada por um conjunto de pontos interligados, também foi criado um conjunto de rotinas que possibilita o ajuste de cada ponto dessa estrutura. O especialista também pode mover qualquer ponto da estrutura e adicionar mais pontos para tornar a estrutura mais precisa.

2.2 Marcação e edição dos pontos Cefalométricos:

A marcação dos pontos cefalométricos foi implementada de modo que tornasse a tarefa fácil para o usuário. Inicialmente foi criada uma base de dados em arquivos XML (*Extensible Markup Language*) para armazenar os nomes de todos os pontos cefalométricos que seriam utilizados na ferramenta. Esses pontos cefalométricos são carregados no programa de acordo com a análise que estiver sendo aplicada na imagem. Cada análise cefalométrica contém uma quantidade de pontos que devem ser marcados em locais específicos na imagem a depender das definições desta análise.

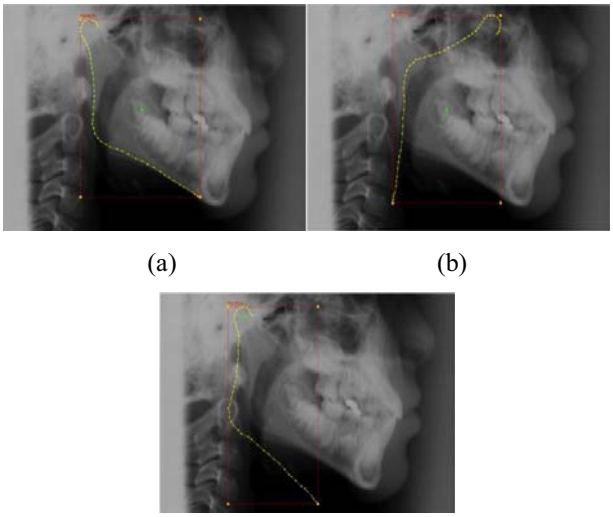
Ao abrir uma imagem para realizar a marcação de pontos, o usuário deve escolher uma análise entre as disponíveis para que sejam marcados apenas os pontos referentes a essa análise. Estes pontos aparecem em uma tabela do lado esquerdo da tela, e para realizar a marcação de um determinado ponto o usuário precisa apenas clicar no ponto desejado e clicar no local da imagem onde esse ponto deve ser posicionado. Caso a marcação tenha sido imprecisa, o usuário tem a opção de mover pontos para que ele possa mudar este ponto para outra região. Após marcar todos os pontos de uma análise, o usuário tem à sua disposição a função de gerar o traçado cefalométrico a partir dos pontos marcados. O traçado é gerado automaticamente a partir do algoritmo referente à análise escolhida. Cada análise possui uma forma de gerar seu traçado cefalométrico que é geralmente composto por retas ligando pontos, prolongamentos de retas, bissetrizes, ou outras relações geométricas que variam de acordo com a análise.

Como auxílio na marcação dos pontos, o usuário tem a opção de tornar visíveis na imagem as estruturas cefalométricas que fazem parte do desenho anatômico (partes do crânio como maxila, mandíbula, tecido mole e dentes). Geralmente os pontos cefalométricos se encontram sobre ou a uma distância conhecida de alguma dessas estruturas, elas podem auxiliar os profissionais na marcação dos pontos cefalométricos.

3. Resultados

A seguir são apresentados os resultado obtidos com a implementação das rotinas para marcação e edição do desenho anatômico e dos pontos cefalométricos, incorporadas à ferramenta computacional em desenvolvimento.

As Figuras 2-a, 2-b e 2-c mostram os resultados das rotinas de rotação. Na Figura 2-a a estrutura encontra-se na posição inicial, na Figura 2-b há uma rotação em torno do ponto pivô (ponto verde) na posição inicial, e na Figura 2-c tem-se uma rotação em torno do ponto pivô posicionado no canto superior esquerdo.

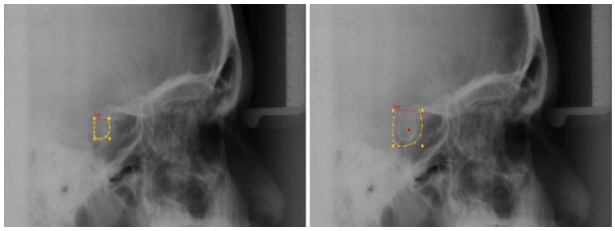


(a) (b)

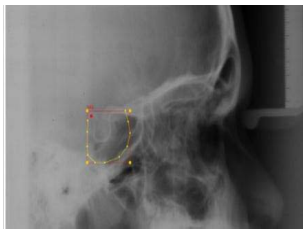
(c)

Figura 2: (a) Estrutura na posição inicial; (b) Estrutura rotacionada; (c) Estrutura rotacionada em relação ao pivô no canto superior esquerdo

Nas Figuras 3-a, 3-b e 3-c, observa-se a aplicação da transformação de escala em outra estrutura. Na Figura 3-a a estrutura está na posição inicial e as outras figuras são respectivamente a escala em torno do pivô (ponto vermelho) na posição inicial e a escala em torno do pivô no canto superior esquerdo.



(a) (b)



(c)

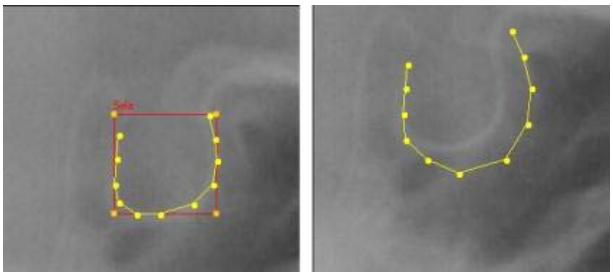
Figura 3: (a) Estrutura na posição inicial; (b) Escala aplicada na estrutura; (c) Escala aplicada na estrutura em relação ao pivô no canto superior esquerdo

Mostra-se nas Figuras 4-a e 4-b uma translação simples. A Figura 4-a mostra a estrutura em uma posição incorreta. A Figura 4-b já mostra a estrutura na posição correta.

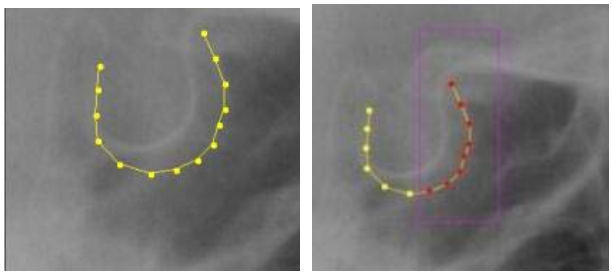
(a) **(b)**

Figura 4: (a) Estrutura na posição inicial; (b) Estrutura transladada para a posição correta

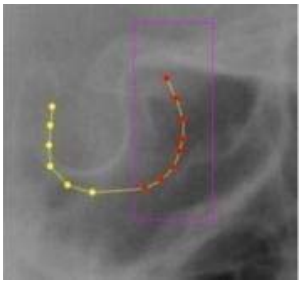
As Figuras 5-a, 5-b, 5-c, 5-d e 5-e são mostradas as rotinas de edição de pontos das estruturas. As Figuras 5-a, 5-b e 5-c mostram a possibilidade de mover cada ponto da estrutura e adicionar novos pontos na mesma. As Figuras 5-d e 5-e mostram as rotinas computacionais que possibilitam a seleção de um conjunto de pontos e posteriormente a movimentação deles.



(a) **(b)**



(c) **(d)**

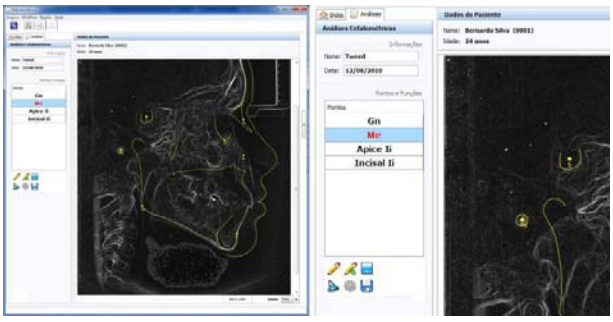


(e)

Figura 5: (a) Estrutura inicial; (b) Modificação nos pontos da estrutura; (c) Adicionando novos pontos à estrutura; (d) Seleção de um conjunto de pontos; (e) Movimentação do conjunto de pontos.

Na Figura 6 é apresentada a marcação dos pontos cefalométricos para a realização de uma análise cefalométrica no sistema. A Figura 6-a mostra uma imagem radiográfica onde alguns pontos cefalométricos já foram marcados. Nessa imagem foram aplicados efeitos de edição de imagem para facilitar a visualização da mesma, além dos contornos anatômicos que estão visíveis para facilitar a localização dos pontos. A Figura 6-b destaca o painel onde são mostrados os pontos a serem marcados na análise além de alguns botões que disponibilizam funcionalidades ao usuário, como marcar

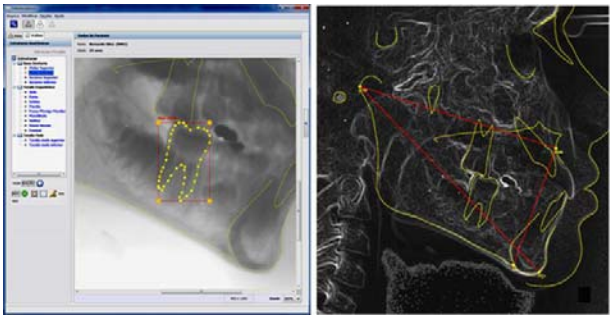
ponto, mover pontos, recursos de edição das estruturas cefalométricas (contornos anatômicos), geração de traçado entre outros. No painel é possível perceber que o ponto que está para ser marcado está realçado na lista.



(a) (b)

Figura 6: Marcação de pontos cefalométricos. (a) Análise cefalométrica sendo executada no sistema; (b) Painel que contém as funcionalidades para marcação de imagens.

A Figura 7-a mostra o recurso de edição das estruturas anatômicas, em que uma determinada estrutura (neste caso, o molar inferior) está sendo modificada para se adequar à imagem que está sendo utilizada. Essas estruturas irão servir como um referencial para o usuário realizar a marcação dos pontos. Na Figura 7-b é apresentado um Traçado Cefalométrico concluído em que os pontos foram posicionados em seu local correspondente e o traçado cefalométrico para a análise utilizada foi gerado (em vermelho).



(a) (b)

Figura 7: Funções relativas ao Traçado Cefalométrico. (a) Estrutura anatômica sendo editada para se adequar ao perfil do paciente; (b) Traçado cefalométrico realizado (marcas vermelhas) a partir da marcação dos pontos e do posicionamento das estruturas (marcas amarelas).

4. Conclusão

Os resultados apresentam rotinas para a realização do desenho anatômico e para a marcação dos pontos cefalométricos que foram incorporados a uma ferramenta computacional que realiza, entre outras funções, traçados cefalométricos.

O uso de uma referência de desenho anatômico auxilia muito o especialista, pois com as rotinas de ajustes de desenho anatômico é bem mais rápido ajustar um desenho a uma imagem de um novo paciente do que fazer todo o processo de desenho de todas as estruturas. Outra possibilidade que já está sendo implementada é a de se utilizar desenhos anatômicos de referência, separados por idade e sexo, onde o especialista escolhe de qual faixa de idade e sexo ele deseja obter o desenho inicial.

Neste trabalho os contornos são representados por um conjunto de pontos interligados entre si, sendo que esta abordagem pode fazer com que o especialista leve muito tempo na realização do ajuste do desenho anatômico até que ele se adeque a ferramenta. Uma alternativa para esse problema seria a utilização das curvas de Bézier para representar as estruturas. A curva de Bézier é uma curva polinomial que é interpolada através de três ou mais pontos, chamados pontos de controle. O grau da curva final depende do número de pontos de controle utilizados [3]. Então, ao invés de modelar as estruturas por um conjunto de pontos interligados, as estruturas anatômicas seriam representadas por um conjunto de curvas de Bézier, tornando mais rápido e mais preciso o processo de ajuste do desenho anatômico, pois, utilizando várias curvas de Bézier concatenadas é possível a criação ou o ajuste de contornos mais suaves com poucos pontos de controle.

Por fim, tanto a rotina de marcação e edição dos pontos cefalométricos como a de desenho anatômico, incorporadas à ferramenta em desenvolvimento, tornou possível a realização de análises cefalométricas já implementadas**.**

5. Referências

[1] Vedovello, M.F. 2007. Cefalometria: Técnicas de Diagnóstico e Procedimentos. 1.ed. São Paulo: Napoleão.

[2] Houston, W.J.B. (1982) A comparision of the reability of measurement of cephalometric radiographs by tracings and direct digitization. Swed Dent J., v.15, p.99-103.

[3] Azevedo E, Conci A. Computação gráfica: teoria e prática. São Paulo, Campus, 2003.

Retrieval of Labeled Shape Classes in Binary Images Using Standardized Granulometric Intervals

Alexandre Yukio Harano and Ronaldo Fumio Hashimoto
Institute of Mathematics and Statistics of University of São Paulo
Rua do Matão, 1010 - Cidade Universitária
São Paulo - SP - Brasil - CEP 05508-090
{ayharano,ronaldo}@ime.usp.br

Abstract

This work proposes an empirical non-interactive method to identify and retrieve Labeled Shape Classes from binary images. The method bases on granulometries extraction from connected components resized by standardized ratios. The extraction procedure is used to compound an array of representative granulometric intervals on setup phase, as well as to test the connected components on retrieval phase. The use of symmetrical convex structuring elements for granulometry is intended to provide rotation invariance for angles multiples of 90° while the standardized resizing is preconceived to provide scale invariance. The corresponding algorithms are presented, along with their results and proposed extensions.

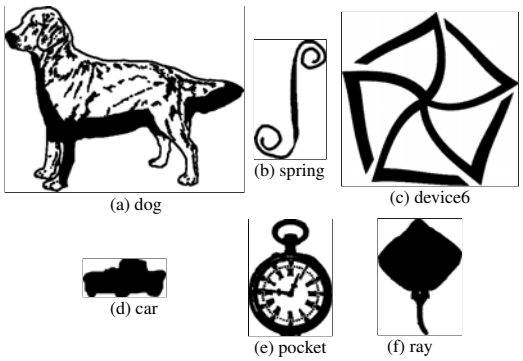


Figure 1. Example of LSCs from *CE-Shape-1 part B* (colors inverted for better visualization).

1. Introduction

A *Labeled Shape Class* (**LSC**) is a set of binary images in $\mathbb{Z}^2$ with the purpose of representing a shape as specified by an arbitrary given label, usually named after an object with similar appearance.

Typically, grayscale images [7] and color images [1] have enough information about the neighborhood similarity of their pixels to generate representative data for shape detection. In case of binary images, the problem becomes more difficult, given that these images do not contain a wealth of information available in other types of images.

The presented method was originally designed to recognize classes of the data set presented by [5], known as *CE-Shape-1 part B*. The data set is notorious for its labeled shape images, containing 70 LSCs and each class has 20 binary images in $\mathbb{Z}^2$, totalizing 1400 images. It is noteworthy that the goal of this method differs from the one presented in [5], so a different protocol was adopted.

Because of the scarcity of information aforementioned, there are few options to represent the data of a class. As a foundational decision, we chose to use the largest connected component of each image included in a class.

Due to the objective of recognizing shapes, we opted to take an approach based on the Mathematical Morphology (MM) framework. Considering the descriptive nature of the problem, a granulometry — also known as *Pattern Spectrum* [6] — meets the requirements. In the context of MM, instead of using a physical deposit grain, a structuring element is used to count the residue pixels removed after an opening operation performed on an image. The distribution of the residues corresponding to the size of the structuring element characterize an image.

The idea proposed in this article resides in representing each class by calculating the minimum and maximum values accepted for each position of a granulometry array indexed by its structuring element. The minimum and maximum values are calculated over the granulometries of the largest connected component of each image from the class. During the retrieval phase, those intervals are used as filters to eliminate unrelated components in an efficient man-

ner.

Targeting to provide invariance to rotation of angles multiples of 90° and improve the robustness to the result, we chose to restrict the structuring elements in such manner that they are required to be symmetrical around both axes. Only rotation of angles multiples of 90° fully preserves information related to each pixel's adjacency, which is important for the discrete approach adopted for binary images in MM.

Given that the sizes of the images provided by CE-Shape-1 part B data set do not follow a rule, an exception to the discrete approach — taken in MM for binary images and their pixels — was designated: form a grayscale image using a bicubic interpolation and generate a binary image after applying a threshold value. The procedure endeavors to standardize the sizes of the connected components from which the granulometries will be calculated.

Owing to the fact that an interpolated resize is used to standardize the connected components, as previously stated, this method attempts to achieve scale transform invariance.

After identifying a component as belonging to a LSC, an Euclidean distance interval metric is used to establish which of the surrounding components also qualify as part of the LSC.

2. Algorithm

The proposed algorithm for retrieval of LSCs in binary images is separated into two parts: the setup of each class representative data and the retrieval of LSCs in binary images itself.

Aiming for the reproducibility of the results obtained during the setup of each class representative data in the retrieval algorithm, we describe a standard procedure named data extraction of a connected component. It is worth mentioning that the connectivity used to determine the neighborhood of a pixel to calculate the connected components is the 8-adjacency [3].

Let S denote an array of integers, whose values are used to define the ratio of resized images; a and z respectively denote the minimum and the maximum values of S; and E denote an array of symmetrical convex structuring elements in $\mathbb{Z}^2$, which convexity is defined according to [8] and their symmetries are detected on both horizontal and vertical axes, e.g., *city-block*, *chessboard* and *octagon* [2].

2.1. Data extraction of a connected component

Let g denote an integer in range of $\big[0,\ 254\big]$; C denote a list of Cartesian coordinates in $\mathbb{Z}^2$; and P denote an array of granulometries indexed by pairs composed by an element of S and an element of E.

The integer g represents a threshold value used to binarize a grayscale image; C represents a connected component and P represents the component's standardized granulometries.

1. Initialize P as an array of invalid granulometries values indexed by a pair composed by an element of S and an element of E.
2. Form a binary image B from the bounding box of C of size $w \times h$. Let m be the maximum value between w and h and let r_z be the ratio of z to m.
3. **(Standardized resizing)** For each integer $s \in S$, let r_s be the ratio of s to m; let G_s form a grayscale image with size $\lfloor w \cdot r_s \rfloor \times \lfloor h \cdot r_s \rfloor$ as the result of a bicubic interpolation from scaling B; and let R_t form a binary image with same size of R_s as the result of the following: $(x, y) \in R_s \iff G_s(x, y) > g$.
4. **(Standardized granulometries calculation)** For each integer $s \in S$ and each structuring element $e \in E$, assign the results of the granulometry of R_s by e to $P_{s,e}$.

   This procedure returns P and r_z.

2.2. Setup of class representative data

Let I denote an array of binary images in $\mathbb{Z}^2$; t denote an integer in range of $\big[0,\ 254\big]$; D denote an array indexed by I containing arrays of integers indexed by pairs composed by an element of S and an element of E; A denote an array of real numbers indexed by I; r_{min} and r_{max} denote real numbers; P_{min} and P_{max} denote arrays of real numbers indexed by pairs composed by an element of S and an element of E.

I represents an array of binary images of the LSC from which is desired to generate the representative data; t represents a threshold value to obtain a binary image from an interpolated grayscale image; D and A represent arrays to store the output granulometries and ratios, respectively, of the data extraction from the largest connected component of each image $i \in I$; r_{min} and r_{max} represent the minimum and maximum standardized Euclidean distance for class pixels' recovery in retrieval phase; and P_{min} and P_{max} represent the minimum and maximum values accepted for granulometries indexed by each possible pair composed by an element of S and an element of E.

1. Initialize P_{min} and P_{max} as arrays filled with invalid integer values indexed by a pair composed by an element of S and an element of E.
2. For each binary image $i \in I$, list the Cartesian coordinates of the largest connected component of i as C_i, whereas the largest is defined as the component with greatest number of elements. In case of two or more

components with same number of elements, pick one at random.

3. For each list C_i, calculate c_i as the centroid of the coordinates in C_i; calculate $d_{\text{min},i}$ as the minimum distance between the coordinates of i and c_i; calculate $d_{\text{max},i}$ as the maximum distance between the coordinates of i and c_i; and assign the resulting P and r_z of data extraction of C_i using t as g (Subsection 2.1) to, respectively, D_i and A_i.
4. **(Standardized Euclidean distance range assignment)** Assign the minimum value among $A_i \cdot d_{\text{min},i}$, $\forall\, i \in I$, to r_{min} and the maximum value among $A_i \cdot d_{\text{max},i}$, $\forall\, i \in I$, to r_{max}.
5. **(Standardized granulometries range assignment)** For each binary image $i \in I$, if P_{min} and P_{max} still have invalid values, assign D_i to both; otherwise, do the following:

(a) For each integer $s \in S$ and for each structuring element $e \in E$: let m_{min} and m_{max} be, respectively, the minimum and the maximum values of $|D_{i,(s,e)}|$ and $|P_{\text{min},(s,e)}|$, whereas $|Z|$ means the number of elements in array Z; and let N_{min} and N_{max} be arrays of integers with m_{max} elements.

i. For each integer j within range of $\big[0, m_{\text{min}}\big[$, assign the minimum of $D_{i,(s,e)}[j]$ and $P_{\text{min},(s,e)}[j]$ to $N_{\text{min}}[j]$ and assign the maximum of $D_{i,(s,e)}[j]$ and $P_{\text{max},(s,e)}[j]$ to $N_{\text{max}}[j]$.

ii. For each integer j within range of $\big[m_{\text{min}}, m_{\text{max}}\big[$, assign 0 to $N_{\text{min}}[j]$ and if $|D_{i,(s,e)}| > |P_{\text{max},(s,e)}|$, then assign $D_{i,(s,e)}[j]$ to $N_{\text{max}}[j]$, otherwise assign $P_{\text{max},(s,e)}[j]$ to $N_{\text{max}}[j]$.

iii. Assign N_{min} to $P_{\text{min},(s,e)}$ and N_{max} to $P_{\text{max},(s,e)}$.

This procedure returns r_{min}, r_{max}, P_{min} and P_{max}.

2.3. Retrieval of LSCs in binary images

Let X denote a binary image in $\mathbb{Z}^2$; L denote an array of labels; R_{min} and R_{max} denote arrays of real numbers indexed by L; G_{min} and G_{max} denote arrays indexed by L containing arrays indexed by pairs composed by an element of S and an element of E; Y denote an array of binary images in $\mathbb{Z}^2$ indexed by L.

X represents the image from which is desired to extract the LSCs' related images. L represents the desired LSCs. R_{min}, R_{max}, G_{min} and G_{max} represent arrays to store the representative data for each label $l \in L$. Y represents the output images according to each label $l \in L$. The integer a is the minimum value of S, as stated at the beginning of this section.

1. For each label $l \in L$, initialize Y_l as an empty binary image with same size as X.
2. Initialize C as an empty list. For each connected component of X, list its Cartesian coordinates: if it has at least a^2 elements, insert the list into C. This value was experimentally chosen to be used as a threshold limit to distinguish candidate components from noise components.
3. For each list of coordinates $c \in C$, do the following:

(a) Let P and r_z be the output of data extraction of c (Subsection 2.1).

(b) For each label $l \in L$, do the following:

i. Assign $G_{\text{min},l}$ to P_{min}; $G_{\text{max},l}$ to P_{max}; and *True* to v.

ii. **(Standardized granulometries range verification)** For each integer $s \in S$ and for each structuring element $e \in E$, if v still holds *True*, then:

A. Let m_{min} and m_{max} be, respectively, the minimum and the maximum values of $|P_{(s,e)}|$ and $|P_{\text{min},(s,e)}|$.

B. If $|P_{\text{min},(s,e)}| < |P_{(s,e)}|$, then assign *False* to v.

C. If v still holds *True*, for each integer j within range of $\big[0, m_{\text{min}}\big[$, verify if v still holds *True* and $P_{(s,e)}[j]$ is within range of $\big[P_{\text{min},(s,e)}[j], P_{\text{max},(s,e)}[j]\big]$: in case of failure, assign *False* to v.

D. If v still holds *True*, for each integer j within range of $\big[m_{\text{min}}, m_{\text{max}}\big[$, verify if v still holds *True* and $P_{\text{min},(s,e)}[j]$ equals 0: in case of failure, assign *False* to v.

iii. **(Class identification)** If v still holds *True*, then c was identified as a component belonging to label l, requiring to recover its surrounding components as described next:

A. Let (c_x, c_y) be the Cartesian coordinate of c's centroid; d_{min} be the ratio of $(R_{\text{min},l} - 1)$ to r_z; and d_{max} be the ratio of $(R_{\text{max},l} + 1)$ to r_z.

B. Form a binary image T with same size of X and containing the elements of X located inside the bounding box delimited by coordinates $(\lfloor c_x - d_{\text{max}} - 1 \rfloor, \lfloor c_y - d_{\text{max}} - 1 \rfloor)$ and $(\lceil c_x + d_{\text{max}} + 1 \rceil, \lfloor c_y + d_{\text{max}} + 1 \rceil)$.

C. Form binary image W as the *inf-reconstruction* [2] of T by the box structuring element of radius 1 conditioned by X.

D. **(Standardized Euclidean distance range verification)** For each connected component of W, verify if the Euclidean distance between each pixel of the component and (c_x, c_y) is within range of $\big[d_{\text{min}} - 1, d_{\text{max}} + 1\big]$: in case of failure, remove the whole component from W.

E. **(Recovery)** Assign $Y_l \cup W$ to Y_l.

This procedure returns Y.

3. Experimental procedure

3.1. Setting the experimental environment

The experiment was designed to use the CE-Shape-1 part B data set. The data set has 1400 GIF image files which can be split into 70 groups with 20 images each, according to the prefix of the file name.

Initially, create a file structure divided by the given prefix and place the corresponding image files into its label directory. For each image file, replace it with an image file which contains an image formed from the content of the minimum bounding box that covers all foreground pixels of the original image.

The variables values used for the algorithms in the experimental procedure are listed below according to the location of their definition:

(Section 2)

$S = [16,\ 32,\ 64,\ 128,\ 256]$; $a = 16$; $z = 256$; and $B = [\textit{city-block}, \textit{chessboard}, \textit{octagon}]$.

(Subsection 2.2)

$t = 127$.

(Step 3a of Subsection 2.3)

$g = 127$.

Let L be an array of labels defined by each directory given name; I be an array of array of binary images indexed by L; and D be an array of representative data indexed by L.

I contains the images with reduced size, sorted according to their labels, and D will contain the representative data of each label at the termination of the following: for each label $l \in L$, assign the output of the setup procedure of I_l to D_l (Subsection 2.2).

3.2. Generating test images

First, choose 4 images from I in a random and independent manner. From the chosen images, generate 4 matrices of images, each one as a 2×2 image grid, arranged in a random order. The minimum space between rows and between columns is one pixel and such proximity was chosen to stimulate possible false results in the retrieval process. Match each generated matrix with one of the transform listed below:

1. The unmodified image.
2. Let k denote a randomly chosen integer in range of $\big[1,\ 3\big]$. Rotate the image by an angle of $k \cdot 90°$.
3. Let s denote a randomly chosen integer in range of $\big[95,\ 300\big]$. Scale the image by a ratio of $s \cdot 0.01$ using bicubic interpolation and forming a binary image like the procedure described in Step 3 of the algorithm presented at Subsection 2.1, using $g = 238$.
4. The image is affected by a combination of the two aforementioned transforms.

Apply the described transform to each element of the corresponding matrix and form a binary image for each matrix.

3.3. Defining experimental evaluation

During the composition of a set of test images, its ground truth images are also generated for evaluation purposes, given that a test image is artificially made. A ground truth image is produced for each label used in the test image.

For each test image i, list the retrieved and the ground truth images and labels. Every image in the list and the test image itself have size $W_i \times H_i$. For each label ℓ in the list, count the number of correct pixels; missing pixels; extra pixels and remaining pixels. Let $Y_{i,\ell}$ be the retrieved image for a label ℓ and $T_{i,\ell}$ be the ground truth image for the corresponding label.

The correct pixels are the pixels detected in a retrieved image present in the ground truth image (true positive – $\mathrm{TP}_{i,\ell}$ – Eq. 1); the extra pixels are the pixels detected in a retrieved image but not present in the ground truth image (false positive – $\mathrm{FP}_{i,\ell}$ – Eq. 2); the missing pixels are the pixels not detected in a retrieved image but present in the ground truth image (false negative – $\mathrm{FN}_{i,\ell}$ – Eq. 3); and the remaining pixels are the true negative ($\mathrm{TN}_{i,\ell}$ – Eq. 4) ones [4].

The following ratios are used to measure the performance of the method on the pair of images. Accuracy is the ratio of the sum of true results to the sum of each result ($\mathrm{A}_{i,\ell}$ – Eq. 5) and recall is the ratio of true positive results to the sum of the predicted (ground truth) positive results ($\mathrm{R}_{i,\ell}$ – Eq. 6).

$$\mathrm{TP}_{i,\ell} = |Y_{i,\ell} \cap T_{i,\ell}| \tag{1}$$

$$\mathrm{FP}_{i,\ell} = |Y_{i,\ell} \setminus T_{i,\ell}| \tag{2}$$

$$\mathrm{FN}_{i,\ell} = |T_{i,\ell} \setminus Y_{i,\ell}| \tag{3}$$

$$\mathrm{TN}_{i,\ell} = W_i \cdot H_i - \mathrm{TP}_{i,\ell} - \mathrm{FP}_{i,\ell} - \mathrm{FN}_{i,\ell} \tag{4}$$

$$\mathrm{A}_{i,\ell} = \frac{\mathrm{TP}_{i,\ell} + \mathrm{TN}_{i,\ell}}{\mathrm{TP}_{i,\ell} + \mathrm{TN}_{i,\ell} + \mathrm{FP}_{i,\ell} + \mathrm{FN}_{i,\ell}} \tag{5}$$

$$\mathrm{R}_{i,\ell} = \frac{\mathrm{TP}_{i,\ell}}{\mathrm{TP}_{i,\ell} + \mathrm{FN}_{i,\ell}} \tag{6}$$

The equations described above applies to a pair of images $Y_{i,\ell}$ and $T_{i,\ell}$ for a single test image i. The measurements given by Eq. 7 to Eq. 12 are calculated and evaluates the method for a set of test images, indexed by the labels which were present or retrieved.

$$\mathrm{TP}_\ell = \sum_i \mathrm{TP}_{i,\ell} \tag{7}$$

$$\mathrm{FP}_\ell = \sum_i \mathrm{FP}_{i,\ell} \tag{8}$$

$$\mathrm{FN}_\ell = \sum_i \mathrm{FN}_{i,\ell} \tag{9}$$

$$\mathrm{TN}_\ell = \sum_i \mathrm{TN}_{i,\ell} \tag{10}$$

$$\mathrm{A}_\ell = \frac{\sum_i (\mathrm{TP}_{i,\ell} + \mathrm{TN}_{i,\ell})}{\sum_i (\mathrm{TP}_{i,\ell} + \mathrm{TN}_{i,\ell} + \mathrm{FP}_{i,\ell} + \mathrm{FN}_{i,\ell})} \tag{11}$$

$$\mathrm{R}_\ell = \frac{\sum_i \mathrm{TP}_{i,\ell}}{\sum_i (\mathrm{TP}_{i,\ell} + \mathrm{FN}_{i,\ell})} \tag{12}$$

4. Results

For each listed transform, as defined in Subsection 3.2, we generated 30 images.

Among the identified and the ground truth labels, there was 54 LSCs. The results are divided according to the listed transforms. Each table displays the values related to labels which recall rates were not zero.

For the case of test images composed by unmodified images, each label was correctly identified and recovered as we can observe Table 1, as each label has a recall rate of 1.0.

The method was expected to provide invariance for angles multiple of 90° due to the use of structuring elements which are required to be symmetrical around both axes. However, the experimental results observed in Table 2 does not reflect such behavior since the recall rate of 1.0 was not achieved for each one of the 54 LSCs arranged in the test images.

As expected, the images affected by the scale transform obtained a worse performance then the ones affected only by the rotation transform as it can be verified in Table 3.

Accordingly, the test images created from the combination of both transforms had unsatisfactory results, as showed in Table 4.

A proper choice for the threshold level variables t and g presented in Subsection 3.1 is needed to obtain a better performance for the scale transform.

Label	TP_ℓ	FN_ℓ	FP_ℓ	TN_ℓ	A_ℓ	R_ℓ
Bone	90089	0	32	1715225	1.0000	1.0000
Comma	128682	0	0	1453424	1.0000	1.0000
Glas	102013	0	0	931595	1.0000	1.0000
HCircle	42495	0	0	641825	1.0000	1.0000
Heart	106395	0	0	433053	1.0000	1.0000
apple	31066	0	0	833785	1.0000	1.0000
bat	148732	0	0	1418875	1.0000	1.0000
beetle	34227	0	0	1550376	1.0000	1.0000
bell	153488	0	0	2060082	1.0000	1.0000
bird	144938	0	0	1162157	1.0000	1.0000
bottle	20229	0	0	1949896	1.0000	1.0000
brick	51870	0	0	1239233	1.0000	1.0000
butterfly	249461	0	119164	2224475	0.9540	1.0000
camel	46371	0	0	385247	1.0000	1.0000
car	34395	0	0	1052818	1.0000	1.0000
cattle	282905	0	0	2531189	1.0000	1.0000
chicken	75975	0	0	1101984	1.0000	1.0000
chopper	26995	0	0	1378058	1.0000	1.0000
classic	473898	0	0	1779381	1.0000	1.0000
crown	23362	0	215379	1980185	0.9029	1.0000
cup	54491	0	0	439669	1.0000	1.0000
device0	199348	0	0	1139304	1.0000	1.0000
device2	74720	0	0	790131	1.0000	1.0000
device4	85003	0	65344	1333884	0.9560	1.0000
device5	39576	0	0	579804	1.0000	1.0000
device6	57636	0	285968	2054366	0.8807	1.0000
device7	100743	0	0	764108	1.0000	1.0000
device8	75053	0	0	1293550	1.0000	1.0000
device9	130488	0	0	334343	1.0000	1.0000
dog	62687	0	147620	2778206	0.9506	1.0000
elephant	274778	0	3823	1005829	0.9970	1.0000
face	38762	0	0	481366	1.0000	1.0000
fish	5406	0	0	711849	1.0000	1.0000
flatfish	153262	0	37	704601	1.0000	1.0000
fly	37499	0	0	924075	1.0000	1.0000
fork	26795	0	0	592585	1.0000	1.0000
frog	41779	0	0	337059	1.0000	1.0000
guitar	184826	0	26507	2799741	0.9912	1.0000
hammer	20372	0	0	1677595	1.0000	1.0000
hat	175738	0	9472	4241121	0.9979	1.0000
horse	202241	0	0	2401876	1.0000	1.0000
horseshoe	28002	0	12250	3114272	0.9961	1.0000
jar	170174	0	0	1423134	1.0000	1.0000
key	24683	0	0	880909	1.0000	1.0000
lizzard	93972	0	0	2089530	1.0000	1.0000
lmfish	91289	0	0	593031	1.0000	1.0000
octopus	68736	0	0	1016069	1.0000	1.0000
personal_car	52987	0	0	216077	1.0000	1.0000
pocket	44964	0	523240	6644062	0.9275	1.0000
ray	100017	0	0	788387	1.0000	1.0000
sea_snake	53101	0	0	1805660	1.0000	1.0000
shoe	131619	0	0	920579	1.0000	1.0000
spring	51026	0	0	1417364	1.0000	1.0000
tree	30239	0	9702	1530328	0.9938	1.0000

Table 1. Results for unmodified images.

5. Conclusion

This article presents an empirical method related to identification and retrieval of LSCs and tries to guarantee that at least every pixel belonging to a connected component located within the normalized Euclidean distance range of the class element will be recovered during retrieval phase.

The method is mostly based on a discrete approach for images and their contents. It should provide invariance for angles multiple of 90°, however, according to the experimental results, such behavior was not accomplished. The reasons behind the non achievement will be carefully studied to improve the proposed method.

The extension addressed to provide scale invariance is tightly related to the threshold levels adopted to generate

Label	TP_ℓ	FN_ℓ	FP_ℓ	TN_ℓ	A_ℓ	R_ℓ
Bone	90089	0	0	1847441	1.0000	1.0000
Comma	128682	0	0	1492632	1.0000	1.0000
HCircle	42495	0	0	693136	1.0000	1.0000
Heart	106395	0	0	431805	1.0000	1.0000
bell	114501	38987	0	2289605	0.9840	0.7460
bird	72183	72755	0	1194806	0.9457	0.4980
butterfly	82554	166907	119164	2116909	0.8849	0.3309
camel	46371	0	0	477885	1.0000	1.0000
car	13380	21015	0	1095970	0.9814	0.3890
cattle	112512	170393	0	2374115	0.9359	0.3977
chicken	39492	36483	0	1049498	0.9676	0.5198
classic	473898	0	0	1843514	1.0000	1.0000
crown	13580	9782	215379	1662784	0.8816	0.5813
cup	54491	0	0	518763	1.0000	1.0000
device0	78932	120416	0	1172087	0.9122	0.3960
device5	39576	0	0	386523	1.0000	1.0000
dog	62687	0	147620	2623152	0.9479	1.0000
elephant	274778	0	0	1019244	1.0000	1.0000
fly	20066	17433	0	1057781	0.9841	0.5351
fork	26795	0	0	399304	1.0000	1.0000
guitar	132060	52766	33300	2414511	0.9673	0.7145
hammer	20372	0	0	1416077	1.0000	1.0000
hat	113254	62484	9472	4230211	0.9837	0.6444
horse	31013	171228	0	2281972	0.9312	0.1533
horseshoe	14862	13140	12250	3022575	0.9917	0.5307
jar	170174	0	0	1454977	1.0000	1.0000
key	24683	0	0	885565	1.0000	1.0000
lizzard	66976	26996	0	2114841	0.9878	0.7127
lmfish	91289	0	0	644342	1.0000	1.0000
personal_car	52987	0	0	230322	1.0000	1.0000
pocket	44964	0	540567	6699910	0.9258	1.0000
ray	28296	71721	0	834475	0.9232	0.2829
sea_snake	9048	44053	0	1848688	0.9768	0.1704
shoe	51020	80599	0	931983	0.9242	0.3876
spring	6214	44812	0	1463933	0.9704	0.1218
tree	20225	10014	0	833239	0.9884	0.6688

Table 2. Results for rotated images.

Label	TP_ℓ	FN_ℓ	FP_ℓ	TN_ℓ	A_ℓ	R_ℓ
Comma	77140	400418	0	7082178	0.9470	0.1615
bell	91908	400203	0	10063441	0.9621	0.1868
dog	1621	119305	714469	11284587	0.9312	0.0134
fish	10370	0	256693	4023703	0.9402	1.0000
fork	35391	0	0	1126634	1.0000	1.0000
guitar	393452	373623	34116	9422656	0.9602	0.5129
lizzard	216812	127048	0	6786376	0.9822	0.6305
lmfish	302088	0	0	1668456	1.0000	1.0000
octopus	220036	247325	0	4214409	0.9472	0.4708
personal_car	80383	0	0	561567	1.0000	1.0000
pocket	124238	35557	1614630	26863112	0.9424	0.7775
sea_snake	11393	126910	0	8443581	0.9852	0.0824
shoe	508978	0	0	4188566	1.0000	1.0000

Table 3. Results for scaled images.

a binary image from a standardized interpolated grayscale image or even another possible scaling method, which was not the focus of the article. The effects of the choice for the threshold levels require further research on the subject.

The proposed method uses interval filters to identify parts of an image. The intervals are rigidly composed by the provided images of a LSC and were designed to select the component as a binary decision, i.e., it belongs to a LSC or not. Instead of intervals, one may find the use of means and standard deviations more suitable.

Label	TP_ℓ	FN_ℓ	FP_ℓ	TN_ℓ	A_ℓ	R_ℓ
bell	700466	317580	0	9810168	0.9707	0.6880
brick	129750	151945	0	4583342	0.9688	0.4606
butterfly	603924	962056	565752	13305230	0.9010	0.3857
crown	14485	109218	0	3643016	0.9710	0.1171
cup	99320	0	0	2527924	1.0000	1.0000
elephant	131383	1397242	0	6562733	0.8273	0.0859
fork	140658	0	0	1841735	1.0000	1.0000
guitar	349319	461598	60902	9890847	0.9515	0.4308
hat	371074	230729	871768	22278092	0.9536	0.6166
lmfish	339362	0	11589	3784933	0.9972	1.0000
octopus	117125	95045	0	7032198	0.9869	0.5520
personal_car	265190	0	0	1077274	1.0000	1.0000
pocket	32760	221829	2412630	29943747	0.9192	0.1287
sea_snake	32094	83529	0	9586100	0.9914	0.2776

Table 4. Results for images affected by rotation and scale.

The values presented in this article can be easily adapted and the proposed method can be employed if the representative images for LSCs are selected by a morphology specialist of the desired application.

6. Acknowledgments

This work was supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) under Grants 137135/2009-2 and 307376/2010-9; and by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) under Grant 2009/04274-2.

References

[1] E. Aptoula and S. Lefèvre. Morphological description of color images for content-based image retrieval. *Trans. Img. Proc.*, 18:2505–2517, November 2009.

[2] E. R. Dougherty and R. A. Lotufo. *Hands-on Morphological Image Processing*. SPIE Press, Bellingham, WA, 2003.

[3] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[4] R. Kohavi and F. Provost. Glossary of terms. *Machine Learning*, 30:271–274, 1998.

[5] L. Latecki, R. Lakämper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:1424, 2000.

[6] P. Maragos. Pattern spectrum and multiscale shape representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:701–716, 1989.

[7] E. R. Urbach, J. B. T. M. Roerdink, and M. H. F. Wilkinson. Connected shape-size pattern spectra for rotation and scale-invariant classification of gray-scale images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:272–285, February 2007.

[8] J. Xu. Decomposition of convex polygonal morphological structuring elements into neighborhood subsets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:153–162, 1991.

NURBS Interpolation with FPGA Acceleration

Luiz Marcelo Chiesse da Silva
Federal University of Technology - Paraná
Electrical Engineering Department
Cornélio Procópio - Paraná - Brazil
chiesse@utfpr.edu.br

Maria Stela Veludo de Paiva
USP - University of São Paulo
Electrical Engineering Department
São Carlos - São Paulo - Brazil
mstela@sel.eesc.usp.br

Abstract

NURBS are used mainly to represent 3D models, being included in standards like OpenGL (Phigs) and used in computer graphics and manufacturing systems, representing adequately, in a compact way, almost any shape. For this sake, NURBS algorithms are implemented in graphical processing units (GPUs) hardware, but its rigid processing architecture and power consumption makes it inadequate in embedded systems implementations. An alternative to this case, instead of another technologies like microcontrollers or dedicated integrated circuits, is the use of reconfigurable logic with FPGAs, more adequate to embedded systems in relation to GPUs. This work proposes a Soc - System on a chip implementation, using FPGA and the Verilog sysnthesis language, aiming implement an embedded system for computer graphics NURBS interpolation tools and visualization.

1. Introduction

Curve or surface interpolation is a fundamental task in graphic systems, like CAD/CAM, for example, where the resolution between the design and the manufacturing systems should be adjusted [17]. When a set of points are given or received from a graphical unit, and is essential to fit this set with a curve or straight lines coincident with the given points, is done the interpolation (if the points are not coincident, the fitting is made by aproximation) [14]. There are several methods for the interpolation of a set of points, ranging from simple and efficient triangulations [8] to modified methods using RBF [3] and others [7]. Among these methods, NURBS - Non Uniform Rational B-Splines are adopted in graphic standards like IGES [11], STEP [1] and OpenGL [9] (PHIGS) for curve and surface representation between graphical systems. The main advantages of the rational b-splines (affine transformations, for example) make them the most suitable choice for standardization, despite the lack of compression in the representation of conic sections [13], is widely used too in generic mathematical applications. The use of piecewise polynomials require a minimal number of procedures, namely orderly parameterization, linear system solution and the curve/surface fitting. Manufacturing systems, like CNC, and 3D data acquisition systems makes use of NURBS to provide greater efficience, being implemented in embedded systems [10] based in microcontrollers, DSPs or a application specific integrated circuit. The FPGA technology provide an all-in-one chip solution for the data pre-processing and control in embedded systems, an architecture specified by the system designer, and reconfigurable logic, capable to perform a custom processor [16] for a specific task. This work proposes a Soc in FPGA system, with NURBS local curve and surface interpolation cores, based in the fast Cox-de Boor implementation [18], and a basic graphic pipeline [6], for visualization purposes. Optionally, is included two cores for the generation of straight lines and circles [2]. A Wishbone based bus [12] is used to connect and synchronize the cores, following its conventions of a open source logic bus. The use of FPGAs in the area of video and image processing is consolidated [4], despite the fact of current technologies in this area of application, like the Cell processor [15] [5] integrated circuit, there is a gap in the graphics processing that leads to aplications making use of mixed technologies, like GPGPUs and FPGAs working together.

2. NURBS

A NURBS curve of degree *p* is a piecewise polynomial curve defined as:

$$C(u) = \sum_{i=0}^{n} w_i P_i N_{i,p}(u) \qquad (1)$$

where *u* is the parameter value, P_i form the so called control polygon points, weighted by w_i, and $N_{i,p}(u)$, *i*= 0,...,n, are the B-spline basis functions defined over the knot vec-

tor U, where:

$$U = \{u_0, u_1, ..., u_m\},\ u_i \leq u_{i+1},\ i = 0, ..., m-1 \quad (2)$$

$$N_{i,p}(u) = \begin{cases} 1 & \text{if} \quad u_i \leq u_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u) \quad (4)$$

We assume throughout this paper that the knot vector has the following form:

$$U = \{\underbrace{a, a, ..., a}_{p+1}, u_{p+1}, u_p, ..., u_{m-p-1}, \underbrace{b, b, ..., b}_{p+1}\} \quad (5)$$

where, in most practical applications, $a = 0$ and $b = 1$. A NURBS surface of degree (*p*,*q*) is defined similarly as:

$$S(u, v) = \sum_{i=0}^{n} \sum_{j=0}^{m} w_{i,j} P_{i,j} N_{i,p}(u) N_{j,q}(v) \quad (6)$$

where *u* and *v* are the parameter values in the longitudinal and isoparametric directions of surface construction, $P_{i,j}$, $i = 0, ..., n$; $j = 0, ..., m$, form the so-called control net defined by a set of points weighted by $w_{i,j}$ and the basis functions $N_{i,p}(u)$, *i* = 0, ... , n, and $N_{j,q}(v)$, *j* = 0, ... , m, are defined as above (the construction of $N_{j,q}(v)$ is similar), over the knot vectors:

$$U = \{u_0, u_1, ..., u_r\},\ u_i \leq u_{i+1},\ i = 0, ..., r-1 \quad (7)$$

$$V = \{v_0, v_1, ..., v_s\},\ v_j \leq v_{j+1},\ j = 0, ..., s-1 \quad (8)$$

2.1. Cox-de Boor Algorithm

The NURBS interpolation could be divided in local and global interpolation methods. The first constructs a curve by rational segments (rational polynomials), in the case of surfaces by rational patches, such that the endpoints of each segment are the given data points. Neighboor segments are joined with some continuity level between the junctions, with the curve construction proceeding segment wise. The global interpolation makes the curve as a whole, using all the given points in a matrix calculation, and the control points are obtained by the inversion in the matrix form of the NURBS equation (1). For both methods, is necessary the points to be interpolated (data points), the number of control points (segments), the knot vector and the parameter values. In figure 1, D_i are the interpolated data points, and the knots are calculated based in the chord length between each data point given by the module of vectors q_i. Another method for the interpolation is to satisfy the conditions given for the curve tangent vectors T_i in each data point D_i. For a given set of data points, the best method to set up the knots is calculating initial parameters values given by the chord length method:

$$t_0 = 0\ ;\ t_n = 1 \quad (9)$$

$$t_k = \frac{1}{L} \sum_{i=1}^{k} |D_i - D_{i-1}| \quad (10)$$

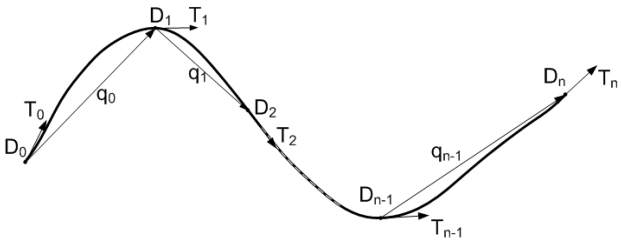


Figure 1. Data points, junctions (knots), distance and tangent vectors in a NURBS curve.

where *t* is the parameter value, *L* the total chord length between the data points, $|D_i - D_{i-1}|$ the chord length between two adjacent data points, *k* the paramater index and *n* the number of data points. For the knot vector is recommended the technique of averaging:

$$u_0 = \ldots = u_p = 0\ ;\ u_{m-p} = \ldots = u_m = 1 \quad (11)$$

$$u_{j+p} = \frac{1}{p} \sum_{i=j}^{j+p-1} t_i \quad (12)$$

where *t* is the parameter in the equation (10). The values u_0 at u_p and u_{m-p} at u_m reflects the knot multiplicities required for the spline beginning and end conditions. The property of local control allows the local interpolation, given by the suport region of the basis functions, restricting the influence of the basis functions only in a limited number of piecewise polynomials. So, many polynomials segments can be computed concurrently to generate the final curve, given the desired continuity at the polinomials junctions (knots), feature exploited by the de Cox-de Boor algorithm given in equation (13), where *t* is the parameter, *u* are the knots, *P* the control points for each layer (data points in the first layer), *i* the control point index in the layer, *k* the order of the polynomial segments and *j* the layer number.

$$C_i^j(t) = \left(1 - \frac{t - u_i}{u_{i+k-j} - u_i}\right) P_{i-1}^{j-1} + \left(\frac{t - u_i}{u_{i+k-j} - u_i}\right) P_i^{j-1} \quad (13)$$

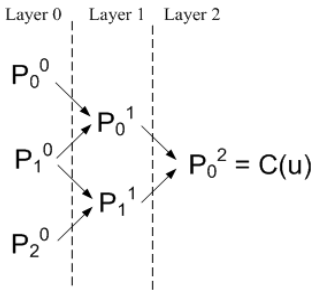


Figure 2. Cox-de Boor algorithm control point layers.

The control point in a b-spline curve is the convex combination of another two control points in the previous layer, as illustrated in figure 2, and its influence in the curvature is given by the term in brackets in equation (13). If a control point and the respective knot are repeated, the curvature tends towards its position, until the curve pass through the point, resulting in the interpolation.

3. Graphics Pipeline

A basic visualization pipeline is used, and by the sequential nature, this process is divided in serial stages, whose number could vary between implementations, but follows the general arrangement given by fig 3. In computer graphics systems, the last three stages are managed by a API - Application Programming Interface. The vertex reader reads 3D data obtained from a cloud of points, in a form of a triple indicating the euclidean coordinates of the points. This data is stored in a RAM memory and, according to the primitive drawer cores, the required points are bufferized in the FPGA. In the viewport transformation stage, the computation of each transformation matrix could be parallelized, but between them are serialized. The primitive drawer include the NURBS, straight line and circle generation cores. The video memory is made in a built-in RAM, for the purpose of different display resolutions suport and faster data transfer, executing the scanning to a D/A video converter independently of the remaining systems. The vertex reader is the data input of the pipeline, obtaining the data in the form of coordinates in euclidean space, from another built-in memory of the FPGA. The data could be inserted in the synthetization process, mapped from a graphical user interface to a memory initialization file. The primitive drawer is responsible for the effective data processing to originate the graphics, remaining to the viewport transformation maps the data to fit the video memory. The video memory writer send the data in 10 bits size for the video resolution of XGA (1024 columns by 768 lines), generating the timing synchronization signals for the video DAC. This pipeline could be used for the 2D or 3D view modifing the viewport transformation stage. Figures 4 and 5 shows examples of NURBS curve and surface obtained from the video memory system.

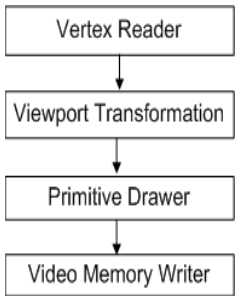


Figure 3. Basic 3D visualization pipeline.

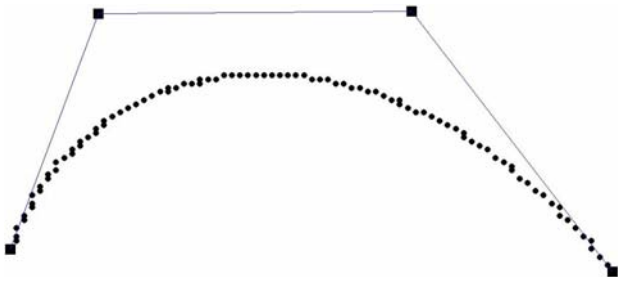


Figure 4. NURBS curve with control points (square dots).

4. FPGA Implementation

Given the set of data points to be interpolated, the initial parameters for the knots are calculated by equations (9-10), with one core for euclidean distance between the points, and the knot vector is calculated by the core of equation (11-12), as demonstred in the diagram given by figure 6, excluding the enabling gates from the bus signals, wich increments in two clock cicles each core processing (the blocks between dashed lines are parallelized). For each parameter t is given a core that recursively calculates the point C in the curve by the Cox-de Boor algorithm, with the recurrence for interpolation of the data points. In figure 7 there is an example for the generation of one point by the Cox-de Boor algorithm for a degree 3 NURBS, resulting in 3 cores layers, showing that the degree determines the number of layers,

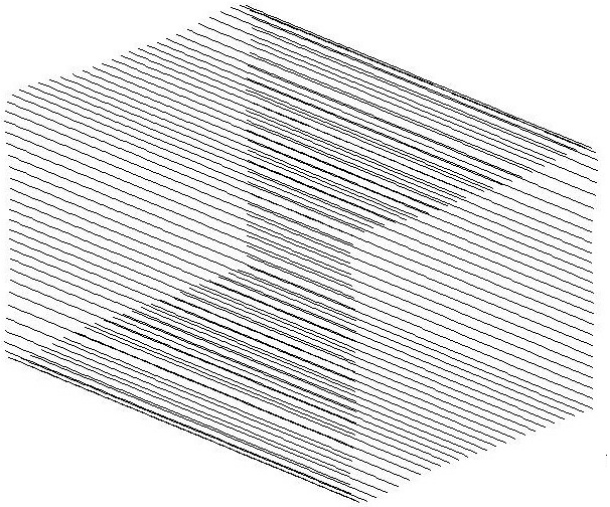


Figure 5. NURBS surface example.

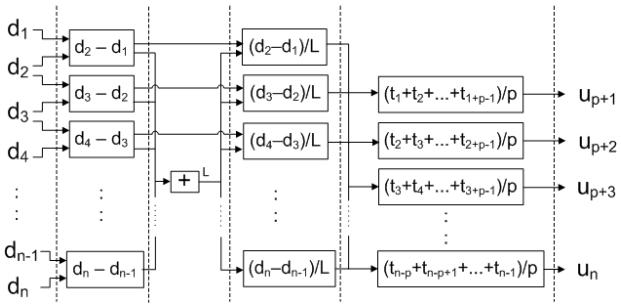


Figure 6. Chord length and averaging cores for the knot vector generation for a degree *p* curve.

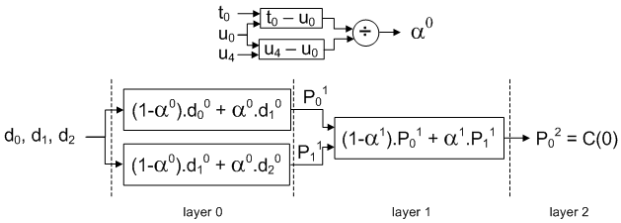


Figure 7. Core for the generation of one point in the degree 3 NURBS curve of p=3 layers.

and this core is repeated for each parameter (point) generated in the curve. If the control points and the respective knot vector are given instead of the data points to be interpolated, is used only the Cox-de Boor core to generate the NURBS curve, whithout the repetition for interpolation.
A wishbone bus is used to connection, with a round-robin arbiter to control the cores requesting. Two cores are included for straight line and circle generation, that has compact and efficient representations in relation to NURBS. For the FPGA system working as a accelerator, the data networking is made by a Nios II embedded processor core, i.e., a core acting as a general purpose microcontroller. The FPGA IDE environment provides a interface specific to custom processor design, allowing data bus width, performance and peripherals configuration, being used to set the communication with a ethernet interface. The Nios II is designed with a 8 instruction set, only for the data communication over a point to point ethernet link, in this case for the FPGA system and a computer. An C++ console application was developed and used for the data transfer between a PC computer and the FPGA system, being possible too constrain the number of parameters sent to the FPGA due to limitation in the number of logic elements.

5. General Purpose GPU

The Graphics Processing Units was created with the multicore processing capability, but the first ones was made for the computer graphics processing applications only. Actually, it has an architecture made for the multidata processing, open to the programmers and suitable for high performance computing, so that the manufacturers allows actually GPGPUs with more than 500 cores devoted to general purpose applications. The use of GPU here is based in the CUDA implementation of NURBS, that consists in a multitthread processing of the Cox-de Boor and knot addiction algorithm to the generation for each parameter in the curve/surface, despite the devoted circuits in the board. The performance is compared with the FPGA implementation, up to 16 cores, being made in CUDA-C language. The CUDA implementation follows these items:
- dividing the task in blocks, each of them consisting of a thread, defining 8 to 32 threads for each block;
- passing the serial processing to the computer processor;
- 16 bits data size (compatible with the FPGA system).
The GPU used has 2 multiprocessors with 8 CUDA cores each (enabling 16 processors). The NURBS is implemented multithreading each layer of the Cox-de Boor algorithm, just like the FPGA implementation, regarding the hardware limitation.

	CPU Vectorized code	GPU CUDA	FPGA
NURBS interpolation	47p(p+1)	16p	8p
Visualization pipeline	12p+2	4p	10p+20
Knot addiction (16 knots)	10p	16p	4p

Table 1. Total clock cycles for 32 interpolation points and 100 parameters (p is the NURBS degree).

Core	LEs
Vertex reader	57
Viewport transformation	110
NURBS interpolation	148
Straight line	24
Circle	48
Vídeo memory writer	62
Wishbone	54

Table 2. Cores and number of logic elements.

6. Results

The NURBS local interpolation algorithm and the visualization system for FPGAs, is compared with a single GPU by the number of clock cicles for each core. The GPU processing still results in a more efficient manner to deal with the data interpolation, being a dedicated circuit, designed to optimally perform graphic functions. The cores are synthesized to perform similar functionallity like the GPU, with 16 simultaneous threads and a independent clock counter synchronized with the beginning and end of the process.

7. Conclusions

The use of FPGAs in computer graphic is still incipient, being confirmed despite the fact that circuits dedicated to this aim leading to the following items:
1. while the GPU is a highly specialized processor that can get great performance (for a specific subset of the problems), actually most of them are not suitable for embedded applications in respect to FPGAs due to the power dissipation, sometimes requiring more cooling than computer processors;
2. the GPU is limited by the built-in hardware and firmware, despite of the multiprocessing power;
3. the processors based in the traditional computer architecture, are restricted by the demand of a higher clock frequency, giving arising to multicore processors. PLAs technologies didn't reach the higher frequency, being developed with higher densities too.
Actually, General Purpose GPUs (GPGPUs) provide more flexibility to the system designer, still locked to the hardware architecture, being some operations, like fixed point operations, efficiently done in FPGAs. Future works are devised to match reconfigurable systems with GPGPUS.

References

[1] I. 10303. Industrial automation systems and integration - product data representation and exchange, multipart standard. *International Organization for Standardization - ISO*.

[2] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30, 1965.

[3] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3d objects with radial basis functions. *Proc. 28th annual Conference on Computer Graphics and Interactive Techniques*, pages 67–76, 2001.

[4] B. Cope, P. Y. K. Cheung, W. Luk, and S. Witt. Have gpus made fpgas redundant in the field of video processing? *Proc. IEEE International Conference on Field-Programmable Technology*, pages 111–118, Dec. 2005.

[5] L. W. Howes, P. Price, O. Mencer, and O. Beckmann. Pgas, gpus and the ps2 - a single programming methodology. *14th Annual IEEE Symposium on Field-Programmable Custom Computing Machines*, pages 313–314, Apr. 2006.

[6] N. Knutsson. *An FPGA-based 3D Graphics System*. Master's thesis in Electronics Systems, Linkoping Institute of Technology, 2005.

[7] S. K. Lodha and R. Franke. Scattered data techniques for surfaces. *Proc. Conference on Scientific Visualization*, page 181, 1997.

[8] S. Mann, M. Lounsbery, C. Loop, D. Meyers, J. Painter, T. DeRose, and K. Sloan. A survey of parametric scattered data fitting using triangular interpolants. *Curve and surface Design - SIAM*, 1992.

[9] D. S. OpenGL Architecture Review Board, M. Woo, J. Neider, and T. Davis. *The OpenGL Programming Guide: The Official Guide to Learning OpenGL - Version 2.1*. Addison-Wesley Professional, Boston, Massachusetts, 2008.

[10] P. Orenstein. High-speed cam of 3-d sculpted surfaces. *Time Compression Magazine*, Mar. 2002.

[11] I. Organization. Initial graphics exchange specification - iges 5.3. *ANSI 1996, U. S. Product Data Association*, Sep. 1996.

[12] O. Organization. *WISHBONE System-on-Chip (SoC) Interconnection Architecture for Portable IP Cores*. Revision: B.3, 2002.

[13] L. Piegl. On nurbs: A survey. *IEEE Computer Graphics and Applications*, pages 55–71, Jan. 1991.

[14] L. Piegl and W. Tiller. *The NURBS Book*. Springer, New York, 1997.

[15] M. L. Stokes. A brief look at fpgas, gpus and cell processors. *ITEA Journal*, pages 09–11, Jun./Jul. 2007.

[16] H. Styles and W. Luk. Customising graphics applications: techniques and programming interface. *IEEE Symposium on Field-Programmable Custom Computing Machines*, pages 77–87, Apr. 2000.

[17] M. C. Tsai, C. W. Cheng, and M. Y. Cheng. A real-time nurbs surface interpolator for precision three-axis cnc machining. *International Journal of Machine Tools and Manufacture*, 43(12):1217–1227, May 2003.

[18] H. T. Yau, M. T. Lin, and M. S. Tsai. Real-time nurbs interpolation using fpga for high speed motion control. *Computer-Aided Design*, (38):1123–1133, 2006.

VII Workshop de Visão Computacional

2011

Organização:

UFPR

PUCPR

UTFPR
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

Patrocínio:

FINEP

CAPES

CNPq

opto
Science in Sight

LACTEC

Apoio:

unesp

USP

SBC
Sociedade Brasileira de Computação

SOCIEDADE BRASILEIRA DE REDES NEURAIS

omnipax
editora

BWT OPERADORA

Simply Tour
INTERNACIONAL