LUIZ ANTÔNIO PEREIRA NEVES
HUGO VIEIRA NETO
ADILSON GONZAGA

Avanços em Visão Computacional

www.omnipax.com.br

omnipax editora

Luiz Antônio Pereira Neves
Hugo Vieira Neto
Adilson Gonzaga
(Editores)

Avanços em Visão Computacional

2012

Caixa Postal: 16532 - 81520-980 Curitiba, PR

Digital Object Identifier (DOI): 10.7436/2012.avc.0

Capa:
Sérgio Alexandre Prokofiev

Projeto gráfico e editoração:
Omnipax Editora Ltda

Ficha catalográfica:
Adriano Lopes (CRB9/1429)

Dados Internacionais de Catalogação na Publicação

A946 Avanços em Visão Computacional / editores: Luiz Antônio Pereira Neves, Hugo Vieira Neto, Adilson Gonzaga. — Curitiba, PR: Omnipax, 2012
406 p. : il.

Vários autores
Inclui biografias
ISBN: 978-85-64619-09-8
eISBN: 978-85-64619-08-1

1. Visão por computador. 2. Processamento de imagens. 3. Sistemas de reconhecimento de padrões. 4. Informática na medicina. I. Neves, Luiz Antônio Pereira, ed. II. Vieira Neto, Hugo, ed. III. Gonzaga, Adilson, ed. IV. Título.

CDD (22. ed.) 006.37

Dedicatória

Meus agradecimentos aos colegas Adilson, Hugo e aos demais pesquisadores que colaboraram para tornar possível a organização deste livro, bem como à minha esposa Jane e filho Luiz Felipe pelo apoio que sempre me dedicam durante minhas realizações profissionais.
L.A.P.N.

À minha valorosa esposa Michele, por sua resiliência e companheirismo, e ao nosso precioso filho Victor Hugo, por sua doçura e sagacidade.
H.V.N.

Dedico este livro a minha esposa Suely e aos meus filhos Bruno, Tiago e Rafael.
"Eu ouço e esqueço. Eu vejo e me lembro. Eu faço e compreendo."
(Confucius)
A.G.

Prefácio

A visão computacional procura integrar as áreas de processamento digital de imagens e inteligência artificial, tendo como ojetivo a obtenção de algoritmos capazes de interpretar o conteúdo visual de imagens. Suas aplicações estão presentes em diversos segmentos tecnológicos que envolvem análise de imagens, reconhecimento de padrões e controle inteligente, abrangendo múltiplas áreas do conhecimento, tais como agronomia, astronomia, biologia, biometria, medicina e muitas outras. Constitui, portanto, uma área multidisciplinar com muitas aplicações práticas.

Os capítulos deste livro correspondem a trabalhos selecionados entre os que mais se destacaram no VII Workshop de Visão Computacional (WVC 2011), realizado na Universidade Federal do Paraná, em Curitiba, de 22 a 25/05/2011. A finalidade dos eventos WVC é servir de plataforma de divulgação de trabalhos científicos que envolvam aplicações e técnicas de visão computacional, de pesquisadores e alunos de instituições brasileiras. O WVC 2011 ofereceu um ambiente propício ao compartilhamento de novas ideias e à divulgação de pesquisas recentes, levando ao ensejo da elaboração deste livro contendo capítulos com versões ampliadas e revisadas de trabalhos previamente apresentados na forma de artigo durante o evento.

Dos 33 trabalhos inicialmente submetidos à avaliação pelos pares, somente 20 capítulos foram selecionados para compor este livro. Embora a natureza dos capítulos seja inerentemente multidisciplinar, as principais áreas de aplicação contempladas foram: análise de imagens médicas (capítulos 1 a 6), agronomia (capítulos 7 e 8), biometria (capítulos 9 e 10), processamento de vídeo (capítulos 11 a 12), reconhecimento de caracteres (capítulos 13 a 15), segmentação (capítulos 16 a 18) e visualização (capítulos 19 e 20).

Os editores agradecem a todos que contribuiram para a concretização deste livro, seja por meio da submissão de seus trabalhos ou por meio da árdua tarefa de avaliação dos capítulos submetidos. Um agradecimento especial é devido a Terumi Paula Bonfim Kamada, pelo seu competente e incansável trabalho no gerenciamento das comunicações entre autores, avaliadores e editores envolvidos na elaboração desta obra.

Apresentamos o livro "Avanços em Visão Computacional", desejando uma relevante e produtiva leitura a todos.

Luiz Antônio Pereira Neves – UFPR
Hugo Vieira Neto – UTFPR
Adilson Gonzaga – USP/SC

Sumário

Capítulo 1

Classificação Estatística e Predição da Doença de Alzheimer por meio de Imagens Médicas do Encéfalo Humano

Michel Pereira Fernandes,* João Ricardo Sato,
Geraldo Busatto Filho e Carlos Eduardo Thomaz

Resumo: O aumento da expectativa de vida populacional ocasionou um aumento na prevalência de demências degenerativas como a Doença de Alzheimer (DA). Como não há cura, o desafio é a antecipação do diagnóstico e tratamento através dos casos do Transtorno Cognitivo Leve (TCL), que tem alto grau de conversão para DA. Identificar nos casos do TCL quais converterão para DA é o principal objetivo. É proposto o uso de ferramentas de automação computacional através de modelos estatísticos para avaliar as informações extraídas das imagens médicas. Análises mostram que a metodologia é promissora com confirmação clínica de 80% e antecedência de até 4 anos.

Palavras-chave: Mapeamento Estatístico; Imagem por Ressonância Magnética; Morfometria Baseado em Voxel.

Abstract: *The increase in life expectancy has led to a population increase in the prevalence of degenerative dementias such as Alzheimer's Disease (AD). Since there is no cure, the challenge nowadays is to anticipate the diagnosis and treatment by analysis of cases of Mild Cognitive Impairment (MCI), which is known to have a high degree of conversion to AD. Thus, identifying which MCI cases will convert to AD has been the main goal. In this work, we propose the use of computational tools based on statistical models to evaluate the information extracted from medical images. Our experimental results show that the methodology is promising with clinical confirmation of 80% and, in some subjects, with up to four years in advance.*

Keywords: *Statistical Mapping; Magnetic Resonance Imaging; Voxel Based Morfometry*

*Autor para contato: michelpf@gmail.com

 DOI: 10.7436/2012.avc.1 ISBN 978-85-64619-09-8

1. Introdução

O avanço da sociedade nas áreas de saneamento, medicina, educação e alimentação, dentre outras áreas, está ocasionando um crescimento acelerado da população idosa no mundo, aumentando a expectativa de vida das pessoas.

No Brasil, em 1900, segundo dados do IBGE (Instituto Brasileiro de Geografia e Estatística), a expectativa de vida ao nascer do brasileiro era de 33,7 anos. Em 2010, passou para 73,4 anos e a projeção para 2041 é de 80,09 anos. Em função destes dados apontarem para uma população idosa em grande número, é cada vez mais claro que a área dos transtornos neuropsiquiátricos associados ao envelhecimento representa um campo de grande importância para a saúde pública nacional, especialmente sobre os custos envolvidos na assistência das doenças (Forlenza & Almeida, 2006). Neste contexto, verifica-se que dentre as doenças relacionadas a transtornos neurológicos e a idade avançada, a Doença de Alzheimer (DA) é a que mais prevalece chegando a responder pela quarta parte de todas as demências quando a pessoa acometida tem 85 anos ou mais (Forlenza & Almeida, 2006).

A DA é caracterizada pelo comprometimento de funções cognitivas, inicialmente ligadas a memória e avançando para outras áreas como atenção, linguagem e funções executivas. Por essa razão, à medida que a doença avança mais severos são os seus efeitos. Ainda que não tenham sido desenvolvidas até hoje estratégias de tratamento de grande eficácia, várias alternativas terapêuticas que postergam a evolução dos sintomas estão disponíveis. Assim, é um desafio a antecipação do diagnóstico para que o tratamento possa ser iniciado precocemente. Uma estratégia importante é a de identificação dos casos de Transtorno Cognitivo Leve (TCL), que representam um estágio de pré-demência. Como os casos de TCL apresentam alto grau de conversão para a DA, o desenvolvimento de métodos diagnósticos que permitam prever a conversão de TCL para DA é de grande relevância.

Este trabalho tem por objetivo a realização de um estudo estatístico sobre predição da DA utilizando as informações discriminantes extraídas de imagens médicas de ressonância magnética (RM) do cérebro humano. As informações extraídas das imagens correspondem ao ponto de vista anatômico e não através de avaliação clínica. Adicionalmente, objetiva-se validar na prática se o modelo adotado está correto por meio dos dados de seguimento clínico disponíveis no banco de dados utilizado. A validação do modelo é importante, pois os casos de classificação intermediária poderão ter uma nova análise quando for realizado o sistema de predição. Além disso, é possível confirmar os casos de conversão de pacientes com TCL para DA do modelo utilizado por meio das informações clínicas disponíveis.

Este capítulo está dividido em 4 seções. A próxima seção apresenta a metodologia que foi utilizada sobretudo nos processos de pré-processamento de imagens e dos modelos estatísticos univariado e multivariado além de apresentar a base de dados investigada. Na seção 3 são abordados os experimentos realizados e seus resultados. Por fim, na seção 4 é apresentada a conclusão do trabalho.

2. Metodologia

Nesta seção são apresentados os métodos investigados neste trabalho, que incluem os métodos de pré-processamento de imagens, extração das características relevantes, classificação e mapeamento estatístico das informações baseados em três metodologias, que utilizam o modelo massivamente univariado, o modelo multivariado e a projeção de hiperplanos que é responsável, principalmente, por validar o modelo de classificação utilizado. Por fim é apresentada a base de dados utilizada.

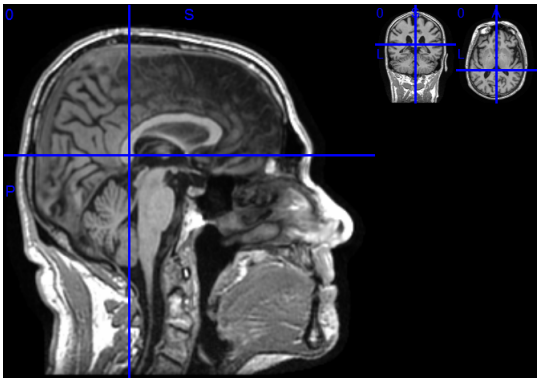


Figura 1. Imagem de RM estrutural disponibilizada pelo ADNI (sem pré-processamento).

2.1 Pré-processamento de imagens

As etapas de pré-processamento de imagens médicas adotadas são as seguintes: (i) normalização, (ii) segmentação e (iii) suavização. Estas etapas foram executadas dentro de um arcabouço computacional de análise de neuroimagens, denominado Mapeamento Estatístico Paramétrico (MEP), que torna possível a investigação de diferenças locais na anatomia do cérebro humano. Este arcabouço implementa o método de morfometria baseada em voxel (Friston et al., 2007; Good et al., 2001) que, de forma sucinta, estima a distribuição de probabilidade das intensidades dos voxels para gerar mapas estatísticos paramétricos que possam determinar diferenças relevan-

tes ou testar hipóteses sobre efeitos regionais específicos das amostras de interesse (Friston et al., 1991).

2.1.1 Normalização espacial

Na primeira etapa de pré-processamento é realizada a normalização espacial, que tem o objetivo de corrigir possíveis erros de aquisição e permitir o registro de imagens de pessoas diferentes através da estimativa de parâmetros de deformação de um espaço anatômico padronizado, como o atlas de Talairach & Tournoux (1987) e que pode aumentar o número de graus de liberdade permitido em um modelo estatístico. O processo de normalização espacial é evidenciado na Figura 2.

A normalização espacial padrão determina a transformação espacial que minimiza a soma da diferença de quadrados entre uma imagem e uma combinação linear de um ou mais modelos padronizados (Friston et al., 2007). Inicia com um registro afim para combinar com o tamanho e a posição da imagem, seguido por uma deformação global não-linear para combinar a forma total do cérebro. Utiliza um modelo Bayesiano para simultaneamente maximizar a suavização das deformações.

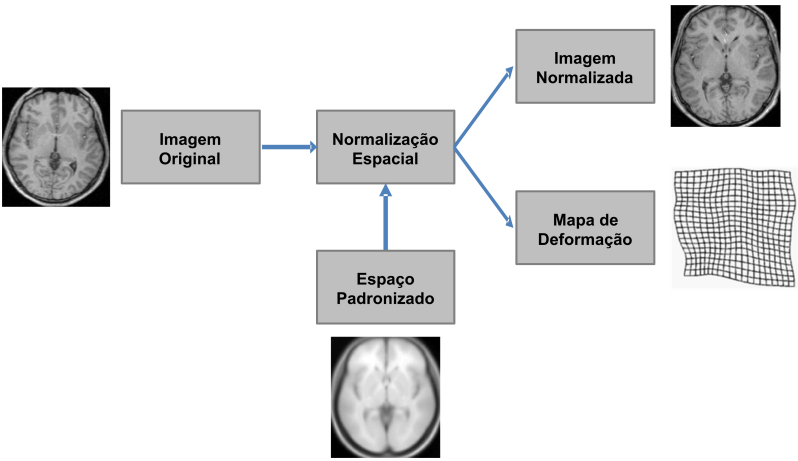


Figura 2. Processo de normalização espacial.

A estimativa de deformação, na normalização espacial, pode usar uma variedade de modelos para o mapeamento, dentre eles: (i) a transformação afim de 12 parâmetros, que constituem uma matriz de transformação espacial (translação, rotação, escala e cisalhamento) nas três direções possíveis, (ii) as funções espaciais base de baixa frequência (geralmente um conjunto cosseno ou polinômios) e (iii) um campo vetorial que especifica o mapeamento para cada ponto de controle (por exemplo, o *voxel*). A estimativa

dos parâmetros de todos estes modelos pode ser acomodada em um modelo Bayesiano simples, em que se está procurando encontrar os parâmetros de deformação θ que tem a máxima probabilidade *a posteriori*, definida por $p(\theta|y)$. A deformação é atualizada iterativamente usando o método de regressão linear Gauss-Newton para maximizar $p(\theta|y)$ utilizando o princípio de verossemelhança (Friston et al., 2007).

2.1.2 Segmentação

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 Equação 1 a seguir:

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

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. 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.

2.1.3 Suavização

Na última etapa, 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 que faz com que a intensidade de cada *voxel* da imagem seja dada pela média ponderada dos valores dos *voxels* adjacentes, determinada pelo valor de comprimento do filtro, suavizando as bordas (Good et al., 2001). Além disso, esta etapa permite a uniformização dos dados estatisticamente para que os erros associados a cada imagem sejam normalmente distribuídos, portanto tende a aumentar a eficiência do teste estatístico paramétrico (Friston et al., 1995).

2.2 Modelo estatístico massivamente univariado

O modelo estatístico univariado é baseado no Modelo Linear Geral (MLG). O MLG expressa uma variável de resposta verificada em termos de uma

combinação linear das variáveis em conjunto com um termo de erro associado (Friston et al., 2007). O MLG é também conhecido como Análise de Covariância ou Análise de Regressão Múltipla dentre outros e inclui variantes mais simples como o Teste de Hipóteses (teste t) para as diferenças entre médias.

O MLG pode ser descrito sucintamente pela Equação 2:

$$Y = X\beta + \varepsilon, \tag{2}$$

onde Y é a matriz de resultados (dados de observação), X é a matriz de variáveis que representa o desenho experimental ou matriz de delineamento, β é a matriz dos parâmetros estimados (contribuição de cada componente na matriz de delineamento) e ε é a representação do erro ou resíduo (diferença entre os dados de observação, Y, e o que foi predito pelo modelo, $X\beta$) associado à amostra utilizada, que se assume que seja independente e normalmente distribuída (Friston et al., 1995a). A forma matricial da Equação 3 considera o modelo MLG aplicado para cada n *voxel* analisado, em um estudo de imagem médica por RM, por exemplo, descrito por:

$$vetor(Y) = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 \cdots 0 \\ \vdots \ddots \vdots \\ 0 \cdots x_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \tag{3}$$

2.2.1 Teste de hipóteses

Para geração dos mapas estatísticos univariados, utiliza-se comumente o teste t de *Student* para identificar diferenças estatisticamente significantes entre os grupos de imagens pré-processadas (Leão et al., 2010). 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:

$$t_i = \frac{\overline{x}_{1,i} - \overline{x}_{2,i}}{\sigma_i \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}, \tag{4}$$

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 os números totais de amostras do grupo 1 e do grupo 2. O desvio padrão ponderado do conjunto de amostras é definido como:

$$\sigma_i = \sqrt{\frac{(N_1 - 1)(\sigma_{1,i})^2 + (N_2 - 1)(\sigma_{2,i})^2}{N_1 + N_2 - 2}}, \tag{5}$$

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

Calculados 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) em favor da hipótese alternativa (H_a), descritas como:

$$H_0 : \mu_1 = \mu_2, \tag{6}$$

$$H_a : \mu_1 \neq \mu_2. \tag{7}$$

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 (Girardi et al., 2009). Normalmente, determina-se um nível de significância α que representa na prática uma taxa de erro tipo I aceitável dado o grau de liberdade do estudo. Usualmente, o valor de α é fixado em 5%, 1% ou 0,1%. Assim, determinada arbitrariamente a probabilidade α 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 um determinado valor crítico teórico, então considera-se que a diferença encontrada no voxel i é estatisticamente relevante.

2.3 Modelo estatístico multivariado

O modelo multivariado tem o objetivo de extrair as informações discriminantes relevantes entre diferentes grupos utilizando todas as informações disponíveis em uma mesma análise. O método SVM (*Support Vector Machines*) (Vapnik, 1999) binário para a separação dos grupos analisados em duas classes tem sido o mais utilizado em classificação de imagens médicas por RM (Cuingnet et al., 2011; Sato et al., 2009; Klöppel et al., 2008; Mourão-Miranda et al., 2006). No contexto deste trabalho, as classes de interesse são DA versus controle e TCL versus controle.

2.3.1 Máquinas de vetores suporte para classificação linear

As Máquinas de Vetores Suporte ou *Support Vector Machines* (SVM) (Vapnik, 1999, 1998) são um método de classificação amplamente utilizado em bioinformática além de outras áreas de conhecimento devido a sua alta precisão, capacidade de lidar com alta-dimensionalidade e quantidade de informação e flexibilidade na modelagem de diversas fontes de dados (Schölkopf et al., 2004).

O classificador SVM também é chamado de modelo não-paramétrico, pois ao contrário da inferência estatística clássica, os parâmetros não são pré-definidos e seu número de vetores suporte depende dos dados de treinamento utilizado. Este classificador é baseado na minimização do risco estrutural, que é uma função que depende do ajuste de um modelo aos dados. Este ajuste leva em consideração a complexidade do modelo. Isto é importante para que o classificador tenha boas propriedades de generalização, ou seja, poder de predição.

Nos casos onde as classes não podem ser separadas de forma linear, o SVM transforma a entrada original em um espaço característico de alta dimensionalidade por meio de mapeamentos não-lineares ou multiquadráticos com a utilização de uma função *kernel* apropriada. Após esta etapa de transformação não-linear é encontrado o hiperplano linear ótimo de separação deste espaço em duas classes (Huang et al., 2005).

Para uma classificação linear binária, seja o seguinte conjunto de treinamento formado por N pontos, onde cada entrada $\overrightarrow{x_i}$ tem n atributos e está em uma das duas classes, ou seja, $y_i = -1$ ou $y_i = +1$. O hiperplano de separação pode ser descrito pela seguinte equação:

$$f(\overrightarrow{x}) = \overrightarrow{w} \cdot \overrightarrow{x} + b = 0, \tag{8}$$

onde $\overrightarrow{w}$ é o vetor normal ao hiperplano e $\frac{b}{\|\overrightarrow{w}\|}$ é a distância perpendicular do hiperplano até a origem, conforme a Figura 3.

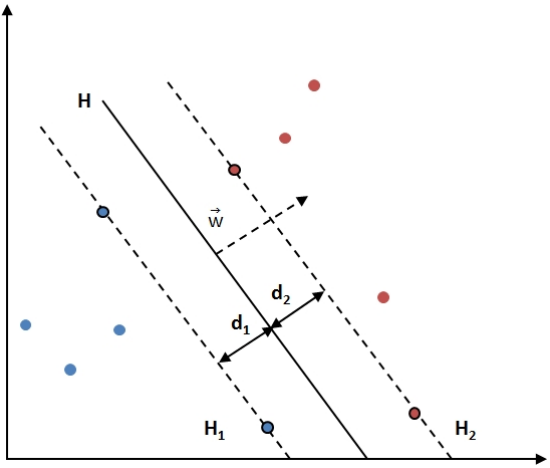


Figura 3. Representação da separação linear de hiperplano utilizando o SVM.

Os dados de treinamento podem ser discriminados em:

$$y_i = +1\text{: } \overrightarrow{x_i} \cdot \overrightarrow{w} + b \geq +1, \tag{9}$$

$$y_i = -1\text{: } \overrightarrow{x_i} \cdot \overrightarrow{w} + b \leq -1. \tag{10}$$

Considerando os pontos mais próximos da separação do hiperplano, isto é, os vetores suporte, pode-se descrever os planos como:

$$\text{para } H_1\text{: } \overrightarrow{x_i} \cdot \overrightarrow{w} + b = +1, \tag{11}$$

$$\text{para } H_2\text{: } \overrightarrow{x_i} \cdot \overrightarrow{w} + b = -1. \tag{12}$$

Para obter a separação ótima de hiperplano é necessário maximizar a margem do SVM, ou seja, manter os vetores suportes mais distantes possíveis do hiperplano. Minimizando o vetor $\overrightarrow{w}$ e aplicando uma função de otimização quadrática com solução por Lagrange, primeiro obtem-se o vetor normal ao hiperplano de separação na Equação 13, depois os vetores suporte por meio das Equações 14 e 15:

$$\overrightarrow{w} = \sum_{i=1}^{N} \alpha_i y_i \overrightarrow{x_i}, \tag{13}$$

$$\alpha_s y_s (\overrightarrow{x_s} \cdot \overrightarrow{w} + b) = 1, \tag{14}$$

$$\alpha_s y_s (\sum_{m \epsilon S} \alpha_m y_m \overrightarrow{x_m} \cdot \overrightarrow{x_s} + b) = 1, \tag{15}$$

onde os multiplicadores de Lagrange, representados por α, dada a sua maximização pelas condições de otimização de Karush-Kuhn-Tucker (KKT), são valores positivos que correspondem aos vetores suporte, porque nesta condição se encontram nos hiperplanos H_1 e H_2.

Por fim, a função de classificação é definida pela Equação 16 a seguir:

$$g(\overrightarrow{x}) = sgn(f(\overrightarrow{x})) = sgn\left(\sum_{\overrightarrow{x_i} \epsilon VetoresSuporte} \alpha_i y_i \overrightarrow{x_i} \cdot \overrightarrow{x} + b \right). \tag{16}$$

2.3.2 Projeção de hiperplano para predição de modelo classificatório

A projeção de hiperplano pode ser uma maneira versátil de verificar o comportamento de casos intermediários de classificação e apontar a direção de evolução para um determinado elemento. Este aspecto pode ser utilizado especialmente na área médica para monitorar a evolução de uma determinada enfermidade e antecipar o caminho de deterioração para começar o quanto antes a estratégia de medicação, por exemplo.

Por meio dos hiperplanos gerados pelo classificador baseado em SVM pode-se utilizar a técnica de projeção em hiperplanos para verificar o comportamento de determinados elementos que não fizeram parte do conjunto de treinamento original afim de avaliar seu comportamento devido à semelhança ou correlação com os elementos originais. A Figura 4 mostra a representação da projeção de amostra em hiperplanos previamente treinados no modelo de classificação SVM.

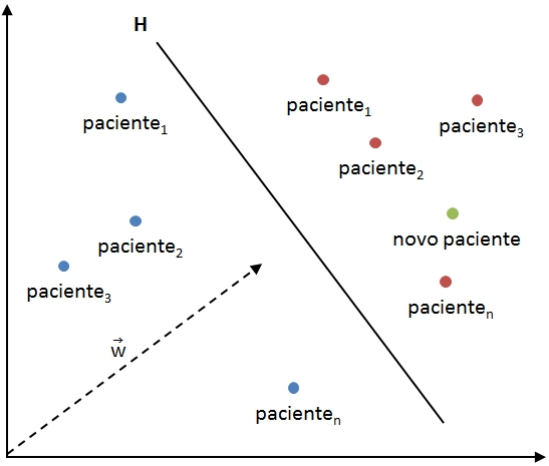


Figura 4. Representação da projeção de amostra em hiperplanos previamente treinados no modelo de classificação SVM. Adaptado de (Mourão-Miranda et al., 2006).

A projeção de um elemento externo ao treinamento no hiperplano ótimo utiliza a função da Equação 17 a seguir:

$$f(\overrightarrow{x}) = (\overrightarrow{x} \cdot \overrightarrow{w}) + b = 0. \tag{17}$$

Utilizando a técnica de validação cruzada (Weston, 1999), é possível avaliar a generalização dos resultados da classificação para verificar qual o desempenho do modelo preditivo. Este tipo de validação se baseia, em um primeiro momento, em gerar os hiperplanos de separação com amostras previamente treinadas, depois projetar uma amostra que não participou da etapa do treinamento, para, posteriormente, o classificador definir em qual lado do hiperplano a mesma será projetada. Esta característica é explorada para a validação do modelo classificatório e preditivo dos pacientes com TCL.

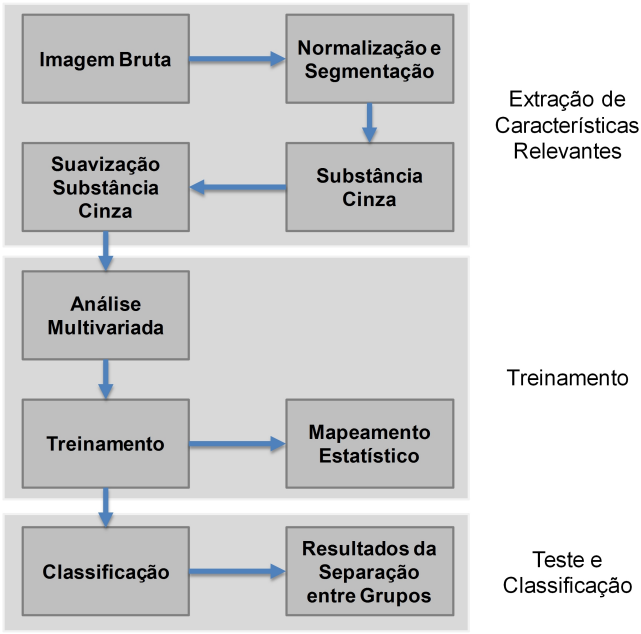


Figura 5. Fluxo das etapas envolvidas na classificação multivariada.

2.4 Classificador multivariado

A classificação multivariada foi realizada com a utilização do método SVM e é composta das seguintes etapas, ilustradas esquematicamente na Figura 5: (i) extração de características relevantes, (ii) treinamento dos grupos que se deseja classificar, (iii) teste e classificação.

A extração das características relevantes das imagens médicas é realizada pelos processos de pré-processamento. Foi priorizado a substância cinzenta devido aos efeitos das enfermidades estudadas se concentrarem, principalmente, neste tecido. As demais componentes cerebrais resultantes da segmentação, como a substância branca e o líquor são descartadas. Após a separação das imagens, cada uma delas é transformada em um vetor de informações cuja dimensão é dada pelo número de *voxels* da imagem pré-processada. Neste trabalho o tamanho da imagem tridimensional é de 91 x 109 x 91 (x, y, z). A próxima etapa consiste no treinamento e aprendizado das imagens envolvidas. Nesta etapa o classificador realizará a melhor separação linear possível entre duas classes de acordo com os vetores de informações provenientes das imagens médicas, segmentadas e suavizadas. Por fim, na fase de testes, cada amostra é avaliada de acordo com o treinamento prévio realizado e desta forma o classificador determinará sua

acurácia de classificação. A projeção corresponderá a quão distante uma classe da outra uma certa amostra estará de acordo com suas características relevantes. Assim pode-se estimar se uma dada amostra se aproxima mais de um determinado grupo ou de outro.

2.5 Base de dados internacional ADNI

Para a realização dos experimentos foi utilizada a base de dados internacional ADNI (*Alzheimer's Disease Neuroimaging Initiative*) tanto de imagens quanto de informações clínicas a respeito dos seus participantes. A base do ADNI é uma importante fonte de dados padronizada e suas imagens e demais informações são utilizadas em vários trabalhos relacionados na área em toda a comunidade científica (Spulber et al., 2010; Yakushev et al., 2009; Cuingnet et al., 2011), principalmente nas pesquisas mais recentes.

O banco de dados ADNI é uma iniciativa público-privada financiada a partir de várias fontes nos Estados Unidos da América, incluindo agências de fomento à pesquisa governamentais e não governamentais bem como diversas empresas, em grande parte do setor farmacêutico, que visa coletar o maior número de imagens por RM estrutural em boas condições e de forma padronizada para serem analisadas em pesquisas que poderão auxiliar novas descobertas para o tratamento de doenças relacionadas a DA. O acervo contempla imagens estruturais por RM, informações adicionais como, por exemplo, sexo, idade, pontuações em testes cognitivos, seguimento clínico e informações genéticas para três grupos distintos e conhecidos *a priori*: DA, TCL e controles (clinicamente sem nenhuma desordem cerebral). O banco de dados é hospedado no acervo de imagens do LONI IDA (*Laboratory of Neuro Imaging Image Data Archive*) da Universidade da Califórnia[1].

Os participantes do banco de dados foram recrutados em 50 localidades nos Estados Unidos e no Canadá, sendo 233 controles cognitivamente saudáveis em idade avançada seguidos por 4 anos, 408 indivíduos com TCL amnésico seguidos por 3 anos e 200 indivíduos com DA inicial seguidos por 2 anos. Estes números refletem a situação do banco no final de 2010, pois há atualização constante. Os critérios de inclusão no ADNI foram: indivíduos entre 55 e 90 anos de idade, sem histórico de depressão ou outra condição psiquiátrica, ausência de medicações psicoativas, pontuação 4 ou menor no teste de Hachinski modificado, 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 a DA foram pré-requisitos: estar

[1] http://www.loni.ucla.edu/ADNI

de acordo com os critérios de provável diagnóstico do *National Institute of Neurological Disorders and Stroke* e *Alzheimer's Disease and Related 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 diversos equipamentos de fabricantes diferentes, por isso durante a fase de planejamento do ADNI foram desenvolvidos protocolos específicos para cada fabricante, com o objetivo de maximizar a utilidade científica das imagens coletadas e minimizar possíveis artefatos relacionados à aquisição das imagens.

3. Experimentos e Resultados

3.1 Mapeamento visual das diferenças univariadas estatisticamente relevantes

Por meio da aplicação do SPM (Friston et al., 1995b), foi realizado o mapeamento estatístico paramétrico que consistiu em uma sobreposição de uma imagem de um cérebro de referência com o mapa gerado pelo teste de hipóteses. Esse mapa é traduzido como regiões de atrofia cerebral significantes entre os grupos de estudo.

Para melhor compreensão dos resultados e das regiões cerebrais demarcadas como coincidentes ou não, foi sobreposto o mapa de contrastes de cada análise em uma imagem de RM estrutural de um cérebro de controle, através do software MRIcro (Rorden & Brett, 2000). A Figura 6 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. Para visualização das diferenças estatísticas optou-se utilizar o nível de confiança de 95%.

No mapeamento univariado pode-se verificar as regiões do cérebro com diferenças estatísticas mais evidenciadas e concentradas em certas estruturas do que outras, como por exemplo, nas estruturas do hipocampo, lobos temporais, corpo caloso, cerebelo e sistema ventricular quando compara-se controles com DA. Na comparação entre controles e TCL, as diferenças são mais sutis, coincidindo em muitas regiões como aquelas detalhadas na DA, mas não tão localizadas. Este resultado evidencia o efeito da atrofia no tecido cerebral em maior escala de acordo com o biomarcador clínico baseado em neuroimagem utilizado atualmente (Cuingnet et al., 2011).

As informações extraídas com este mapeamento coincidem com as informações clínicas esperadas para cada tipo de enfermidade (Killiany et al., 2000; Xu et al., 2000; Bobinski et al., 1998). O método de mapeamento univariado, que analisa a imagem por RM *voxel*-a-*voxel*, constrói um mapa estatístico com várias regiões que compreendem as diferenças entre os grupos analisados. Essas regiões confirmam as estruturas cerebrais afetadas e

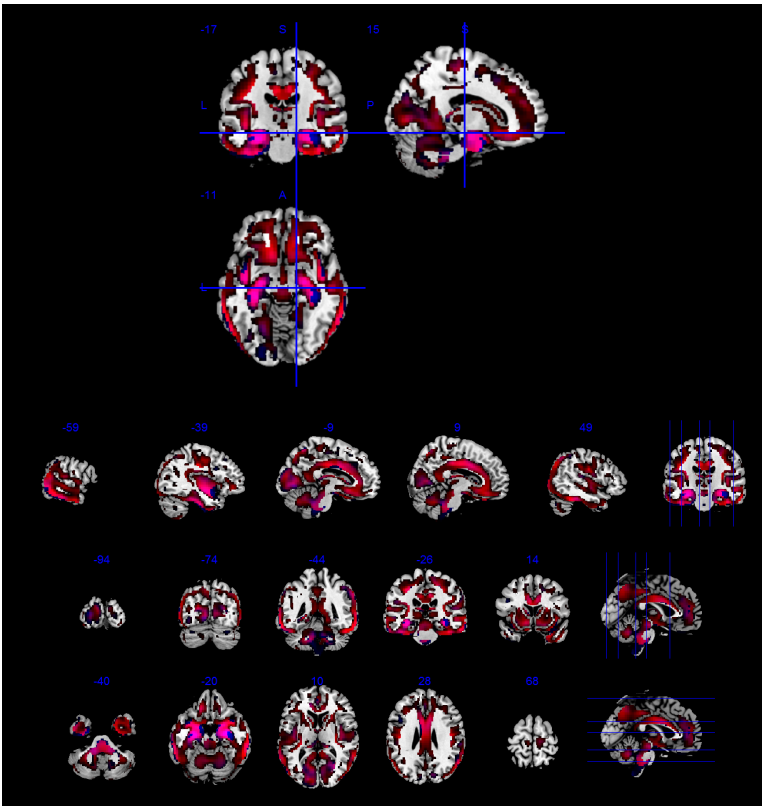


Figura 6. Mapeamento visual das diferenças univariadas estatisticamente relevantes: DA em vermelho, TCL em azul e regiões mútuas em lilás.

constituem uma ferramenta estatística importante em evidenciar as principais características relevantes entre os grupos analisados.

3.2 Mapeamento e classificação do modelo estatístico multivariado

O mapeamento e classificação estatística de imagens no modelo multivariado foram realizadas por meio do aplicativo PROBID (*Pattern Recognition of Brain Image Data*) (Marquand et al., 2010), utilizando o classificador SVM binário com validação cruzada baseada no método *Leave One Out*. As imagens foram classificadas em DA versus controle e TCL versus controle. Esta análise é baseada na classificação binária dos pacientes, que permite não somente avaliar o desempenho do classificador em separar os grupos de

análise, mas também como utilizar a informação para um modelo preditor, especialmente para os casos de TCL que poderão evoluir para a DA.

3.2.1 Classificação estatística multivariada

Para a classificação dos grupos controle versus DA foi obtido um alto nível de desempenho, com acurácia de 86,11%. A classificação do grupo controle versus TCL obteve um nível de desempenho inferior devido ao TCL se encontrar na fronteira entre o envelhecimento natural e a demência (Flint Beal et al., 2005) e obteve acurácia de 71,79%, conforme Tabela 2. A explicação de cada parâmetro de desempenho é exemplificada pela Tabela 1 e os valores consolidados dos indicadores de avaliação foram obtidos por meio das Equações 18 (Sensibilidade), 19 (Especificidade) e 20 (Acurácia), descritas a seguir:

Resultado	Doença Presente	Doença Ausente
Positivo	VP (Verdadeiro Pos.)	FP (Falso Pos.)
Negativo	FN (Falso Neg.)	VN (Verdadeiro Neg.)

Tabela 1. Quadro explicativo para o conceito de avaliação dos resultados pelo sistema classificador.

$$S = \frac{VP}{VP + FN} \tag{18}$$

$$E = \frac{VN}{FP + VN} \tag{19}$$

$$A = \frac{VP + VN}{VP + VN + FP + FN}. \tag{20}$$

Classificação	Sensibilidade	Especificidade	Acurácia
DA x Controles	87,22%	85%	86,11%
TCL x Controles	73,74%	69,83%	71,79%

Tabela 2. Quadro com os principais indicadores de desempenho do processo de classificação multivariado utilizando o SVM.

O classificador conseguiu separar os grupos DA e controles baseado somente na informação contida na imagem por RM com uma baixa taxa de erro. Apesar da imagem por RM ser um importante biomarcador para auxiliar no diagnóstico da DA, ela não é o único critério adotado (McKhann et al., 1984), e isso pode ajudar a explicar os erros de classificação encontrados. Ainda é possível verificar que a classificação não somente separa as duas classes como também pode indicar a severidade dos efeitos da DA

sob o ponto de vista do classificador. Por exemplo, pacientes classificados no grupo DA com índice de classificação mais negativo podem indicar uma progressão maior da doença enquanto os pacientes que se encontram na região de separação (índices próximos de zero) podem estar no início da DA.

Na classificação entre os grupos TCL e controles, a separação entre os dois grupos não é tão evidente quanto nos grupos DA e controles. Isso é explicado pela natureza da TCL ser caracterizada pela transição gradual entre o envelhecimento natural e a demência. Seu diagnóstico é considerado complexo sob o ponto de vista clínico, mesmo assim o classificador conseguiu separar os grupos de pacientes com taxa de acerto de aproximadamente 72%. Da mesma maneira, a severidade dos efeitos do TCL pode estar associada à sua classificação, logo os pacientes com índices de classificação mais negativos podem indicar uma evolução maior da doença. Os pacientes que se encontram na região de separação podem indicar a própria transição entre o envelhecimento natural e a demência.

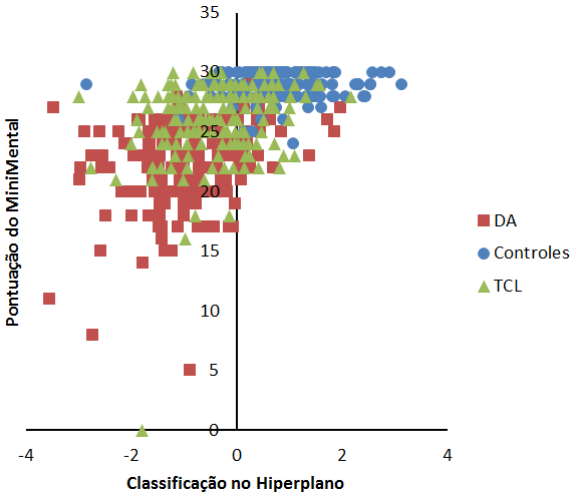


Figura 7. Gráfico de projeção dos grupos controle, DA e TCL junto ao resultado do teste de cognição Mini Mental (classificação no eixo x e resultado do Mini Mental no eixo y)

O gráfico ilustrado pela Figura 7 fornece a plausibilidade do modelo, pois mostra o cruzamento das informações do resultado do teste de cognição Mini Mental com a informação obtida pelo classificador para os três grupos envolvidos (DA, TCL e controle) na primeira visita de aquisição. Fica comprovado que, quanto menor o resultado do teste Mini Mental, mais a classificação converge para casos de DA, ou seja, há um agrupamento

dos pacientes da DA em zonas de pontuação baixa. Os pacientes com TCL ficam concentrados na zona intermediária e, por fim, os casos de controle, ficam no topo da pontuação. O Mini Mental é um importante teste de avaliação cognitivo utilizado para acompanhamento e subsídio para o diagnóstico de pacientes com TCL e DA.

3.3 Projeção de amostras e predição da classificação

A projeção de amostras tem por objetivo validar em que nível de severidade os efeitos no TCL estão e se é possível verificar sob o ponto de vista da base de treinamento dos grupos DA e controles, se o paciente está no início do TCL ou com aspecto de estabilização, podendo ser classificado pelo hiperplano como uma amostra de controle. De forma contrária, se estiver em estado de progressão, poderá ser classificado como DA ou na região de fronteira. Este experimento foi realizado com as 179 amostras de cada grupo (DA e controle) na primeira aquisição de imagem. Adicionalmente, o experimento utilizou imagens por RM em várias etapas de aquisição para realizar o acompanhamento da progressão da enfermidade e relacionar com a informação obtida exclusivamente pelo classificador. Portanto, é possível evidenciar casos de conversão em etapas diferentes de aquisição, de acordo com a progressão da enfermidade. As informações obtidas neste experimento são validadas pelos dados clínicos que estão disponíveis na base do ADNI, assim pode-se também constatar se houve confirmação dos casos de conversão ou não previstos pelo classificador desde a primeira aquisição.

O experimento de acompanhamento utilizou 30 pacientes escolhidos aleatoriamente de cada grupo (DA e TCL) que foram testados em três etapas diferentes (aquisição ou *screening*, seis meses depois e 24 meses depois). Foi constatado que nem todos os pacientes possuem o mesmo acompanhamento, ou seja, podem ter a aquisição inicial e somente após 12 meses realizam outra aquisição, de modo que tentou-se priorizar neste experimento os pacientes que possuem as três etapas de aquisição para permitir uma melhor comparação entre eles.

Para a validação da relação do parâmetro obtido na classificação com o diagnóstico clínico e também pelas características evidenciadas pela progressão foi projetada a classificação do grupo de pacientes com DA ao longo de 24 meses, com as três etapas de aquisição nos hiperplanos de classificação (no hiperplano da DA o valor de classificação é menor do que zero e para o hiperplano de controle o valor de classificação é maior do que zero). No gráfico da Figura 8 é possível constatar que, como a DA é uma doença neurodegenerativa progressiva, e mesmo com a medicação atual, não é possível a sua regressão. As projeções confirmam o progresso constante da doença e nenhum sinal de regressão.

Aplica-se então o mesmo estudo de validação para o grupo de pacientes com TCL (Figura 8), onde é apresentado a evolução da enfermidade para

este grupo de pacientes. Neste gráfico fica evidente a característica da velocidade da progressão da enfermidade ser mais lenta quando comparada com os efeitos da DA. Por esse motivo, verifica-se que as amostras dos pacientes da primeira, segunda (após seis meses) e terceira (após 24 meses) etapas de aquisição ficam próximas umas das outras, diferentemente quando compara-se a evolução da DA, onde para cada visita o aprofundamento dos efeitos da enfermidade é maior. Ainda é possível verificar pacientes com TCL que possuem uma grande probabilidade de evoluir para a DA cujo valor de classificação seja menor que zero ou na zona limítrofe, evidenciado, por exemplo, pelo paciente de número 2. O oposto também é verificado, isto é, pacientes com estabilização na zona de transição ou com valor de classificação maior do que zero, como por exemplo o paciente de número 27, onde todos os exames se encontraram na mesma região, ou seja, não houve uma evolução consistente dos efeitos da DA.

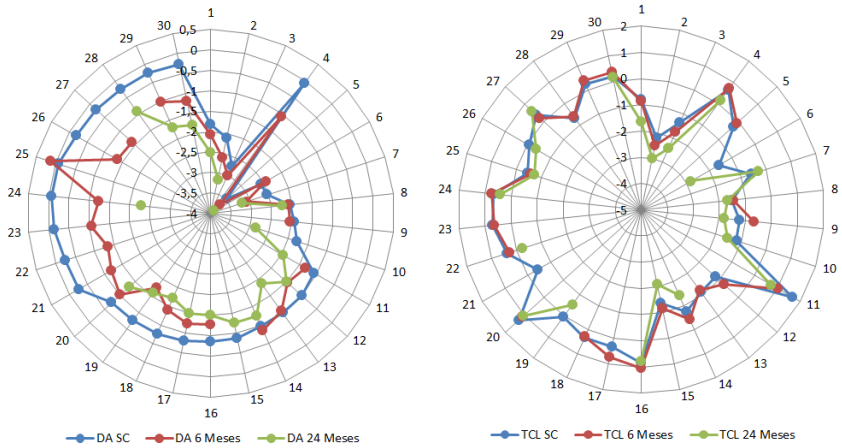


Figura 8. Gráfico radial de projeção da evolução do grupo de pacientes (DA e TCL). Valores negativos sinalizam DA e positivos controles. Cada eixo radial corresponde a um paciente distinto.

Foram confirmados pelo acompanhamento clínico fornecido pela base do ADNI 87,5% de conversão dos pacientes TCL classificados de maneira preditiva para conversão para DA, levando em consideração somente a informação da imagem médica por RM. Em termos de acurácia total, ou seja, confirmação dos pacientes que converteram ou não clinicamente, o desempenho foi de 76,7%.

Avaliando os dados de classificação mostrados no quadro da Tabela 5 é verificada uma taxa elevada de confirmação das informações de classificação, que indicam um bom desempenho do modelo preditivo. Para a

Indicador	Quantidade
Verdadeiro Positivo (VP)	14
Verdadeiro Negativo (VN)	9
Falso Positivo (FP)	5
Falso Negativo (FN)	2

Tabela 3. Quadro com dados de desempenho do modelo estatístico multivariado comparando com os resultados dos testes de cognição clínico.

avaliação da predição de conversão ou não dos pacientes com TCL foram utilizados os mesmos indicadores de desempenho utilizados anteriormente (Tabela 1). Assim, obtém-se para VP o valor de 14 das 16 amostras confirmadas. Para VN obtém-se o valor de nove das 14 amostras confirmadas. Para FP obtém-se o valor de cinco amostras e para FN tem-se o valor de duas amostras. O resumo destes resultados estão descritos na Tabela 4.

Sensibilidade	Especificidade	Acurácia
87,5%	64,3%	76,7%

Tabela 4. Quadro com dados de desempenho do modelo estatístico multivariado de conversão comparando com a confirmação clínica.

Realizando uma avaliação mais completa dos resultados e da sua validação clínica de acordo com o que foi concluído com os dados[2] da Tabela 5, verifica-se que o modelo de predição (ou seja, se um paciente com TCL converterá ou não para DA) é promissor. Das confirmações em que o classificador coincidiu com o dado clínico (14 de 16 ou 87,5% de acerto para os casos de conversão), apenas em um caso o classificador apontou conversão para DA em menos de um ano de antecedência (paciente com identificação 2). Nos demais casos o classificador já apontava conversão para DA com grande antecedência, superior ou igual a um ano, confirmada clinicamente nas etapas posteriores. O poder de predição chegou a ser de 4 anos de antecedência. Também é importante mencionar os casos de não conversão. O classificador constatou nove não conversões das 14 verificadas clinicamente. Como o objetivo principal é promover o tratamento dos pacientes convertedores cada vez mais cedo, estes resultados mostram a relevância de se utilizar tais ferramentas computacionais para avaliação de pacientes com TCL em complemento com o acompanhamento clínico.

[2] As informações ND (não definido) na Tabela 5 são devido à ausência de aquisição de imagem de um determinado paciente. A 2ª visita é após 6 meses a primeira aquisição e a 4ª visita é após 18 meses da primeira aquisição. Nem todos os pacientes possuem exames em todas as visitas.

Paciente		Predição			Conv. Clínica	Acerto
ID.	Idade	1ªVisita	2ªVisita	4ªVisita		
1	74,9	-0,79461	-0,8492	-1,6597	Sim (2 anos)	✓
2	76,67	-2,1939	-2,4957	-3,0149	Sim (0,5 ano)	✓
3	73,5	-1,3687	-1,765	-2,4474	Não	✗
4	75,3	0,63533	0,73644	0,13702	Não	✓
5	79,89	-0,26161	-0,10006	ND	Não	✗
6	86,29	-1,5859	ND	-2,8244	Sim (1,5 anos)	✓
7	82,84	-0,59505	ND	-0,30472	Sim (4 anos)	✓
8	81,77	-1,5307	-1,4724	-1,7035	Sim (3 anos)	✓
9	56,81	-1,2188	-0,66401	-1,8249	Sim (1,5 anos)	✓
10	87,78	-1,1763	ND	-1,5435	Sim (1,5 anos)	✓
11	82,02	1,6732	1,0088	0,73405	Não	✓
12	62,19	-1,1742	-0,76552	ND	Sim (2 anos)	✓
13	77,07	-1,1143	-1,2034	ND	Sim (1,5 anos)	✓
14	61,66	-0,72683	-0,42392	-1,4065	Não	✗
15	70,52	-1,3511	-1,1541	-2,1038	Sim (1 ano)	✓
16	79,75	0,84755	1,0393	0,76782	Sim (0,5 ano)	✗
17	75,66	0,32715	0,75582	ND	Não	✓
18	65,16	0,33904	0,30631	ND	Não	✓
19	78,51	0,03779	ND	-0,52664	Sim (3 anos)	✓
20	75,03	1,2826	1,1003	1,0685	Não	✓
21	79,94	-0,43008	ND	ND	Sim (1 ano)	✓
22	64,71	0,38254	0,28761	-0,2342	Sim (1,5 anos)	✓
23	71,02	0,72424	0,65947	ND	Não	✓
24	75,98	0,68095	0,73789	0,42262	Não	✓
25	71,1	-0,42939	-0,57593	-0,71368	Sim (3 anos)	✓
26	74,16	-0,056789	ND	-0,37667	Não	✗
27	72,47	0,38824	0,22058	0,61694	Não	✓
28	68,02	-0,67679	-0,6302	ND	Não	✗
29	76,64	0,22907	0,39083	ND	Sim (inicial)	✗
30	81,81	0,19246	0,3651	0,15992	Não	✓

Tabela 5. Tabela com a classificação baseada em projeção dos casos de TCL no hiperplano de DA versus controles, comparando os casos de conversão clínica para DA confirmados pelo ADNI.

4. Conclusão

Este trabalho utilizou o arcabouço computacional de análise de imagens médicas por RM que torna possível a investigação de diferenças locais na anatomia do cérebro humano. Este arcabouço implementa o método de morfometria baseado em *voxel* que, de forma sucinta, estima a distribuição de probabilidade das intensidade dos *voxels* para gerar mapas estatísticos paramétricos que possam determinar diferenças relevantes que permitem, antes de realizar qualquer tipo de análise, extrair as informações mais importantes. A extração dessas informações constitui a etapa

de pré-processamento, que tem por objetivo normalizar as imagens com a finalidade de mantê-las em um mesmo espaço anatômico para permitir a comparação de pessoas diferentes em uma mesma análise. Após isso, são extraídas as regiões de interesse do cérebro por meio da segmentação de imagens. A última etapa do pré-processamento, a suavização, tem o objetivo de suavizar as bordas das imagens, processo que auxilia a eficiência dos testes estatísticos.

Neste artigo, por meio de mapeamentos estatísticos, foi possível revelar quais as diferenças estatisticamente relevantes entre as enfermidades DA e TCL nas estruturas cerebrais humanas. As diferenças encontradas podem complementar os estudos atuais relacionados com ambas as doenças e os efeitos que afetam cada uma das enfermidades, sobretudo no TCL, onde a possibilidade de antecipar a evolução para outros tipos de demência permite um tratamento clínico mais personalizado. Nota-se, especialmente, a relevância da região do hipocampo, comum nas duas enfermidades, que estabelece efeitos semelhantes nos pacientes, porém em diferentes tipos de severidade.

Com o classificador multivariado foi possível, além de verificar como o método classifica os grupos analisados, também constatar o efeito preditor do modelo estatístico na prática. Na classificação de grupos, o desempenho de acurácia para separação dos grupos DA e controles obteve pouco mais de 86% de acurácia, nos grupos TCL e controle o resultado foi um pouco inferior, chegando a mais de 71% de acurácia. Estes resultados comprovam que o SVM conseguiu uma boa margem de classificação para todos os grupos analisados. Outros trabalhos na área obtiveram valores de desempenho semelhantes por meio desta e de outras técnicas de classificação quando se compararam os mesmos grupos envolvidos neste trabalho (Cuingnet et al., 2011).

Ao treinar uma base de dados com um número significativo de amostras de pacientes com DA e controles, a avaliação de acompanhamento de pacientes com TCL durante 24 meses em três etapas de aquisição tornou-se um método promissor de predição dos efeitos da DA e posterior possibilidade de conversão. O modelo foi devidamente validado por meio das confirmações clínicas dos pacientes com conversão para DA, por meio do acompanhamento clínico que o banco de dados ADNI proporciona. O alto grau de confirmações verificadas, superior a 80%, e a grande capacidade de predição, confirmada pela previsão de conversões de pelo menos seis meses e até quatro anos de antecedência, podem indicar um caminho promissor do estudo da conversão de pacientes com TCL para DA.

Agradecimentos

Os autores deste trabalho gostariam de agradecer o apoio da FAPESP (2010/01394-4). Os dados utilizados na elaboração deste estudo foram

obtidos do banco de dados ADNI. Como tal, os pesquisadores do ADNI forneceram os dados, mas não participaram da análise e elaboração deste trabalho. Para pré-processamento e classificação das amostras, foram utilizados respectivamente os pacotes SPM e PROBID de computação em imagens médicas e domínio público.

Referências

Bobinski, M.; de Leon, M.J.; Tarnawski, M.; Wegiel, J.; Reisberg, B.; Miller, D.C. & Wisniewski, H.M., Neuronal and volume loss in CA1 of the hippocampal formation uniquely predicts duration and severity of Alzheimer's disease. *Brain Research*, 14(1-2):267–269, 1998.

Cuingnet, R.; Gerardin, E.; Tessieras, J.; Auzias, G.; Lehéricy, S.; Habert, M.; Chupin, M.; Benali, H. & Colliot, O., Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766–781, 2011.

Flint Beal, M.; Lang, A.E. & Ludolph, A.C., *Neurodegenerative Diseases: neurobiology, pathogenesis and therapeutics*. Cambridge, UK: Cambrige University Press, 2005.

Forlenza, O.V. & Almeida, J.R., *Envelhecimento e transtornos psiquiátricos*. São Paulo, SP: Atheneu, 2006.

Friston, K.J.; Ashburner, J.; Frith, C.D.; Poline, J.B.; Heather, J.D. & Frackowiak, R.S.J., Spatial registration and normalization of images. *Human Brain Mapping*, 3(3):165–189, 1995.

Friston, K.J.; Ashburner, J.T.; Kiebel, S.J.; Nichols, T.E. & Penny, W.D. (Eds.), *Statiscal Parametric Mapping: The Analysis of Functional Brain Images*. London, UK: Academic Press, 2007.

Friston, K.J.; Frith, C.D.; Liddle, P.F. & Frackowiak, R.S.J., Comparing functional (PET) images: the assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism*, 11(4):690–699, 1991.

Friston, K.J.; Holmes, A.P.; Poline, J.B.; Grasby, P.J.; Williams, S.C.; Frackowiak, R.S.J. & Turner, R., Analysis of fMRI time-series revisited. *NeuroImage*, 2(1):45–53, 1995b.

Friston, K.J.; Holmes, A.P.; Worsley, K.J.; Poline, J.P.; Frith, C.D. & Frackowiak, R.S.J., Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210, 1995a.

Girardi, L.H.; Cargnelutti Filho, A. & Storck, L., Erro tipo I e poder de cinco testes de comparação múltipla de médias. *Revista Brasileira de Biometria*, 27(1):23–36, 2009.

Good, C.D.; Johnsrude, I.S.; Ashburner, J.; Henson, R.N.; Friston, K.J. & Frackowiak, R.S., A voxel-based morphometry study of ageing in 465 normal adult human brains. *NeuroImage*, 14(1):21–36, 2001.

Huang, T.; Kecman, V. & Kopriva, I., *Kernel Based Algorithms for Mining Huge Data Sets*. v. 17 de *Studies in Computational Intelligence*. Heidelberg, Germany: Springer-Verlag, 2005.

Killiany, R.J.; Gomez-Isla, T.; Moss, M.; Kikinis, R.; Sandor, T.; Jolesz, F.; Tanzi, R.; Jones, K.; Hyman, B.T. & Albert, M.S., Use of structural magnetic imaging to predict who will get Alzheimer's disease. *Annals of Neurology*, 47(4):430–439, 2000.

Klöppel, S.; Stonnington, C.M.; Chu, C.; Draganski, B.; Scahill, R.I.; Rohrer, J.D.; Fox, N.C.; Jack Jr., C.R.; Ashburner, J. & Frackowiak, R.S.J., Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(3):681–689, 2008.

Leão, R.D.; Sato, J.R. & Thomaz, C.E., Extração multilinear de informações discriminantes em imagens de ressonância magnética do cérebro humano. Anais do XXX Congresso da SBC – XX Workshop em Informática Médica, :1740–1749, 2010.

Marquand, A.; Rondina, J.; Mourão-Miranda, J.; Rocha-Rego, V. & Giampietro, V., *Pattern Recognition of Brain Image Data* (PROBID). Technical Report, King's College London, London, UK, 2010. www.brainmap.co.uk/Programs/PROBID_user_guide_v1.03.pdf.

McKhann, G.; Drachman, D.; Folstein, M.; Katzman R., P.D. & Stadlan, E., Clinical diagnosis of alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology*, 34(7):939–944, 1984.

Mourão-Miranda, J.; Reynaud, E.; McGlone, F.; Calvert, G. & Brammer, M., The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. *NeuroImage*, 33(4):1055–1065, 2006.

Rorden, C. & Brett, M., Stereotaxic display of brain lesions. *Behavioural Neurology*, 12(4):191–200, 2000.

Sato, J.R.; Fujita, A.; Thomaz, C.E.; Morais-Martin, M.G.; Mourão-Miranda, J.; Brammer, M.J. & Junior, E.A., Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. *NeuroImage*, 46(1):105–114, 2009.

Schölkopf, B.; Tsuda, K. & Vert, J.P. (Eds.), *Kernel Methods in Computational Biology*. Cambridge, USA: MIT Press, 2004.

Spulber, G.; Niskanen, E.; MacDonald, S.; Smilovici, O.; Chen, K.; Reiman, E.M.; Jauhiainen, A.M.; Hallikainen, M.; Tervo, S.; Wahlund, L.O.; Vanninen, R.; Kivipelto, M. & Soininen, H., Whole brain atrophy rate predicts progression from MCI to Alzheimer's disease. *Neurobiology Aging*, 31(9):1601–1605, 2010.

Talairach, J. & Tournoux, P., *Co-planar stereotactic atlas of the human brain*. Stuttgart, Germany: Thieme Medical Publishers, 1987.

Vapnik, V.N., *Statistical Learning Theory*. New York, USA: J. Wiley & Sons, Inc., 1998.

Vapnik, V.N., *The nature of statistical learning theory*. 2a edição. New York, USA: Springer-Verlag, 1999.

Weston, J., Leave-one-out support vector machines. Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, :727–731, 1999.

Xu, Y.C.; Jack Jr., C.R.; O'Brien, P.C.; Kokmen, E.; Smith, G.E.; Ivnik, R.J.; Boeve, B.F.; Tangalos, R.G. & Petersen, R.C., Usefulness of MRI measeures of entorhinal cortex vs. hippocampus in AD. *Neurology*, (9):1760–1767, 2000.

Yakushev, I.; Hammers, A.; Fellgiebel, A.; Schmidtmann, I.; Scheurich, A.; Buchholz, H.; Peters, J.; Bartenstein, P.; Lieb, K. & Schreckenberger, M., SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnestic mild cognitive impairment from healthy aging. *NeuroImage*, 44(1):43–50, 2009.

Notas Biográficas

Michel Pereira Fernandes tem graduação e mestrado em Engenharia Elétrica (FEI, 2006 e 2011, respectivamente). Atualmente é professor da Universidade Paulista (UNIP), da Faculdade de Informática e Administração Paulista (FIAP) e consultor de projetos de sistemas voltados a inovação na Telefônica Vivo, atuando principalmente nos seguintes temas: sistemas de classificação de imagens médicas e mineração de dados, além de sistemas de acessibilidade a deficientes visuais.

João Ricardo Sato tem graduação em Estatística, mestrado e doutorado (Universidade de São Paulo, 2002, 2004 e 2007, respectivamente). É Coordenador e Professor do Núcleo de Cognição e Sistemas Complexos na UFABC. Atua principalmente em projetos de pesquisa multidisciplinares envolvendo os seguintes temas: modelagem estatística e computacional em neurociências, neuroimagem, mapeamento funcional do cérebro humano, análise de séries temporais, bioestatística, estatística não-paramétrica e modelos de regressão.

Geraldo Busatto Filho tem graduação em Medicina (Universidade de São Paulo, 1987), residência médica (Instituto de Psiquiatria do Hospital das Clínicas da USP – HC-FMUSP, 1990), doutorado em Psiquiatria pelo (Institute of Psychiatry, University of London, 1995), e pós-doutorado (University of London, 1996). Atualmente é Professor Associado do departamento de Psiquiatria junto à FMUSP, sendo o coordenador do Laboratório de Investigação Médica 21 (LIM21) do HC-FMUSP, cujo tema central de investigações são estudos de neuroimagem aplicados aos transtornos neuro-psiquiátricos. Suas linhas de pesquisa principais envolvem o uso de métodos de ressonância magnética, PET e SPECT para a investigação de esquizofrenia e outros transtornos psicóticos, demências e transtornos do humor, associando os dados de neuroimagem a outros biomarcadores.

Carlos Eduardo Thomaz tem graduação em Engenharia Eletrônica e mestrado em Engenharia Elétrica (Pontifícia Universidade Católica do Rio de Janeiro, 1993 e 1999, respectivamente), doutorado e pós-doutorado em Ciência da Computação (Imperial College London, 2005). Atualmente é professor adjunto do Centro Universitário da FEI. Tem experiência na área de Ciência da Computação, com ênfase em reconhecimento de padrões em estatística, atuando principalmente nos seguintes temas: Visão Computacional, Computação em Imagens Médicas e Biometria.

Capítulo 2

Análise e Caracterização de Lesões de Pele para Auxílio ao Diagnóstico Médico

Alex Fernando de Araujo [*], João Manuel R. S. Tavares, Roberta Barbosa Oliveira, Ricardo Baccaro Rossetti, Norian Marranghello, Aledir Silveira Pereira

Resumo: Neste capítulo, propõe-se uma metodologia híbrida para detectar e extrair os contornos de lesões de pele a partir de imagens, bem como a definição de características usualmente utilizadas no diagnóstico de lesões. O método de segmentação por divisão e união (*Split and Merge*) foi adotado para detectar a lesão e obter o seu contorno inicial. Em seguida, este contorno é refinado pelo modelo de contorno ativo tradicional. Características da lesão usadas na regra ABCD são definidas a partir do contorno refinado. Os resultados experimentais indicam que o método proposto é promissor para detectar as áreas com lesão e extrair seus contornos a partir de imagens, mantendo suas características.

Palavras-chave: Segmentação de Imagens Médicas, Lesões de Pele, Crescimento de Regiões, Contornos Ativos.

Abstract: *A hybrid methodology for detecting and extracting skin lesion contours from images as well as the definition of common lesion diagnosis features are presented in this paper. The split and merge segmentation method has been applied to lesion detection and to the extraction of its initial contour. This contour is then adjusted using the traditional active contour model. The final contour characteristic features are defined according to the ABCD rule. Experimental results show that the proposed method is promising in detecting ill areas as well as extracting their contours from images, while keeping lesions' features.*

Keywords: *Medical Image Segmentation, Skin Lesions, Region Growing, Active Contour.*

[*]Autor para contato: fa.alex@gmail.com

Neves et al. (Eds.), Avanços em Visão Computacional (2012) DOI: 10.7436/2012.avc.2 ISBN 978-85-64619-09-8

1. Introdução

Metodologias computacionais para processamento e análise de imagens 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 graves, tal como o câncer de pele. Geralmente, o diagnóstico inicial destas lesões é feito a partir da análise de imagens obtidas através do exame de dermatoscopia ou de fotografias tiradas usando câmeras digitais convencionais. Para obter o diagnóstico inicial destas lesões, os dermatologistas analisam visualmente as imagens das regiões suspeitas. No entanto, alguns fatores, como o cansaço visual, a pequena dimensão de algumas 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 áreas de lesões. Os elementos que podem auxiliar no diagnóstico médico são a rugosidade e irregularidade das bordas das áreas lesionadas, a sua assimetria, diâmetro e variação de cor. 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 pelo desenvolvimento de manchas ou lesões, que são afetadas por fatores externos, como a exposição excessiva ao sol (Guide, 2012). Assim, técnicas de processamento e análise computacional de imagem 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 (Ma et al., 2010). Com o objetivo de detectar e extrair os contornos de lesões de pele a partir de imagens, mantendo sua integridade de detalhes, propõe-se desenvolver um método para processamento e análise das imagens de pele. Para atingir este objetivo aplicou-se o método de crescimento de regiões, seguido de uma etapa de pós-processamento, onde este contorno inicial é melhor ajustado à lesão pelo modelo de contorno ativo tradicional.

O crescimento de regiões foi implementado utilizando o algoritmo de "divisão e união"(*split and merge*), 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. Após a divisão é realizada a

união baseada 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 aplicado deformando este contorno inicial, de maneira a obter o novo contorno, sendo o mais próximo do real possível.

Neste capítulo, uma proposta de caracterização das lesões a partir das bordas obtidas das imagens é proposta. No método desenvolvido, para definir a rugosidade, a assinatura das bordas é extraída, analisando-se o tamanho e a quantidade de ispículos presentes na mesma. O diâmetro e a assimetria das áreas lesionadas são calculados a partir da distância entre os pixels do contorno, e a partir da localização do tecido lesionado obtem-se a variação de sua coloração.

Na sequência, tem-se na Seção 2 o panorama geral da área em pesquisa. Na Seção 3, descreve-se o método proposto. Uma discussão sobre os testes realizados e os resultados experimentais obtidos são incluídos na Seção 4, seguido das conclusões e sugestões para trabalhos futuros.

2. Enquadramento

Para efetuar o processamento e análise de imagens de lesões é importante conhecer o problema a ser tratado, bem como as metodologias a serem utilizadas na resolução do mesmo.

2.1 Diagnóstico de lesões de pele

O alto índice de lesões de pele adquiridas de várias formas, e as consequências que elas podem trazer para o paciente, podendo vir a ser um câncer de pele, torna a sua detecção precoce muito importante para permitir a definição do plano de tratamento mais adequado. As lesões de pele provocadas por células cancerosas podem levar vários anos para se manifestar. No entanto, após se manifestarem, algumas lesões cancerosas podem crescer lentamente enquanto outras (os melanomas, por exemplo) podem crescer e se espalhar rapidamente pelo corpo (Guide, 2012). A demora na manifestação das lesões decorre do fato de que as cancerígenas geralmente se desenvolvem a partir de uma célula doente, a qual sofreu alguma mutação desordenada, provocando danos no seu DNA.

Para o diagonóstico das lesões de pele algumas características visuais como assimetria, irregularidade da borda, variação da cor interna e o diâmetro são observadas. Estas características são conhecidas como ABCD das lesões de pele (ou simplemente regra do ABCD). Na Tabela 1 apresenta-se o guia ABCD.

O fato dos melanomas crescerem e se modificarem rapidamente tem levado a uma reavaliação do guia ABCD, onde pretende-se adicionar a

característica "evolução", transformando o guia em ABCDE (Barcelos & Pires, 2009). Uma pesquisa desenvolvida por Morris-Smith (1996) revelou que os profissionais da área médica enfatizam muito estas características quando analisam as imagens de lesões, principalmente a irregularidade do contorno (Jia-Xin & Sen, 2005). Este guia permite ao dermatologista fazer uma análise prévia da lesão, definindo os procedimentos a serem seguidos até que exames mais detalhados sejam feitos.

Tabela 1. Guia ABCD das lesões de pele (Adaptado de Barcelos et al. (2003)).

Característica	Descrição
Assimetria	**Simétrica:** lesões que geralmente são não cancerosas e/ou não malignas, sendo mais arredondadas. **Assimétrica:** lesões que possuem grande probabilidade de serem cancerosas e malignas.
Borda	**Lisa:** lesões não malignas possuem bordas lisas e suaves. **Rugosa ou cortada:** frequentemente lesões malignas possuem bordas irregulares.
Cor	**Regular:** lesões com cor interna homogênea que, na maioria das vezes, são não malignas. **Irregular:** geralmente os melanomas possuem coloração interna com grande variação das intensidades marrom e preta, podendo ter regiões isoladas em branco.
Diâmetro	$<$ **6mm:** lesões não malignas geralmente possuem diâmetro inferior a 6 mm e não variam de tamanho. $\geq$ **6mm:** frequentemente os melanomas são maiores do que 6 mm de diâmetro, crescem e mudam de forma rapidamente.

As imagens de lesões de pele são obtidas, de forma não invasiva, a partir de máquinas fotográficas ou, mais detalhadamente por meio de um exame de dermatoscopia. Neste segundo caso, um instrumento óptico chamado dermatoscópio é usado para fazer a captura da imagem. Este equipamento possui lentes de aumento associadas a um sistema de iluminação, permitindo analisar as camadas mais profundas da pele. A principal vantagem de fazer a captura das imagens por dermatoscopia é o diagnóstico das lesões em seus estágios iniciais. No entanto, um fator negativo relacionado a este exame é que a maioria dos pequenos municípios não dispõem de tal tecnologia. Tal fato aumenta a importância de técnicas que permitam uma análise inicial das lesões a partir de imagens obtidas por máquinas fotográficas comuns.

Através da visualização das imagens de pele, o dermatologista pode analisar a lesão, e suas características, a fim de ter uma idéia sobre a sua classificação. No entanto, esta análise não é tarefa fácil, devido a fatores como o cansaço visual do dermatologista, e a interferência causada por pêlos ou bolhas sobre a região lesionada. Uma forma de minimizar estes problemas é extrair o contorno destas regiões. Assim, a irregularidade das lesões fica mais evidente, evitando que a principal característica do guia ABCD seja distorcida em caso de cansaço visual.

2.2 Segmentação de imagens médicas

Geralmente, as técnicas de segmentação de imagens médicas são usadas para 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 análise 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, modelos de contornos ativos e métodos híbridos (Ma et al., 2010).

Limiarização

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 (Gonzalez et al., 2003).

Crescimento de região

Na tentativa de unir o maior número possível de pixels em regiões homogêneas, têm sido propostos métodos baseados em crescimento de regiões (Lee et al., 2005; Celebi et al., 2008). Uma abordagem frequentemente usada é dividir a imagem de entrada em conjuntos de regiões disjuntas, tal como realizado pelo método de divisão e união. 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 quatro nós filhos, os quais podem ser unidos de acordo com sua similaridade (Gonzalez et al., 2003). Diversas outras abordagens para crescimento de regiões têm sido propostas.

Algoritmos genéticos

Algoritmos genéticos (AG) têm sido usados na segmentação de imagens com características variadas (Hashemi et al., 2010; Mukhopadhyay & Maulik, 2011). 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 de mutação atua na manutenção da diversidade genética da população, evitando a estagnação na evolução dos indivíduos, que pode gerar resultados falsos-positivos.

Redes neurais artificiais

As redes neurais artificiais (RNA) são sistemas inteligentes, cuja meta é interpretar e resolver problemas computacionais baseando seu funcionamento no do cérebro humano (Babini & Marranghello, 2007). Elas são capazes de adquirir conhecimento utilizando um grande conjunto de unidades de processamento, tratados por neurônios artificiais, interligados entre si, formando o que se conhece por sinapses artificiais. Essa estrutura confere às RNA uma capacidade de processamento paralelo e distribuído, garantindo-lhes características interessantes, do ponto de vista computacional, tais como: adaptabilidade, capacidade de aprendizado, habilidade de generalização, tolerância a falhas, possibilidade de armazenamento distribuído e facilidade de prototipagem (da Silva et al., 2010). Técnicas baseadas em RNA têm sido muito utilizadas para o reconhecimento de padrões (Bishop, 2004), em particular, dentro do escopo deste capítulo, para o processamento e análise de imagens médicas (Hudson & Cohen, 2000).

Modelos de contornos ativos

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 de Kass et al. (1988) 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 na análise de imagens médicas (Yoon et al., 2008; Lu & Shen, 2006).

Métodos híbridos

Existe também uma tendência para unir diferentes técnicas (Jianli & Baoqi, 2009; Dokur & Ölmez, 2002; Araujo et al., 2011). Estas metodologias híbridas têm ganhado atenção especial devido à sua habilidade para produzir resultados mais precisos, além de processar imagens mais complexas. Por meio da combinação das características de duas ou mais técnicas, estas metodologias são usadas para superar algumas dificuldades da segmentação, tais como a heterogeneidade das regiões, as interferências de ruídos, as variações de posicionamento ou oclusões.

3. Método proposto

Nas imagens de lesões de pele existe a interferência causada por ruídos e outros artefatos, como pêlos, bolhas e interferências da iluminação. Além disto, as lesões de pele possuem formas, dimensões e tons variados. Assim, realizar a caracterização visual destas lesões é uma tarefa complexa, o que torna a extração eficiente das bordas de lesões crucial para facilitar o diagnóstico pelo dermatologista. Aqui é proposto um método híbrido para extrair contornos de lesões de pele a partir de imagens convencionais, usando as técnicas de crescimento de regiões e contornos ativos. Optou-se por adotar as imagens obtidas por câmeras de imagem convencionais devido ao fato de uma boa parte dos exames iniciais ainda serem feitos a partir destas imagens, por falta de acesso ao dermatoscópio.

Na Figura 1, apresenta-se o diagrama de fluxo do método proposto para realizar as tarefas de processamento, segmentação e caracterização das lesões a partir de imagens de pele. O método é baseado nos seguintes passos: pré-processamento, processamento, pós-processamento e caracterização.

As imagens de lesões de pele são geralmente coloridas e variadas, com diferentes formas e intensidades de ruído, como pêlos, bolhas e interferência da iluminação, dificultando sua caracterização visual. Desta forma, um filtro de redução de ruídos é aplicado para pré-processar a imagem original e tentar reduzir a interferência destes artefatos na etapa de segmentação. Nesta etapa, é desejável que as interferências sejam removidas e ao mesmo tempo as regiões dos contornos dos objetos mantidas. Diante disto, o método de suavização proposto por Barcelos et al. (2003) é aplicado por apresentar bons resultados na remoção de ruídos e na manutenção das bordas em imagens de pele, como pode ser verificado em Barcelos & Pires (2009).

A extração das bordas das lesões de forma eficiente é importante para facilitar a análise da rugosidade do contorno das regiões lesionadas pelo dermatologista. Várias técnicas de segmentação podem ser aplicadas para realizar esta tarefa. No entanto, para obter bordas mais precisas, as tarefas de extrair os contornos é dividida em duas etapas: detecção do contorno da lesão e refinamento deste contorno. Em cada etapa foi aplicado um método de segmentação, na tentativa de obter contornos que mantivessem a integridade de suas características.

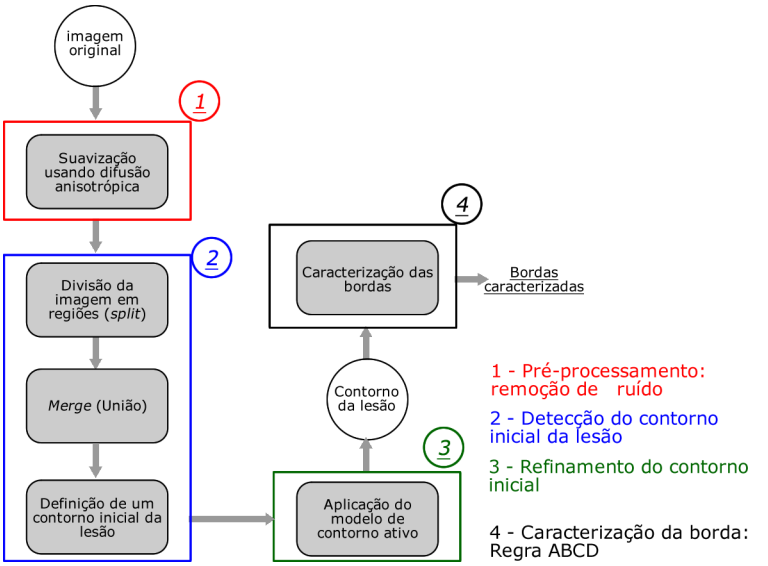


Figura 1. Diagrama de fluxo do método proposto.

Após a redução das interferências causadas pela presença de ruídos, o segundo passo do método é aplicado na imagem para detectar a área lesionada e extrair um contorno inicial para esta. Partindo-se do conhecimento a priori sobre as lesões de pele analisadas, nota-se que geralmente as áreas lesionadas apresentam intensidades de cores mais escuras do que as regiões não lesionadas. Assim, o método de segmentação baseado em divisão e união foi adotado para esta etapa, por permitir dividir a imagem em regiões distintas de acordo com a similaridade de seus pixels. 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. 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 seguinte permite que as imagens suavizadas sejam representadas por um conjunto de regiões homogêneas, facilitando a detecção das regiões de pele que são potenciais áreas lesionadas. Para unir as várias regiões segmentadas e isolar o fundo da imagem, um algoritmo de união é aplicado considerando a distância entre as intensidades das regiões e agrupando as regiões com intensidades similares. O resultado é uma imagem binarizada, a qual torna possível a extração dos contornos aproximados para as lesões pela avaliação das intensidades dos seus pixels.

Na maioria das imagens testadas, os contornos aproximados obtidos pelo método de crescimento de regiões foram coincidentes com a topologia das bordas das lesões. No entanto, apesar destes contornos possuírem topologia semelhante à topologia das lesões, eles não envolveram toda área doente das mesmas, necessitando de um melhor ajuste para coincidir com o contorno desejado. Estes contornos iniciais são então melhor ajustados para a fronteira da lesão pelo uso de um método de contorno ativo. Para este refinamento foi adotado o modelo de contorno ativo tradicional (*snakes*) (Kass et al., 1988), que tem como uma de suas características, a manutenção da topologia da curva inicial. O peso das forças externas adotado experimentalmente foi $0,1$, e os parâmetros de elasticidade de rigidez foram considerados iguais e com valor $0,05$ para todas as imagens. É importante ressaltar que os parâmetros adotados foram baixos para garantir a manutenção da topologia borda inicial da lesão e também para que a deformação da curva com o passar do tempo ocorresse lentamente, uma vez que a curva inicial pode-se encontrar próximo do limite desejado.

Com a última etapa do método proposto tenta-se extrair características das imagens que possam ser usadas para ajudar o especialista médico na definição das lesões a partir da regra ABCD, as quais são importantes por possibilitarem um diagnóstico inicial do tipo das lesões cancerosas.

A partir da borda da lesão detectada e extraída a caracterização pode ser feita calculando-se primeiramente o diâmetro do contorno (D), seguidos do cálculo da assimetria (A), irregularidade da borda (B) e variação da cor interna (C). De acordo com os especialistas médicos, o diâmetro da lesão é considerado como sendo o maior segmento de reta que corta a região lesionada, ligando dois pontos do seu contorno. Assim, adotou-se a maior distância euclidiana entre os pares de pontos do contorno para definir o diâmetro da lesão.

O cálculo do diâmetro foi o primeiro passo da caracterização devido à importância da maior diagonal da lesão para a definição de sua assimetria.

Aqui é proposto classificar as lesões em três classes de acordo com a sua assimetria: simétrica, levemente assimétrica e acentuadamente assimétrica. Esta classificação é feita a partir da análise das semi-retas perpendiculares à diagonal maior. Primeiramente, são definidas todas as linhas que ligam os pontos do contorno, cruzando a diagonal maior, e que sejam perpendiculares a esta. Em seguida, divide-se cada linha no ponto de interseção com a diagonal maior, obtendo duas novas linhas, uma de cada lado da diagonal maior, como ilustrado na Figura 2 pelas linhas (L_1) na cor rosa, e (L_2) em amarelo.

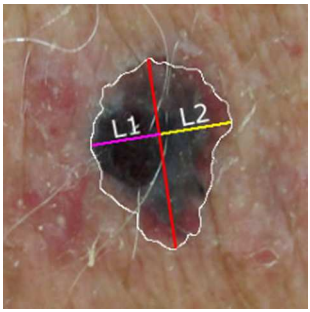


Figura 2. Exemplo de comparação de uma linha perpendicular à diagonal maior usada para definir a assimetria da lesão.

Após a divisão da linha perpendicular à diagonal maior, calcula-se a diferença de tamanho entre os seus segmentos (L_1) e (L_2), definindo a porcentagem desta diferença em relação ao maior dos segmentos. Assim, a diferença fica definida em função do maior lado da linha perpendicular. Este cálculo é feito para todas as perpendiculares. Para definir à qual classe de assimetria a lesão pertence, foram definidos dois limiares T_1 e T_2, usados para dividir o conjunto de perpendiculares à diagonal principal em três grupos: $Grupo_1$ contendo as perpendiculares cuja diferença for menor ou igual a T_1; $Grupo_2$ para as perpendiculares cuja diferença for maior do que T_1 e menor ou igual a T_2; e $Grupo_3$ com as perpendiculares cuja diferença for maior do que T_2.

Com o objetivo de definir estes limiares de forma que pudessem ser usados para todas as imagens do conjunto de testes, analisou-se a distribuição das porcentagens obtidas através das diferenças entre os segmentos L_1 e L_2 das retas perpendiculares à diagonal maior. O histograma obtido é formado por 100 colunas (referentes aos 100%), onde em cada coluna armazena-se a quantidade de diferenças com cada porcentagem. A partir da análise dos histogramas de uma amostra de 9 imagens do conjunto de testes, percebeu-se que ocorreu uma maior concentração entre os valores

0% (zero) e 30% para as bordas simétricas, e acima de 45% para as acentuadamente assimétricas. É importante ressaltar que as lesões usadas para análise dos histogramas tiveram suas assimetrias previamente definidas por uma análise visual de um dermatologista, sendo 3 imagens pertencentes a cada um dos grupos de assimetria definidos no trabalho. Assim, adotou-se $T_1 = 30$ e $T_2 = 45$. Com estes limiares as lesões foram classificadas dividindo-as em classes da seguinte forma: simétrica se o $Grupo_1$ contiver mais elementos do que os demais; levemente assimétrica se o $Grupo_2$ for maior que os restantes; e acentuadamente assimétrica se o $Grupo_3$ contiver mais elementos.

Para definir a irregularidade das bordas, propôs-se, uma abordagem baseada nos pontos de inflexão. Estes pontos são aqueles em que as curvas mudam de côncavas para convexas, ou vice-versa. Para encontrar estes pontos em que as bordas mudam de direção, desenvolveu-se duas técnicas. A primeira retira a assinatura da borda, e através desta procura os pontos em que ocorrem mudanças de direção, definindo seus pontos de inflexão (picos e vales). A segunda técnica percorre a assinatura da borda, calculando o produto vetorial entre dois vetores definidos por pontos pertencentes a esta.

Para a primeira técnica percorre-se todos os pontos p_i da assinatura da borda, analisando os quatro vizinhos do lado esquerdo e os quatro do lado direito de p_i. Na imagem (a) da Figura 3, tem-se o esquema de um ponto de inflexão na cor vermelha, e seus quatro vizinhos à esquerda (na cor azul) e à direita (na cor verde).

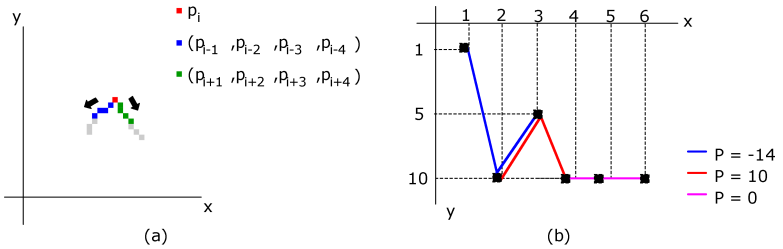


Figura 3. Exemplo de um ponto de inflexão e seus vizinhos direitos e esquerdos na imagem (a), e exemplo de produto vetorial na imagem (b).

Para determinar se o ponto p_i é uma inflexão, definiram-se pesos para cada um dos oito vizinhos de p_i (os quatro da direita e os quatro da esquerda) da seguinte forma: se o vizinho estiver abaixo de p_i em relação à coordenada y, ele recebe peso 1, senão, ele recebe peso -1. Para os casos onde a soma dos pesos for superior ou igual a 2, ou inferior ou igual a -2, para ambos os lados, considera-se a existência de uma inflexão. Caso

seja detectada uma mudança de direção e a soma dos pesos seja positiva, a inflexão é um pico, caso seja negativa é um vale. Assim, cada ponto de inflexão está associado a uma oscilação da borda.

Cada ispículo é formado por um pico cercado por um vale à direita e outro à esquerda. Como este método faz uma análise apenas dos vizinhos próximos de p_i, ele retorna as pequenas variações da borda, dificultando a detecção dos ispículos maiores. Na tentativa de obter um método para definição dos ispículos maiores, desenvolveu-se uma segunda técnica para calcular as inflexões, baseando-se no produto vetorial. Considerando três pontos, $p_1 = (x_1, y_1)$, $p_2 = (x_2, y_2)$ e $p_3 = (x_3, y_3)$ pertencentes à borda da lesão, pode-se definir o direcionamento da curva formada pelos pontos $(\widehat{p_1 p_2 p_3})$, a partir do produto vetorial entre os vetores definidos pelos pontos $p_1\vec{p}_2$ e $p_2\vec{p}_3$. Usando a convenção de leitura da imagem, isto é, da esquerda para a direita e de cima para baixo, tem-se que se o produto vetorial p for maior que 0 (zero), p_1, p_2 e p_3 constituem um pico, se for menor que 0 (zero) constituem um vale, e se for igual a 0 (zero) formam uma reta.

$$p = (x_2 - x_1).(y_3 - y_1) - (y_2 - y_1).(x_3 - x_1) \quad (1)$$

Na imagem (b) da Figura 3, têm-se exemplos de três sequências de pontos e o seu produto vetorial. Pode-se observar que a curva azul forma um vale (produto vetorial negativo), enquanto a vermelha forma um pico (produto vetorial positivo) e a rosa forma um segmento de reta (produto vetorial igual a 0 (zero)). Vale ressaltar que apenas na imagem (b) desta figura foi exibido o sistema de coordenadas com a origem no canto superior esquerdo. A mudança para esta imagem foi feita apenas para facilitar o entendimento de como o produto vetorial foi usado para definir os picos e vales.

A técnica que utiliza o produto vetorial considera os vetores formados pelos pontos $(p_1 - 15; p_i; p_i + 15)$, sendo $1 \leq i \leq n$, e n o número de pontos da borda. Esta variação de 15 pixels para a esquerda e para a direita foi adotada para descartar os ispículos muito pequenos, que já foram calculados pela técnica anterior. Assim, esta etapa do método calcula a quantidade de ispículos e vales grandes, e também a quantidade de ispículos e vales pequenos. As duas informações são importantes para definir, respectivamente, a irregularidade e a rugosidade dos contornos das lesões.

A variação da coloração interna das lesões foi calculada usando o canal H (matiz) do sistema de cores HSV. Este canal foi escolhido por conter a variação de cor de uma imagem. Para definir a quantidade de cores do tecido doente, a matiz foi dividida linearmente em 10 intervalos, chamados de classes de cores. Todas as cores pertencentes a um mesmo intervalo foram consideradas semelhantes. Assim, para calcular a quantidade de cores da lesão, percorre-se a região limitada pelo contorno, contando quantos

pixels existem em cada um dos intervalos de cor. Após fazer esta contagem, descartam-se as classes que contêm menos de 100 pixels. Isto é feito para evitar que pixels isolados, ou regiões muito pequenas (como pequenos artefatos) interfiram na contagem do número de cores da lesão.

4. Resultados

Os testes experimentais do método proposto foram feitos em um conjunto de imagens de pele com lesões, formado por 40 imagens coloridas, com resolução de 256x256 pixels, retiradas de Dermatlas (2012) e Goshtasby (2012). Foram utilizadas imagens contendo lesões de três classificações: lesões atípicas, lesões malignas e lesões não-malignas. As imagens desta fonte foram usadas porque, além das imagens, são disponibilizadas também informações sobre o diagnóstico final das regiões afetadas. 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. A validação dos resultados foi feita por um especialista em doenças dermatológicas.

A Figura 4 contém algumas imagens processadas pelo método desenvolvido. Nesta figura tem-se as imagens originais e suas respectivas bordas, resultantes da aplicação da etapa de segmentação proposta. A lesão da imagem 1 é uma lesão atípica, da 2 e 3 são não-malignas e da 4 e 5 foram diagnosticadas como lesões malignas.

Nas imagens apresentadas, nota-se que as lesões foram detectadas e envolvidas pelo contorno final. Todos os contornos finais obtidos foram visualmente avaliados pelo especialista que confirmou que as lesões existentes nas imagens em estudo foram detectadas com sucesso. Após detectar as lesões e extrair suas bordas, aplicou-se a etapa de caracterização das lesões. Nesta etapa, o objetivo é extrair características que possam ser usadas para diagnosticar as áreas doentes de acordo com a regra ABCD. Na Tabela 2, tem-se os dados retornados por esta etapa para as mesmas imagens da Figura 4.

Na segunda coluna da Tabela 2, tem-se a classificação das lesões de acordo com a sua simetria. A quantidade de ispículos grandes e pequenos aparece na terceira coluna. Estes dados foram retornados para que o dermatologista tenha a ideia da irregularidade das bordas. Vale ressaltar que, para fazer a analise da irregularidade, deve-se considerar o tamanho da lesão, uma vez que a tendência é que o numero de ispículos seja menor quando a região doente for menor.

Pelos resultados percebeu-se também que nas lesões atípicas e malignas, a quantidade de cores internas foi maior do que nas lesões não-malignas. Ao passo que foram encontradas poucas cores nas lesões não malignas, na maior parte das malignas detectou-se mais de 5 cores, o que também atende a regra ABCD.

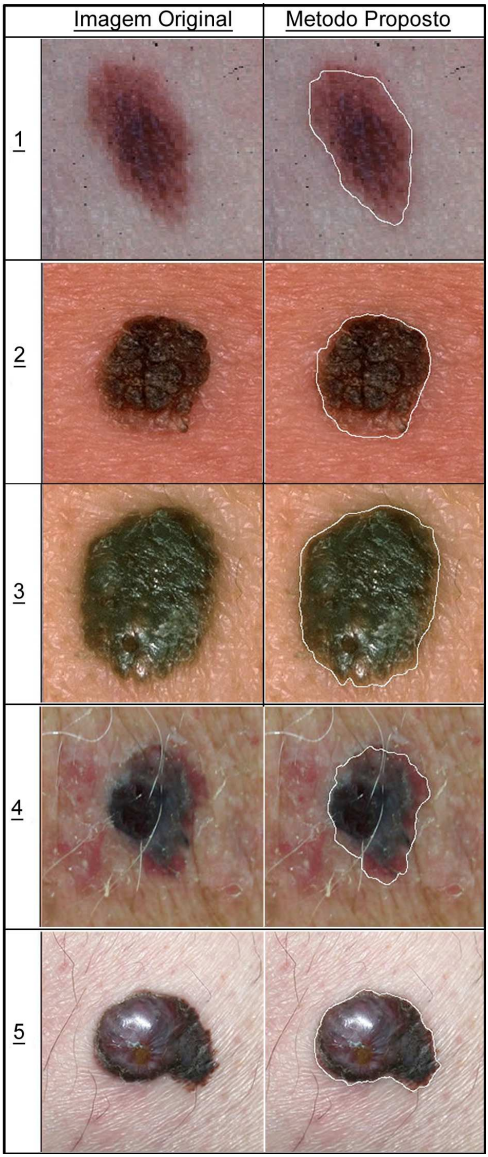


Figura 4. Resultados do método proposto aplicado a imagens com diferentes características.

Tabela 2. Características extraídas das lesões.

Img	Assimetria	Irregularidade da borda	Cores	Diâmetro (Pixels)
1	Levemente Assimétrica	7 ispículos e 5 vales grandes, e 13 ispículos pequenos	9	167
2	Levemente Assimétrica	8 ispículos e 1 vale grandes, e 18 ispículos pequenos	1	150
3	Levemente Assimétrica	11 ispículos e 4 vales grandes, e 25 ispículos pequenos	4	216
4	Acentuadamente Assimétrica	6 ispículos e 3 vales grandes, e 18 ispículos pequenos	10	162
5	Acentuadamente Assimétrica	6 ispículos e 3 vales grandes, e 21 ispículos pequenos	8	149

A medida do diâmetro foi calculada em pixels, pois a maioria das imagens usadas nos testes não trazia informações que permitissem montar uma relação, e definir aproximadamente esta medida em milímetros. Para demonstrar a validade do cálculo do diâmetro usado, apresenta-se na Figura 5 uma imagem com uma referência para cálculo da distância em milímetros. Na Figura 5, em (a) tem-se uma imagem contendo uma lesão de pele com aproximadamente 5 milímetros de diâmetro, como pode-se observar pela régua usada como referência na parte superior da imagem (a). Na imagem (b) tem-se o valor em milímetros (4,73 mm), do diâmetro calculado pelo método de caracterização desenvolvido. Este cálculo foi feito a partir da relação entre o número de pixels da maior diagonal da lesão, e o número de pixels necessários para representar cada milímetro na imagem original. A borda da lesão e o seu diâmetro foram processados adequadamente por meio do método desenvolvido.

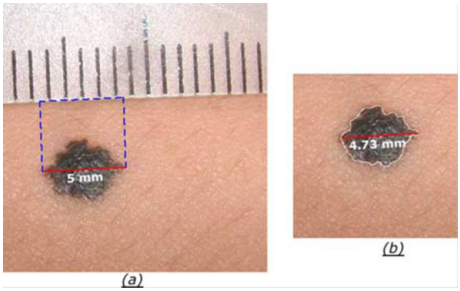


Figura 5. Cálculo do diâmetro da lesão. Em (a) o diâmetro da lesão definido pelo especialista, e em (b) o diâmetro calculado a partir da borda retornada pelo método proposto.

Após obter estes resultados, foi avaliada a qualidade das bordas finais detectadas e extraídas. Os dados desta análise são apresentados na Tabela 3.

Tabela 3. Avaliação da qualidade das bordas finais, considerando o conjunto de imagens adotadas para teste.

Resultados aceitáveis	Resultados ruins
92%	8%

A partir dos dados apresentados nas Tabelas 2 e 3, verifica-se que o procedimento desenvolvido é promissor, e obteve um bom desempenho. Todos os contornos finais obtidos foram visualmente avaliados pelo especialista que confirmou que as lesões existentes nas imagens em estudo foram todas detectadas com sucesso, e em aproximadamente 92% dos casos, os contornos extraídos foram considerados aceitáveis pela avaliação médica, envolvendo todas as áreas lesionadas nas imagens.

5. Conclusões

Um método híbrido para segmentação de imagens coloridas de lesões de pele foi apresentado. O método desenvolvido une as características das técnicas de crescimento de regiões e de contornos ativos para detectar, extrair e refinar as bordas das lesões, preservando suas características principais.

A partir da análise visual dos resultados obtidos, realizada por um especialista médico, pode-se concluir que o método proposto é promissor, sendo capaz de detectar as prováveis regiões doentes em imagens de lesões de pele, obtidas a partir de fotografias convencionais.

No entanto, o método proposto ainda possui algumas limitações quando a imagem original possui transições muito suaves entre as regiões doente e saudável, quando existe a presença de reflexos sobre as transições, ou quando as lesões encontram-se no couro cabeludo e existe uma presença excessiva de pêlos na região lesionada. Para superar o problema das transições muito suaves, pretende-se realizar testes usando outros métodos de contorno ativo para pós-processar as imagens, na tentativa de realizar uma deformação mais adequada do contorno aproximado dado pelo método de crescimento de regiões. Além disto, na tentativa de solucionar as falhas para extrair o contorno das lesões no couro cabeludo e com reflexos, serão testados outros parâmetros para a metodologia de suavização adotada. Outra etapa a ser desenvolvida na sequência é uma comparação estatística do método proposto com os métodos existentes.

6. 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), 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

Araujo, A.F.; Tavares, J.M.R.S.; Oliveira, R.B.; Marranghello, N.; Pereira, A.S. & Rossetti, R.B., Uma metodologia híbrida para segmentação de lesões de pele. In: *Anais do VII Workshop de Visão Computacional.* p. 173–178, 2011.

Babini, M. & Marranghello, N., *Introdução às redes neurais artificiais.* 1^{st} edição. São José do Rio Preto, SP: Cultura Acadêmica Editora, 2007.

Barcelos, C.A.; Boaventura, M. & Silva Jr., E.C., A well-balanced flow equation for noise removal and edge detection. *IEEE Transactions on Image Processing*, 12(7):751–763, 2003.

Barcelos, C.A.Z. & Pires, V.B., An automatic based nonlinear diffusion equations scheme for skin lesion segmentation. *Applied Mathematics and Computation*, 215(1):251–261, 2009.

Bishop, C.M., *Neural networks for pattern recongition.* 1^{st} edição. New York, USA: Oxford University Press, 2004.

Celebi, M.E.; Kingravi, H.A.; H. Iyatomi, Y.A.A.; Stoecker, W.V.; Moss, R.H.; Malters, J.M.; Grichnik, J.M.; Marghoob, A.A.; Rabinovitz., H.S. & Menzies, S.W., Border detection in dermoscopy images using statistical region merging. *Journal of International Society for Bioengineering*, 14(3):347–353, 2008.

Dermatlas, M.I., Dermatology image atlas. Disponível em http://dermatlas.med.jhmi.edu/derm, 2012.

Dokur, Z. & Ölmez, T., Segmentation of ultrasound images by using a hybrid neural network. *Pattern Recognition Letters*, 23:1825–1836, 2002.

Gonzalez, R.C.; Woods, R.E. & Eddins, S.L., *Digital Image Processing Using MATLAB.* 1^{st} edição. Upper Saddle River, USA: Pearson Prentice-Hall, 2003.

Goshtasby, A., Segmentation of skin cancer images. Disponível em http://www.cs.wright.edu/ agoshtas/paper_fig.html, 2012.

Guide, S.C., Skin cancer self-examination. Disponível em http://www.skincancerguide.ca/prevention/self_examination.html, 2012.

Hashemi, S.; Kiani, S.; Noroozi, N. & Moghaddam, M.E., An image contrast enhancement method based on genetic algorithm. *Pattern Recognition Letters*, 31:1816–1824, 2010.

Hudson, D.L. & Cohen, M.E., *Neural networks and artificial intelligence for biomedical engineering*. New York: Wiley-IEEE Press, 2000.

Jia-Xin, C. & Sen, L., A medical image segmentation method based on watershed transform. In: *Proceedings of the Fifth International Conference on Computer and Information Technology*. Piscataway, USA: IEEE Press, p. 634–638, 2005.

Jianli, L. & Baoqi, Z., The segmentation of skin cancer image based on genetic neural network. In: *Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering*. Washington, USA: IEEE Computer Society, v. 5, p. 594–599, 2009.

Kass, M.; Witkin, A. & Terzopoulos, D., Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

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*, 175:177–199, 2005.

Lu, R. & Shen, Y., Automatic ultrasound image segmentation by active contour model based on texture. In: *Proceedings of the First International Conference on Innovative Computing, Information and Control*. Piscataway, USA: IEEE Computer Society, v. 2, p. 689–692, 2006.

Ma, Z.; Tavares, J.M.R.; Jorge, R.N. & Mascarenhas, T., A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. *Computer Methods in Biomechanics and Biomedical Engineering*, 13(2):235–246, 2010.

Morris-Smith, J.D., *Characterisation of the appearance of pigmented skin lesions*. Phd thesis, School of Computer Science, The University of Birmingham, Birmingham, UK, 1996.

Mukhopadhyay, A. & Maulik, U., A multiobjective approach to MR brain image segmentation. *Applied Soft Computing*, 11:872–880, 2011.

da Silva, I.N.; Spatti, D.H. & Flauzino, R.A., *Redes Neurais Artificiais: para engenharia e ciências aplicadas*. 1^{st} edição. São Paulo, SP: Artliber, 2010.

Yoon, S.W.; Lee, C.; Kim, J.K. & Lee, M., Wavelet-based multi-resolution deformation for medical endoscopic image segmentation. *Journal of Medical Systems*, 32:207–214, 2008.

Notas Biográficas

Alex Fernando de Araujo é graduado e mestre em Ciência da Computação (Universidade Federal de Goiás/Catalão, 2007 e Universidade Estadual Paulista "Júlio de Mesquita Filho"/São José do Rio Preto, 2010, respectivamente). Atualmente é doutorando em Engenharia Informática (Universidade do Porto, Portugal).

João Manuel R. S. Tavares é licenciado em Engenharia Mecânica (Universidade do Porto – FEUP, 1992), mestre e doutor em Engenharia Electrotécnica e de Computadores (FEUP, 1995 e 2001, respectivamente). Desde 2001, é Investigador Senior e Coordenador de projeto no Laboratório de Óptica e Mecânica Experimental (LOME), do Instituto de Engenharia Mecânica e Gestão Industrial (INEGI). Foi Professor Auxiliar do Departamento de Engenharia Mecânica (DEMec) da FEUP desde 2001 até 2011 e é Professor Associado do mesmo departamento desde 2011.

Roberta Barbosa Oliveira é graduado em Sistemas de Informação (Fundação Educacional de Fernandópolis, SP, 2008) e atualmente é mestrando em Ciência da Computação (Universidade Estadual Paulista, São José do Rio Preto, SP).

Ricardo Baccaro Rossetti é médico, especialista em dermatologia, pesquisador do laboratório da clínica DERM (São José do Rio Preto, SP).

Norian Marranghello é graduado em Engenharia Eletrônica (Pontifícia Universidade Católica do Rio Grande do Sul, 1982), mestre e doutor em Engenharia Elétrica (Universidade Estadual de Campinas, 1987 e 1992, respectivamente). Tem pós-doutorado em Sistemas de Computação (Universidade de Aarhus, Dinamarca, 1998) e livre-docência em Sistemas Digitais (Universidade Estadual Paulista – UNESP, 1998). Atualmente é Professor Titular da UNESP e tem experiência nas áreas de Engenharia Elétrica e Ciência da Computação, com ênfase em Sistemas Digitais, atuando principalmente nos seguintes temas: sistemas digitais integráveis, modelagem e simulação de sistemas, arquiteturas reconfiguráveis, redes de Petri e síntese de sistemas digitais.

Aledir Silveira Pereira é graduado e mestre em Engenharia Elétrica (Fundação Educacional de Barretos, SP, 1980 e Universidade de São Paulo – USP, 1987) e doutor em Física Aplicada Computacional (USP, 1995). Atualmente é professor assistente doutor na UNESP – Universidade Estadual Paulista junto ao IBILCE – Instituto de Biociências, Letras e Ciências Exatas – Campus de São José do Rio Preto. Tem experiência em processamento digital de imagens, sistemas digitais, automação e controle industrial e ensino de computação.

Capítulo 3

Comparação de Imagens Tomográficas *Cone-Beam* e *Multi-Slice* Através da Entropia de Tsallis e da Divergência de Kullback-Leibler

André Sobiecki, Celso Denis Gallão,
Daniel Cardoso Cosme e Paulo Sérgio Silva Rodrigues *

Resumo: Registro de imagens é uma técnica que compara duas imagens para fins de alinhamento e calibração. Geralmente, podemos realizar o alinhamento das características geométricas, bem como perfil de luminância. O alinhamento de luminância está relacionado com a quantidade de informações entre duas imagens. Até o final dos anos 90 a estratégia mais conhecida para transferência de informação entre duas distribuições de probabilidade de sistemas físicos era através da clássica entropia de Shannon. A melhoria deste tipo de formalismo atualmente é conhecida como entropia de Tsallis. Este capítulo apresenta uma análise da distribuição de probabilidade de luminância entre imagens adquiridas de técnicas *cone-beam* e *multi-slice*, baseadas na divergência de Kullback-Leibler estendida pela estatística de Tsallis.

Palavras-chave: Processamento de imagens, Entropia, Divergência de Kullback-Leibler.

Abstract: *Image registration is a technique that compares two images for alignment and calibration purposes. Generally, we can accomplish alignment of geometric features as well as luminance profile. The luminance alignment is closely 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. This paper presents an analysis of the probability luminance distribution between images acquired from cone-beam and multi-slice techniques based on the Kullback-Leibler divergence, extended for Tsallis statistics.*

Keywords: *Image processing, Entropy, Kullback-Leibler divergence.*

*Autor para contato: psergio@fei.edu.br

1. Introdução

Na medicina atual, é vital o uso de imagens para avaliação e acompanhamento de pacientes, como mostram Maintz & Viergever (1998), cujos métodos de aquisição avançam a medida em que as tecnologias também avançam. A Tomografia Computadorizada (TC) pode ser apontada como um exemplo comum, sendo um dos exames mais utilizados e confiáveis, atualmente.

Diferentes técnicas para a aquisição de imagens são desenvolvidas buscando melhorar a qualidade do exame com o menor dano para o paciente. Apesar de suas vantagens, métodos como o *Multi-Slice Computerized Tomography* (MSCT) expõe o paciente a radiação. Em contrapardida, estudos demonstram que utilizar técnicas específicas para determinadas regiões do corpo humano, como a técnica *Cone-Beam Computerized Tomography* (CBCT) que é utilizada na área de ortodontia, podem ser mais eficientes ao mesmo tempo em que são menos prejudiciais ao paciente, conforme mostrado em Loubele et al. (2007, 2009).

Há, portanto, a necessidade de se avaliar as diferenças entre as duas técnicas, levando-se em consideração a análise das imagens produzidas, visando minimizar erros ou dificuldades de diagnóstico. Para que seja minimizada a diferença entre os dois tipos de imagens, é necessário realizar, primeiramente, o registro entre ambas, buscando o alinhamento espacial e de luminância como sugerem Rodrigues & Giraldi (2009).

O Alinhamento espacial pode ser determinado com técnicas de transformações geométricas, tais como em Rodrigues et al. (2006b), Rodrigues & Giraldi (2009), Rodrigues & Giraldi (2011). Uma vez calculadas as transformações geométricas, as diferenças entre as imagens podem ser computadas através da sobreposição de seus histogramas de luminâncias. As técnicas mais comuns para registro de luminância consideram os respectivos histogramas das imagens.

Na área da Teoria da Informação, sabe-se que a distribuição de probabilidade da luminosidade de imagem pode transmitir informações relacionadas ao seu conteúdo semântico. Outras características tais como cor, textura e relacionamento espacial entre as regiões dominantes também podem carregar informações semelhantes. No entanto, devido ao alto grau de correlação entre as características de pixels, pode ser difícil medir essas propriedades.

A maneira tradicional de comparar a quantidade de informação entre duas imagens é através do cálculo das suas respectivas entropias relativas, melhor dizendo, através da divergência de Kulback-Leibler. Recentemente, nos trabalhos de Albuquerque et al. (2004), Esquef (2002), Rodrigues & Giraldi (2011), Rodrigues & Giraldi (2009) e Rodrigues et al. (2006b), foram apresentadas evidências de que imagens médicas podem ser melhor explicadas se considerarmos seus respectivos sistemas físicos como sendo do

tipo não-extensivo, que significa possuírem interações espaciais e temporais de longo alcance. No que diz respeito ao registro de imagens, esse tema foi pouco explorado na literatura, até o momento.

Este trabalho é uma expansão do artigo Sobiecki et al. (2011). Neste trabalho, é apresentado um estudo de técnicas baseadas em entropia não-extensiva para comparação de imagens médicas adquiridas pelos métodos CBCT e MSCT, em imagens de tomografia computadorizada, utilizadas na área de ortodontia.

Foram realizados estudos através da divergência de Kullback-Leibler estendida pela entropia de Tsallis. Os resultados mostram o poder dessa metodologia recente para medir a relação entre duas distribuições de probabilidade, e sugerem que as interações entre os pixels de CBCT e MSCT podem ter um comportamento sub-extensivo.

2. Entropia Não-Extensiva

O termo entropia surgiu primeiramente no campo da termodinâmica, sendo utilizado para demonstrar comportamentos microscópicos sob processos físicos macroscópicos. Inicialmente, foi considerada como uma propriedade aplicável somente ao contexto da termodinâmica. Posteriormente, Ludwig von Boltzmann e Willard Gibbs mostraram a entropia como uma medida estatística possibilitando utilizá-la em outros contextos, em diversas aplicações, surgindo então a conhecida Equação de Boltzmann-Gibbs,

$$S = k \log W, \tag{1}$$

onde a entropia (S) é o produto da constante de Boltzmann (k) pelo logaritmo do mínimo de estado (W), conforme Tavares (2003) e Esquef (2002).

Mais tarde, Claude Shannon ofereceu uma importante contribuição utilizando o conceito de entropia à luz de um novo contexto. Em sua obra (Shannon (1948)), Shannon derivou a conhecida equação

$$S = -\sum_i p_i \ln(p_i), \tag{2}$$

onde p_i é a probabilidade de encontrar o sistema no estado i, e $P = [p_1, \ldots, p_k]$, $0 \leq p_i \leq 1$ e $\sum_i p_i = 1$, que ficou conhecida como a entropia de *Boltzmann-Gibbs-Shannon* (BGS). Sistemas que podem ser descritos pela entropia de BGS são chamados de sistemas extensivos e possuem, entre outras, a propriedade da aditividade (3), dada por:

$$S(A * B) = S(A) + S(B), \tag{3}$$

onde $S(A * B)$ é a entropia de um sistema composto por duas variáveis independentes, $S(A)$ e $S(B)$, calculado de acordo com a Equação (2).

Entretanto, a entropia de BGS não descreve sistemas físicos que envolvem efeitos não-extensivos, sendo necessária uma generalização. Para sistemas considerados do tipo não-extensivos, Tsallis (1988) definiu em seu trabalho a seguinte equação:

$$S_q(p_1, \ldots, p_k) = \frac{1 - \sum_{i=1}^{k} p_i^q}{q - 1} \tag{4}$$

A entropia de Tsallis é uma fórmula generalizada de entropia, a partir da entropia de Boltzmann-Gibbs-Shannon, onde k é o número de possibilidades do sistema e q é o índice de entropia que caracteriza o grau de não-extensividade do sistema. Note que, quando $q \to 1$, a Equação (4) transforma-se na tradicional Equação (2) da entropia de BGS, segundo Tavares (2003).

A propriedade de aditividade pode explicar melhor os sistemas extensivos, porém falha na explicação de sistemas não-extensivos, onde Tsallis propõe a seguinte generalização, chamada de propriedade da pseudo-aditividade, dada por

$$S_q(A * B) = S_q(A) + S_q(B) + (1 - q)S_q(A)S_q(B). \tag{5}$$

Ambas as equações, da aditividade e da pseudo-aditividade, são amplamente utilizadas em metodologias de limiarização de imagem para extração de informações. No entanto, atingir um limiar ótimo ainda é um desafio.

Em Kapur et al. (1985) foi proposto um método de limiarização utilizando a entropia de BGS. O trabalho considera duas distribuições de probabilidade, sendo uma para o primeiro plano (*foreground*) e outra para o fundo (*background*). Então, o limiar é dado calculado através da Equação (3).

Em Albuquerque et al. (2004) a metodologia proposta por Kapur et al. (1985) foi melhorada com o uso do formalismo não-extensivo. No trabalho de Rodrigues et al. (2006b) foi proposto um algoritmo para a classificação de imagens de cânceres de mama, obtidas por ultra-som. Sua estratégia utiliza a entropia não-extensiva.

Também, em Rodrigues et al. (2006b) foi proposto o primeiro algoritmo para o cálculo automático do índice de não-extensividade q, necessário para o formalismo de Tsallis, otimizando resultados. Em Esquef (2002) são discutidos e apresentados resultados com base na entropia relativa não-extensiva, ou entropia relativa generalizada, ou simplesmente Divergência de Kullback-Leibler Estendida, que será detalhada na seção seguinte.

3. Cálculo do índice q

Considerando o *background* e o *foreground* de uma imagem como subsistemas físicos independentes, a estratégia proposta por Pun (1981) para

segmentação de imagens, utiliza a propriedade de aditividade dos sistemas extensivos, dada pela Equação (3), para obter o *threshold* ótimo entre os sub-sistemas. Essa ideia vem do fato de que atingi-se o máximo de informação possível de ser transferida quando se calcula a entropia global máxima, através da soma de ambos os sub-sistemas.

O mesmo argumento funciona para sistemas não-extensivos, onde o formalismo utilizado aplica-se de acordo com a Equação (5). O formalismo de Tsallis é uma generalização da entropia de Shannon, reduzindo-se ao sistema tradicional somente quando $q \to 1$. Assim, podemos concluir que a q-entropia, como também é chamada, permite capturar ambos comportamentos, extensivo e não-extensivo, sendo portanto razoável investigar abordagens de segmentação entrópicas sob os dois contextos.

A utilização de um novo parâmetro traz um custo computacional extra, e apesar de sua classe, cada imagem ou região pode demandar um valor de q diferente (incluindo $q = 1$), a fim de obter-se a maximização da informação. Desta forma, torna-se interessante avaliar o valor da entropia computada para cada imagem em diversos intervalos de q; considerando sistemas sub-extensivos ($q < 1$), extensivos ($q = 1$) e super-extensivos ($q > 1$), conforme discutido em Tavares (2003).

Sob o ponto de vista da Teoria da Informação, quanto menor a entropia máxima S_q produzida por um valor q relacionado com a entropia teórica máxima S_{max} de um sistema físico (neste caso, uma imagem), maior é a auto-informação contida nesse sistema. Este é um princípio bem conhecido da Teoria da Informação, que nos conduz à idéia de que o valor ótimo de q pode ser alcançado minimizando a relação S_q/S_{max}. Então, calculamos o valor ótimo de q, ressaltando a imagem, conforme segue.

Para cada valor de q no intervalo $[0,01, 0,02, ..., 2,0]$ calculamos o q ótimo como sendo aquele que minimiza a razão S_q/S_{max}. Trabalhamos aqui com a hipótese de que, não apenas cada imagem natural pode comportar-se como um sistema não-extensivo singular – e, como tal, exigindo um valor de q diferente para a segmentação – mas também as suas regiões internas também podem ser não-extensivas singulares, também exigindo diferentes valores de q, conforme apresentado no trabalho de Rodrigues & Giraldi (2009).

A fim de aplicar valores diferentes de q para segmentar diferentes regiões de uma imagem, e para atingir a maioria das principais regiões envolvidas, realizou-se dois níveis de segmentação. Inicialmente, calculamos o valor de q minimizando S_q/S_{max} e aplicando a Equação (6) para obter um primeiro limiar ótimo t_{opt}, obtendo uma primeira segmentação que separa *background* (R_B) do *foreground* (R_F). Então, para cada região considerada (R_B e R_F) calculamos novos valores de q, tratando R_B e R_F como sistemas físicos diferentes. Aplicamos o algoritmo novamente, obtendo dois novos $t_{opt}s$. Assim, podemos alcançar, no máximo, quatro separações de intensidade e várias regiões na imagem.

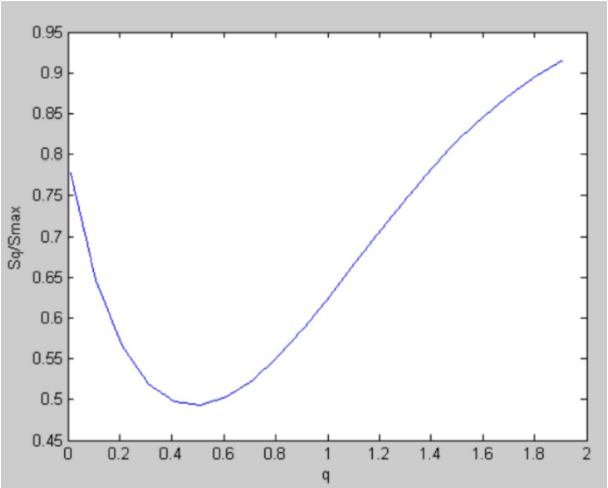


Figura 1. Gráfico de S_q/S_{max} em função dos valores de q. O menor valor, correspondendo a $q = 0.46$, é o q ótimo utilizado para a segmentação inicial.

A Figura 2 mostra um exemplo onde, a esquerda está a imagem original e no centro a sua primeira segmentação em duas regiões (R_B e R_F), conseguido o melhor $q = 0,46$, que corresponde ao valor mínimo da curva da Figura 1. Seguindo a mesma idéia para regiões R_B e R_F, calculamos novos valores de q minimizando novas curvas S_q/S_{max} e alcançar dois novos limiares t_{opt}. O resultado pode ser visto na Figura 2 a direita. Neste caso encontramos $q = 0,15$ para R_B e $q = 0,73$ para R_F, sugerindo um sistema com comportamento sub-extensivo para todas as regiões.

4. Divergência de Kullback-Leibler

A divergência de Kullback-Leibler, ou simplesmente Entropia Relativa como também é conhecida, apresentada no trabalho de Kullback & Leibler (1951), é semelhante á entropia de BGS. Além disto, considera duas distribuições de probabilidade e calcula a divergência entre elas. Através da divergência de Kullback-Leibler podemos medir o ganho de informações entre duas regiões a partir das mesmas imagens. A divergência de Kullback-Leibler tradicional é aplicável á sistemas considerados extensivos, sendo definida pela Equação 6:

$$D_{KL}(P : P') = \sum_{i}^{n} p_i \cdot \log \frac{p_i}{p_i{}'}, \tag{6}$$

Figura 2. Ilustração comparativa de uma imagem natural: (esquerda) a imagem original. (centro) a primeira segmentação com $q = 0.46$ alcançando R_B e R_F. (direita) a segmentação final com $q = 0.15$ e $q = 0.73$ para R_B e R_F, respectivamente.

onde P e P' são duas distribuições de probabilidades:

$$P = (p_1, p_2, \ldots, p_n)$$

e

$$P' = ({p_1}', {p_2}', \ldots, {p_n}') \ .$$

A divergência de Kullback-Leibler possui uma equivalência ao utilizar o formalismo não-extensivo proposto por Tsallis, conforme apresentado no trabalho de Borland et al. (1998), e definida como

$$D_{KL_q}(P : P') = \sum_i^n \frac{p^q}{q-1} \cdot (p_i^{1-q} - {p_i}'^{1-q}). \tag{7}$$

Semelhante à equação (6), que mede a divergência em sistemas extensivos, a Equação (7) calcula a divergência entre duas distribuições de probabilidade para sistemas não-extensivos, a chamada Divergência de Kullback-Leibler Estendida. No limite, quando $q \to 1$, pode-se mostrar que a Equação (7) reduz-se a Equação (6), ou seja, a Equação (7) é uma generalização da Equação (6).

Os trabalhos de Albuquerque et al. (2004), Giraldi et al. (2008), Giraldi et al. (2006) e Rodrigues et al. (2006a) sugerem que a estatística de Tsallis é uma poderosa ferramenta para a segmentação de imagens médicas, e os trabalhos de Erdmann (2009), Lessa (2010), Albuquerque & Esquef (2008) e Olívio (2009) mostram que a divergência de Kullback-Leibler, para sistemas não-extensivos, tem resultados promissores quando aplicado ao problema de classificação de imagens.

Normalmente, não é possível saber *a priori* o comportamento dos sistemas para classificá-los como extensivos ou não-extensivos. A solução usual é calcular a divergência para vários valores de q, incluindo $q = 1$, e escolher o valor de q com o melhor desempenho. Quando $q < 1$, diz-se que o sistema tem um comportamento sub-extensivo, quando $q = 1$ observa-se

que o sistema é extensivo, e quando $q > 1$, classifica-se o sistema como super-extensivo, conforme apresentado em Tsallis (2001).

Em nosso trabalho, propomos o uso de divergência de Kullback-Leibler Estendida e a investigação do valor de q, que minimiza a quantidade de informação entre os dois tipos de imagens de TC: *Cone-Beam* e *Multi-Slice*, ambas aplicadas em ortodontia.

5. Metodologia Proposta

Por estarmos interessados em investigar o poder da divergência de Kullback-Leibler como uma ferramenta para medir somente a informação de luminância, e também por as imagens utilizadas nos testes estarem em nível de cinza, a comparação aqui proposta é realizada somente em imagens em tons de cinza representados por $L = [0 : 255]$. Então, a escala completa de Hounsfield (1973) (HU) para imagens TC foi mapeada para esta faixa, incluindo: água $= 0HU$; ar $= -1000HU$; ossos = acima de $1000HU$, entre outros.

A Figura 3 mostra o diagrama da metodologia proposta, sendo cada etapa descrita a seguir.

Na Etapa 1 temos as imagens tomográficas de entrada, CBCT e MSCT, ainda não-normalizadas.

A normalização é realizada manualmente na Etapa 2, onde as imagens capturadas pelo método CBCT foram aparadas, e sem necessidade de re-dimensionamento ou rotações. Por outro lado, as imagens MSCT foram aparadas e rotacionadas, a fim de convergir com as imagens CBCT. Uma vez que as distribuições de probabilidades são invariantes à rotação, bem como a translação, essas tarefas geométricas não tem qualquer influência sobre os nossos resultados, e foram realizadas apenas com o propósito de melhorar a visualização. Pela mesma razão, não há necessidade de adequar seus respectivos centros geométricos.

A Figura 4 (a) superior mostra um exemplo de imagem obtida por CBCT, e a Figura 4 (a) inferior mostra um exemplo de uma imagem correspondente obtida por MSCT. A Figura 4 (b) mostra as respectivas imagens já normalizadas, como representação da Etapa 3 do diagrama da metodologia.

Na Etapa 4, são obtidas as distribuições de probabilidades de cada imagem através de seus respectivos histogramas. Ambos os histogramas são mostrados na Figura 5, como representação da Etapa 5, onde pode-se perceber que as imagens são semelhantes no que se refere à probabilidade de pixels.

Com os histogramas das imagens calculados, efetua-se a divergência de Kullback-Leibler, na Etapa 6.

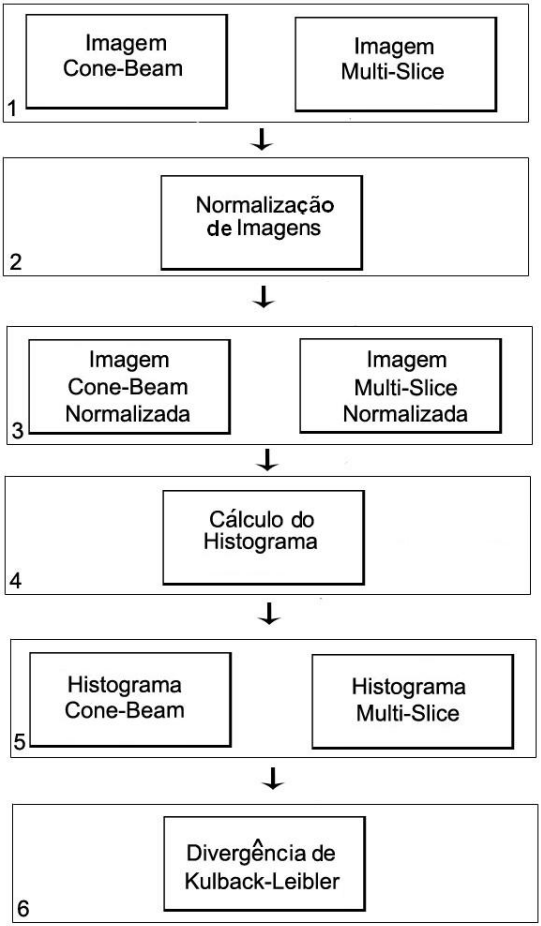


Figura 3. Metodologia proposta.

6. Resultados Experimentais

Para os experimentos foram utilizadas 16 imagens, sendo 8 imagens de exame *Cone-Beam* e 8 imagens de exame *Multi-Slice*. As imagens foram normalizadas através dos softwares de edição de imagens Corel Photo-Paint

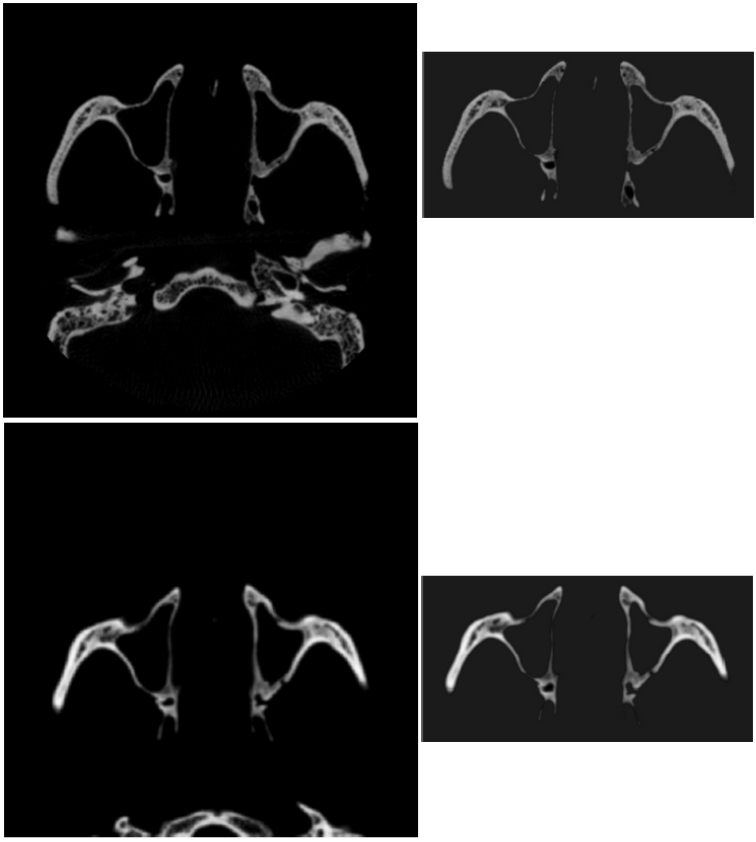


Figura 4. Comparação de imagens originais com imagens normalizadas: (a) imagens antes da etapa de normalização. (b) as correspondentes imagens após a etapa de normalização.

X3 e Corel Draw X3 [1], exportadas na resolução de 300 dpi (tanto na vertical quanto na horizontal), medindo 500 pixels de largura por 250 pixels de altura, em tons de cinza (8 bits), no formato *jpeg*.

[1] Corel Photo-Paint X3 e Corel Draw X3: é 2005 Corel Corporation, todos os direitos resevados.

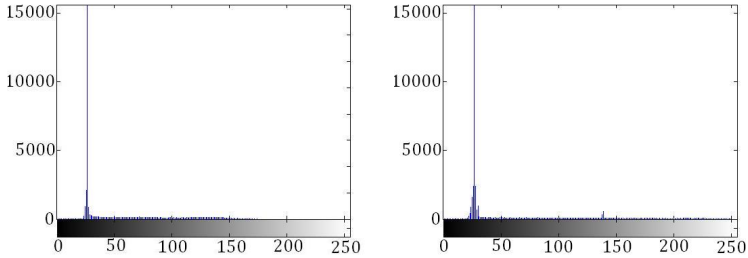


Figura 5. Histograma das imagens normalizadas: (esquerda) histograma da imagem CBCT. (direita) histograma da imagem MSCT.

A Figura 6 ilustra algumas das imagens utilizadas nos experimentos: as imagens superiores (a) referem-se aos exames *Cone-Beam* e as imagens inferiores (b) referem-se aos exame *Mult-Slice*. Percebe-se de modo geral que as imagens são muito semelhantes, porém com algumas diferenças de estrutura e de luminância.

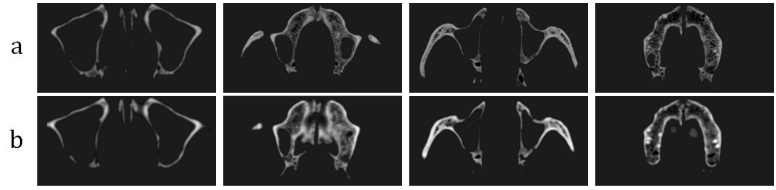


Figura 6. Ilustração de algumas das imagens já normalizadas que compõem a base de testes: (a) *Cone-Beam*. (b) *Multi-Slice*.

Para ilustrar o comportamento da entropia de Tsallis, a Tabela 1 apresenta a diferença de entropia entre imagens CBCT e MSCT, aplicando valores crescentes de q (de $0,1$ a $0,9$), expressos na Coluna 1. A Coluna 2 mostra a entropia para CBCT, a Coluna 3 mostra a entropia para MSCT, e a Coluna 4 mostra a diferença relativa correspondente.

Nota-se na Tabela 1 que quanto menor o valor de q, maior é o grau de não-extensividade da entropia, aumentando suas diferenças correspondentes. Isso ocorre porque, quanto menor o valor de q, maior é a importância dada as pequenas distribuições de probabilidade, aumentando as entradas do histograma, associadas aos valores de baixa probabilidade.

Para avaliar a robustez da entropia de Tsallis sob mudanças da relação de sinal-ruído (SNR), foi aplicado o aumento de ruído Gaussiano com

Tabela 1. Diferenças entre imagens obtidas por CBCT e por MSCT, utilizando a entropia de Tsallis.

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

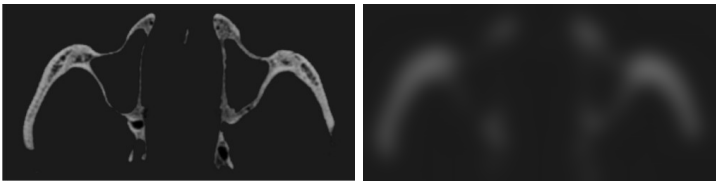


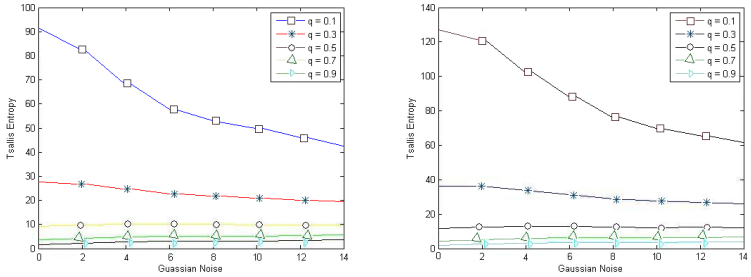
Figura 7. Ilustração comparativa utilizando imagem CBCT: (esquerda) imagem original, sem aplicação de ruído. (direita) a mesma imagem com ruído Gaussiano, sendo aplicado desvio-padrão 14.

desvio-padrão variando de 2 a 14, sobre as imagens CBCT e MSCT. Esta estratégia simula os casos em que há histogramas ”espalhados”.

A Figura 7 mostra duas imagens de CBCT, uma sem ruído (à esquerda) e outra que foi aplicado um ruído com desvio padrão 14 (à direita). Além disso, aplicamos cinco valores para q, a saber: $(0,1; 0,3; 0,5; 0,7$ e $0,9)$.

Nos gráficos da Figura 8, percebe-se que para $q < 0,5$ o valor de entropia aumenta significativamente, e para $q > 0,5$, aproximando-se de $0,9$, o valor da entropia decresce com o aumento de ruído Gaussiano. Isto sugere que, mesmo sob grande ruído Gaussiano, como na Figura 7, a entropia de Tsallis pode gerar valores observáveis.

Combinando esta informação com a Tabela 1, conclui-se que quanto maior a diferença entre duas imagens tendo valores próximos a $q = 0,5$, sugere-se um sistema não-extensivo, para as imagens utilizadas neste trabalho. Quando vemos um sistema físico como um sistema extensivo tradicional, com ruído Gaussiano grave, é completamente impossível distinguir diferenças entre eles.



(a) (b)

Figura 8. Entropia de Tsallis: (a) com incremento de ruído Gaussiano para CBCT. (b) com incremento de ruído Gaussiano para MSCT.

As diferenças destacadas pela entropias de Shannon e de Tsallis podem ser projetadas sobre suas respectivas divergências de Kullback-Leibler. Para mostrar isso, aplicamos a Equação (7) com a finalidade de medir a informação relativa entre CBCT e MSCT invertendo as imagens.

A Tabela 2 mostra a divergência de Kullback-Leibler de CBCT em relação a MSCT (Coluna 1), e de MSCT em relação a CBCT (Coluna 2), mostrando também as respectivas distâncias Euclidianas (Coluna 3), através do incremento do valor de q.

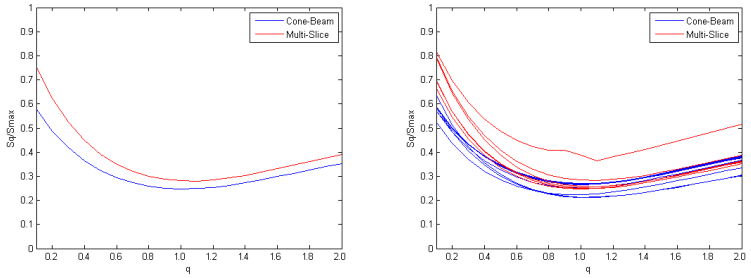
Tabela 2. A divergência de Kullback-Leibler com a inversão das imagens: a Coluna 4 mostra a distância Euclidiana entre as Colunas 2 e 3, para diferentes valores de q.

q	***K(cb:mt)***	***K(mt:cb)***	**Distância Euclidiana**
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

Pode-se observar que, na medida em que se aproxima aos sistemas extensivos tradicionais aumentando o valor de q, maior é a divergência de Kullback-Leibler. Assim, conclui-se que, quando as imagens apresentam diferenças significativas e se comportam como sistemas não-extensivos, é recomendada a utilização da divergência de Kullback-Leibler Estendida de modo a medir as suas informações de conteúdo relativo.

Para os testes sobre S_q/S_{max} ilustrados na Figura 9, nota-se que o comportamento dos valores de S_q/S_{max} é muito semelhante entre imagens

de exames *Cone-Beam* com imagens de exame de *Multi-Slice*, mesmo que as imagens de exame *Multi-Slice* apresentem valores mais altos. A Figura 9 (a) apresenta os valores de S_q/S_{max} médios de todas imagens CBCT e MSCT, enquanto a Figura 9 (b) apresenta os valores de S_q/S_{max} de cada imagem CBCT e MSCT.



(a) (b)

Figura 9. Comparação do comportamento de S_q/S_{max} entre imagem *Cone-Beam* e *Multi-Slice*: (a) valor médio de S_q/S_{max} entre as imagens da base de testes. (b) comportamento do S_q/S_{max} de cada imagem.

7. Conclusões

Este trabalho mostra a sensibilidade da entropia não-extensiva de Tsallis em um estudo comparativo entre imagens de Tomografia Computadorizada obtidas através das técnicas *Cone-Beam* (CBCT) e *Multi-Slice* (MSCT). Pela simples comparação entre histogramas é possível investigar a contribuição de cada *pixel* relacionado, ou seja, de cada tom de cinza, para cada probabilidade, em todo o sistema físico. Mas, não é possível ver claramente as diferenças entre os histogramas. No entanto, como o uso de exames pela técnica *Cone-Beam* tem se tornando popular e pode substituir as imagens em *Multi-Slice* em diversos exames, há uma crescente necessidade de investigar as suas diferenças. Assim, a distância Euclidiana simples ou a entropia relativa tradicional não podem realçar as principais diferenças, pelo menos quando os pixels importantes são aqueles relacionados com as pequenas probabilidades.

Quando se precisa investigar a contribuição de pequenas probabilidades, a divergência de Kullback-Leibler Estendida pode ser a escolha adequada. A divergência de Kullback-Leibler Estendida permite o ajuste fino do parâmetro entrópico q para combinar duas imagens através de seus histogramas de distribuição de probabilidade. Esta correspondência pode calcular a quantidade de informação relativa entre as amostras e também

permitir a sensibilidade sob ruído, até em pequenas probabilidades quando se tem algum significado exigido.

Nos resultados aqui apresentados, mostra-se que a divergência de Kullback-Leibler Estendida pode ser uma poderosa ferramenta para investigar as diferenças entre as informações extraídas em cada situação, como foi realizado na Etapa 6 da metodologia (Seção 5).

Pode-se notar que, quanto mais o parâmetro q afasta-se de 1 em direção ao 0, mais perceptível o valor da entropia se torna. Este comportamento foi menos invariável, mesmo sob a forte presença de ruído. Isto sugere que esta ferramenta de comparação é melhor usada ao considerar a imagem como um sistema não-extensível; o que significa que os estados estatísticos (ou probabilidades de luminância) podem ter interações de longo alcance espacial e temporal, mesmo para pequenas probabilidades. Isto é interessante para sistemas onde as pequenas probabilidades são semânticamente importantes.

Os resultados aqui apresentados somente reforçam aqueles encontrados na literatura para outros tipos de aplicações, necessitando entretanto uma grande investigação para provar sua eficiência em imagens médicas.

Como trabalhos futuros de comparação de imagens médicas prentende-se dividir as imagens de entrada em alguns pedaços e aplicar o método de entropia não-extensiva em cada um destes pedaços para fazer as comparações. Acredita-se que assim é possível obter resultados mais precisos identificando-se as regiões que apresentam maior semelhança e diferença.

8. Agradecimentos

Os autores agradecem ao CNPq pelo apoio financeiro (bolsa de mestrado, processo 148531/2010-5).

Referências

Albuquerque, M.P. & Esquef, I.A., Image segmentation using nonextensive relative entropy. *IEEE Latin America Transactions*, 6(5):477–483, 2008.

Albuquerque, M.P.; Esquef, I.A.; Mello, A.R.G. & Albuquerque, M.P., Image thresholding using Tsallis entropy. *Pattern Recognition Letters*, 25(9):1059 –1065, 2004.

Borland, L.; Plastino, A.R. & Tsallis, C., Information gain within nonextensive thermostatistics. *Journal of Mathematical Physics*, 39(12):6490–6501, 1998.

Erdmann, H.E.R., *Detecção e Classificação de Imagens Sintéticas Utilizando Entropia Não-Extensiva*. Dissertação de mestrado em engenharia elétrica, Centro Universitario da FEI, Sao Bernardo do Campo, SP, 2009.

Esquef, I.A., *Técnicas de Entropia em Processamento de Imagens*. Dissertação de mestrado em instrumentação científica, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, RJ, 2002.

Giraldi, G.A.; Rodrigues, P.S.; Kitani, E.C.; Sato, J.R. & Thomaz, C.E., Statistical learning approaches for discriminant features selection. *Journal of Brazilian Computer Society*, 14(2):7–22, 2008.

Giraldi, G.A.; Rodrigues, P.S. & Suri, J., Implicit dual snakes for medical imaging. In: *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. New York, USA, v. 1, p. 3025–3028, 2006.

Hounsfield, G.N., Computerized transverse axial scanning (tomography): Part I. Description of system. *British Journal of Radiology*, 46:1016–1022, 1973.

Kapur, J.N.; Sahoo, P.K. & Wong, A.K.C., A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3):273–285, 1985.

Kullback, S. & Leibler, R., On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

Lessa, V.S., *Classificação de Imagens de Ultrassom de Câncer de Mama Baseada em Informações Híbridas Utilizando Teoria da Informação*. Dissertação de mestrado em engenharia elétrica, Centro Universitario da FEI, Sao Bernardo do Campo, SP, 2010.

Loubele, M.; Bogaerts, R.; Van Dijck, E.; Pauwels, R.; Vanheusden, S.; Suetens, P.; Marchal, G.; Sanderink, G. & Jacobs, R., Comparison between effective radiation dose of CBCT and MSCT scanners for dentomaxillofacial applications. *European Journal of Radiology*, 71(3):461 – 468, 2009.

Loubele, M.; Guerero, M.E.; Jacobs, R.; Suetes, P. & van Steenberghe, D., A comparison of jaw dimensional and quality assessments of bone characteristics with cone-beam CT, spiral tomography, and multi-slice spiral CT. *International Journal of Oral and Maxillofacial Implants*, 22(3):446 – 454, 2007.

Maintz, J. & Viergever, M., A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

Olívio, F.C., *Um Modelo Bayesiano com Divergência de Kulback-Leibler Estendida para Reconhecimento de Objetos 3D Baseados em Múltiplas Visões*. Dissertação de mestrado em engenharia elétrica, Centro Universitario da FEI, Sao Bernardo do Campo, SP, 2009.

Pun, T., Entropic thresholding: A new approach. *Computer Graphics and Image Processing*, 16(3):210–239, 1981.

Rodrigues, P.S.; Chang, R.F. & Suri, J.S., Non-extensive entropy for CAD systems of breast cancer images. In: *Proceedings of the XIX Brazilian Symposium on Computer Graphics and Image Processing*. Los Alamitos, USA: IEEE Computer Society, p. 121–128, 2006a.

Rodrigues, P.S. & Giraldi, G.A., Computing the q-index for Tsallis nonextensive image segmentation. In: *Proceedings of XXII Brazilian Symposium on Computer Graphics and Image Processing*. Los Alamitos, USA: IEEE Computer Society, p. 232–237, 2009.

Rodrigues, P.S. & Giraldi, G.A., Improving the non-extensive medical image segmentation based on Tsallis entropy. *Pattern Analysis & Applications*, 14(4):369–379, 2011.

Rodrigues, P.S.S.; Giraldi, G.A.; Provenzano, M.; Faria, M.D.d.; Chang, R.F. & Suri, J., A new methodology based on q-entropy for breast lesion classification in 3-D ultrasound images. In: *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway, USA: IEEE Press, p. 1048–1051, 2006b.

Shannon, C., A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

Sobiecki, A.; Gallão, C.; Cosme, D. & Rodrigues, P., The power of Kullback-Leibler divergence extendable for Tsallis entropy in comparing cone-beam and multi-slice tomography images. In: *Anais do VII Workshop de Visão Computacional*. Curitiba, PR, p. 296–301, 2011.

Tavares, A.H.M.P., *Aspectos Matemáticos da Entropia*. Dissertação de mestrado em matemática, Departamento de Matemática, Universidade de Aveiro, Aveiro, Portugal, 2003.

Tsallis, C., Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52(1/2):479–487, 1988.

Tsallis, C., Nonextensive statistical mechanics and thermodynamics: Historical background and present status. In: Abe, S. & Okamoto, Y. (Eds.), *Nonextensive Statistical Mechanics and its Applications*. Berlin, Germany: Springer, Lecture Notes in Physics, p. 3–98, 2001.

Notas Biográficas

André Sobiecki é graduado em Tecnologia de Sistemas de Informação (UDESC, 2009) e mestre em Engenharia Elétrica (Centro Universitário da FEI com estágio de 3 meses no LNCC). Atualmente é doutorando em Ciência da Computação (Departamento de Visualização Científica e Computação Gráfica, Universidade de Groningen, Holanda, desde maio de 2012). Tem interesse de pesquisa na área de processamento de imagens, reconhecimento de padrões, restauração digital e métodos de esqueletização 2D e 3D.

Celso Denis Gallão é bacharel em Matemática com ênfase em Análise de Sistemas (Centro Universitário Fundação Santo André – FSA, 1990), especialista em Gestão da Qualidade (Centro Universitário de Santo André - UNIA, 1998) e mestre em Engenharia Elétrica (Centro Universitário da FEI, 2012). Atualmente é doutorando em processamento de sinais (Centro Universitário da FEI, desde junho de 2012) e participante do grupo de inteligência artificial aplicada à automação na mesma instituição, com interesses na área da visão computacional. Desde agosto de 2012 é professor de Inteligência Artificial e Engenharia de Software da Faculdade de Tecnologia do Estado de São Paulo (FATEC - São Caetano do Sul, SP); desde 2011 é professor de Inteligência Artificial da Faculdade Anhanguera (São Caetano do Sul, SP); desde 1994 é professor de computação gráfica e desenvolvimento *web* do curso Técnico em Informática do Colégio Singular (Santo André, SP), onde é coordenador desde 2001.

Daniel Cosme Cardoso é graduado em Ciência da Computação (Centro Universitário da FEI, 2009) e atualmente é mestrando em Engenharia Elétrica com enfase em Inteligência Artificial, na mesma instituição. Trabalha no departamento de pesquisa e desenvolvimento de software da Micromar como coordenador de projetos. Possui interesse nos temas de visão computacional e processamento digital de imagens.

Paulo Sérgio Silva Rodrigues é graduado em Ciência da Computação (Universidade Federal do Pará, 1996), mestre e doutor em Ciência da Computação (Universidade Federal de Minas Gerais, 1999 e 2003, respectivamente, com estágio na Univertità Degli Studi di Ancona, Itália). Entre 2003 e 2006 fez pós-doutorado no Laboratório Nacional de Computação Científica (LNCC). Suas principais áreas de interesse são: visão computacional, processamento de imagens, realidade aumentada e reconhecimento de padrões e tem a área médica como um dos principais alvos dos resultados de seus trabalhos. Em 2005-2006 publicou vários trabalhos na área de análise de imagens de câncer de mama e atualmente vem desenvolvendo técnicas para reconstrução de próteses cranio facial. Desde 2007 é professor do departamento de Ciência da Computação do Centro Universitário da FEI (São Bernardo do Campo, SP) e membro do grupo de Inteligência Artificial da mesma Instituição. É professor do mestrado em Engenharia Elétrica ministrando as disciplinas de Visão Computacional e Geometria Computacional.

Capítulo 4

Classificação e Extração de Características Discriminantes de Imagens 2D de Ultrassonografia Mamária

Albert da Costa Xavier*, João Ricardo Sato, Gilson Antônio Giraldi, Paulo Sérgio Rodrigues e Carlos Eduardo Thomaz

Resumo: Os estudos relacionados à detecção e ao tratamento de cânceres por meio de imagens, incluindo o câncer de mama, vêm se aperfeiçoando ano após ano. Este capítulo aborda a análise discriminante de tumores mamários em imagens 2D ultrassonográficas. No processo de análise destas imagens, métodos estatísticos univariados e multivariados são investigados com o objetivo de extrair informações discriminantes para fins de classificação. Em caráter complementar, este tipo de metodologia evidencia também as diferenças estatísticas mais significativas entre os tumores, indicando no espaço original das imagens os casos mais simples e difíceis de classificação.

Palavras-chave: Imagem de ultrassom da mama, Extração estatística de características discriminantes.

Abstract: *Research on detection and treatment of cancer using images, including breast cancer, has improved in the last years. This work addresses the problem of computer-aided breast cancer diagnosis in 2D ultrasound images by carrying out a statistical discriminant analysis of mammary tumors. To analyze the images, univariate and multivariate statistical methods have been explored with the aim of extracting discriminant information for classification purposes. 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.*

Keywords: *Breast ultrasound image, Statistical extraction of discriminant features.*

*Autor para contato: albert.xavier@gmail.com

Neves et al. (Eds.), Avanços em Visão Computacional (2012) DOI: 10.7436/2012.avc.4 ISBN 978-85-64619-09-8

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 mesmo (INCA, 2009).

Neste 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 (Kopans, 2000).

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.

O capítulo está dividido em 5 seções. A próxima seção, Seção 2, contextualiza o problema abordando assuntos relacionados à patologia da mama e ao diagnóstico do câncer de mama com imagens ultrassonográficas. A Seção 3 apresenta a base teórica que apoia o desenvolvimento do capítulo, descrevendo os métodos estatísticos univariado e multivariado. Os experimentos realizados e os dados das imagens são apresentados na Seção 4 por meio da descrição dos recursos utilizados e das técnicas adotadas na análise das imagens ultrassonográficas de tumores mamários. A Seção 5 contempla a análise dos resultados e, por fim, a Seção 6 conclui este capítulo resumindo e discutindo os principais resultados obtidos.

2. Contextualização do Problema

O câncer de mama é a principal causa mundial de morte não previsível por câncer. A estimativa para novos casos é preocupante. No Brasil, segundo o INCA (Instituto Nacional de Câncer[1]), a previsão para 2010 foi de 49240 novos casos representando aproximadamente 25,5% de incidência perante todas as ocorrências de câncer, estimadas em 192.590. Na Figura 1 pode-se observar as estimativas para o ano de 2010 de número de casos novos por câncer segundo localização primária. Ainda segundo o INCA, 33 mulheres morrem por dia no Brasil em decorrência do câncer de mama.

[1] http://www.inca.gov.br/estimativa/2010/estimativa20091201.pdf

Localização Primária	Estimativa de Casos Novos		
Neoplasia Maligna	Masculino	Feminino	Total
Próstata	52.350	-	52.350
Mama Feminina	-	49.240	49.240
Traqueia, Brônquio e Pulmão	17.800	9.830	27.630
Cólon e Reto	13.310	14.800	28.110
Estômago	13.820	7.680	21.500
Colo do Útero	-	18.430	18.430
Cavidade Oral	10.330	3.790	14.120
Esôfago	7.890	2.740	10.630
Leucemias	5.240	4.340	9.580
Pele Melanoma	2.960	2.970	5.930
Outras Localizações	59.130	78.770	137.900
Subtotal	**182.830**	**192.590**	**375.420**
Pele não Melanoma	53.410	60.440	113.850
Todas as Neoplasias	**236.240**	**253.030**	**489.270**

*Números arredondados para 10 ou múltiplos de 10

Figura 1. Distribuição das estimativas de câncer de mama no Brasil para o ano de 2010 (INCA, 2009, p 42).

Em países desenvolvidos como, por exemplo, nos Estados Unidos, mais de 180000 mulheres são diagnosticadas com câncer de mama invasivo anualmente. Em complemento, são detectados aproximadamente 25000 novos casos de carcinoma ductal e lobular *in situ* (ACS, 2010). Notadamente os 25000 casos de carcinoma representaram percentual significante perante a totalidade de casos diagnosticados de cânceres de mama invasivos. Segundo o INCA, a denominação "carcinoma" é atribuída ao câncer que tem início em tecidos epiteliais como pele ou mucosas. O carcinoma mamário caracteriza-se pela proliferação anormal das células do ducto e do lóbulo.

Nenhum avanço terapêutico pode ser comparado, em termos de benefícios globais, à contribuição que o diagnóstico precoce, por meio dos métodos radiológicos, pode trazer para pacientes portadores do câncer de mama (Kopans, 2000).

Atualmente, não existe nenhuma forma de prevenção da doença. Os dados mostram que o melhor resultado obtido, mesmo distante do ideal, é a detecção e tratamento precoce. Tais medidas implicam diretamente no retardo da morte e na melhor qualidade de vida da paciente que poderá ser poupada do trauma físico e psicológico da mastectomia. Definitivamente, o rastreamento mamográfico não é a solução para o problema do câncer de mama, mas é a melhor solução disponível no presente e no futuro próximo. Historicamente, a detecção precoce proporcionou o uso mais adequado da terapia conservadora com predomínio da excisão e da utilização da radiação.

Como as razões pelas quais a mulher desenvolve o câncer ainda são desconhecidas e obscuras, o entendimento da epidemiologia e da etiologia desta malignidade é o alicerce para qualquer estudo deste assunto. Mas, diante destes dados e considerações brevemente descritos aqui, as condições econômicas do sistema de saúde pública podem influenciar as possíveis escolhas de diagnóstico e tratamento desta doença.

2.1 Fatores de risco e prevenção

Embora sem comprovação científica, investigações indicam que a incidência do câncer tem relação direta com alguns fatores. Dentre estes pode-se destacar a idade da mulher, a primeira gravidez, associada com o período de lactação, o uso de anticoncepcionais e a reposição hormonal. Infelizmente, trata-se apenas de suposições pois até o presente momento não existe nenhuma comprovação concreta que justifique o aparecimento e desenvolvimento das lesões. Mas, independentemente das afirmações anteriores, toda mulher apresenta grau de risco. Em número muito menor também são diagnosticados cânceres de mama em indivíduos do sexo masculino.

A hereditariedade do câncer de mama é outro fator amplamente discutido mas também ainda pouco consolidado. Algumas mulheres herdam um gene anormal e apresentam risco muito elevado de desenvolvimento do câncer. Mesmo sem comprovação do resultado, recomenda-se iniciar o rastreamento 10 anos antes da idade que o familiar foi diagnosticado como portador do câncer de mama (Robson & Offit, 2002; Cohn et al., 2003).

Ainda sem explicação também, Dupont & Page (1985) identificaram que a associação da presença de um cisto com o histórico de câncer familiar provoca um aumento da ordem de 2 a 3 vezes na possibilidade de desenvolvimento da doença. Outro agravante é o próprio processo de detecção utilizando raios-X, que pode aumentar o risco de desenvolvimento da doença. É sabido que a energia da radiação produz radicais livres (partículas carregadas) que podem reagir e causar danos ao DNA. Ainda sem os devidos cuidados, a exposição à radiação pode afetar também o próprio técnico radiologista que executa o rastreamento.

2.2 Diagnóstico do câncer de mama por ultrassom

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. Trata-se de um método baseado na reflexão do som. Os transdutores são dispositivos existentes no aparelho capazes de transmitir e receber ondas sonoras. As ondas produzidas alcançam os tecidos e, de acordo com os tipos dos mesmos, são refletidas de volta ao transdutor para ampliação em monitor. Este método é comumente chamado de ecografia pelo fato da reflexão do som também ser chamada de “eco”.

Para lesões de tamanho maior ou igual a 1(um) cm a utilização do ultrassom é fortemente indicada para diferenciar um cisto de uma possível lesão sólida (Kopans, 2000; Shah et al., 2010).

3. Metodologia

Para analisar as alterações morfológicas estatisticamente significantes e classificar imagens 2D ultrassonográficas de mama que contêm tumores malignos e benignos, duas abordagens estatísticas são consideradas: massivamente univariada e multivariada. Ambas são descritas em detalhes nas subseções seguintes.

3.1 Análise estatística massivamente univariada

A análise estatística univariada de imagens é frequentemente baseada em um modelo linear geral (*General Linear Model - GLM*), que reune vários modelos estatísticos diferentes. Dentre outros, pode-se citar a regressão linear, o teste t, a análise da variância (ANOVA) e a análise da covariância (ANCOVA).

O teste de hipóteses, aqui detalhado, foi adotado para identificação das diferenças visuais entre os grupos de imagens de tumores benignos e malignos. Usando o teste t como teste de significância, tem-se que esta diferença é dada basicamente pelas diferenças entre as médias de cada grupo ponderada por um desvio padrão do espalhamento das amostras assumindo igualdade de variâncias entre as amostras, ou seja:

$$t_k = \frac{\overline{\mathrm{x}}_{1k} - \overline{\mathrm{x}}_{2k}}{\sigma_k \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}, \tag{1}$$

onde t_k é o t-valor da variável k, $\overline{\mathrm{x}}_{1k}$ e $\overline{\mathrm{x}}_{2k}$ são as médias amostrais respectivas da variável k do grupo 1 e do grupo 2, σ_k é o desvio padrão ponderado da variável k, o N_1 e N_2 são respectivamente o total de amostras do grupo 1 e do grupo 2. O desvio padrão ponderado do conjunto de amostras é dado pela equação:

$$\sigma_k = \sqrt{\frac{(N_1 - 1)(\sigma_{1k})^2 + (N_2 - 1)(\sigma_{2k})^2}{N_1 + N_2 - 2}}, \tag{2}$$

onde σ_{1k} e σ_{2k} são respectivamente o desvio padrão da variável k para os grupos 1 e 2.

Trata-se de um teste que visa a identificação de diferenças entre os grupos para um nível de significância fixado. Neste teste, são verificadas as diferenças de médias pixel a pixel (Davatzikos, 2004). O teste de hipótese estatística fornece uma afirmação acerca dos parâmetros de uma ou mais populações (testes paramétricos) ou acerca da distribuição da população

(Magalhães & Lima, 2010). Calculando o mapa de *t-valores* para os grupos analisados, ou seja, os *t-valores* de cada um dos pixels do conjunto amostral, o conceito de hipótese nula (H_0) e hipótese alternativa (H_1) pode ser utilizado para definir as variáveis que apresentam diferenças significativas. As hipóteses H_0 e H_1 podem ser descritas matematicamente pelas seguintes equações:

$$H_0 : \textit{Média do grupo 1} = \textit{Média do grupo 2}$$
$$H_1 : \textit{Média do grupo 1} \neq \textit{Média do grupo 2}$$

Entretanto, podem ocorrer dois tipos de erros no teste de hipóteses: erro tipo I e tipo II. No erro do tipo I, rejeita-se a H_0 quando esta é verdadeira, ou seja, afirma-se que existe diferença estatisticamente significativa quando, na verdade, não existe. No erro do tipo II não rejeita-se H_0 quando esta é falsa, ou seja, afirma-se que não existe diferença estatisticamente significativa quando na verdade existe (Magalhães & Lima, 2010). Neste teste determina-se um valor de significância p que indique que a diferença é estatisticamente significante com um determinado nível de significância.

Por meio da tabela t de Student pode-se obter o *t-valor* crítico correspondente a um nível de significância desejado orientado pelos graus de liberdade do conjunto de amostras (Magalhães & Lima, 2010). O grau de liberdade é obtido pela diferença entre a quantidade total de amostras e a quantidade de grupos analisados. O teste consiste em verificar se o *t-valor* calculado, em módulo, é superior ao *t-valor* observado na tabela t de Student. Se o *t-valor* calculado em módulo for maior que o *t-valor* crítico da tabela, então a hipótese nula é rejeitada e esta diferença encontrada é considerada relevante do ponto de vista estatístico (Leão et al., 2009).

3.2 Análise estatística multivariada

Técnicas em estatísticas multivariadas como LDA (*Linear Discriminant Analysis*) e SVM (*Support Vector Machine*) podem ser utilizadas na etapa de classificação (Fisher, 1936; Giraldi et al., 2008; Hastie et al., 2009; Sato et al., 2009; Vapnik, 1998). 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 simultaneamente e não pixel a pixel como na abordagem anterior.

3.2.1 LDA

A proposta do método LDA, também conhecido como método de Fisher (Fisher, 1936), é encontrar o hiperplano de maior separação entre os grupos analisados. O cálculo deste 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 (Fisher, 1936). 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:

$$S_b = \sum_{i=1}^{g} N_i(\overline{\mathrm{x}}_i - \overline{\mathrm{x}})(\overline{\mathrm{x}}_i - \overline{\mathrm{x}})^T, \tag{3}$$

$$S_w = \sum_{i=1}^{g} \sum_{j=1}^{N_i} (\mathrm{x}_{i,j} - \overline{\mathrm{x}}_i)(\mathrm{x}_{i,j} - \overline{\mathrm{x}}_i)^T, \tag{4}$$

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. A relação proposta por Fisher que deve ser maximizada é dada pela seguinte equação (5):

$$P_{LDA} = \arg\max \frac{|S_b|}{|S_w|}. \tag{5}$$

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. Este critério de Fisher descrito pela Equação 5 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, esta razão somente pode ser calculada se a matriz de espalhamento intra-classe S_w for não-singular. Exatamente neste 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 (Fukunaga, 1990). A quantidade de amostras necessárias para evitar esta instabilidade no cálculo da matriz inversa de S_w deve ser igual ou superior a 5 vezes a quantidade de variáveis destas (Jain & Chandrasekaran, 1982). Considerando o cenário do trabalho desenvolvido aqui, certamente ocorreria este 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, utiliza-se o método denominado MLDA (*Maximum uncertainty Linear Discriminant Analysis*) (Thomaz et al., 2006). Esta 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 (5) por S_w^* e regulariza-se o critério de Fisher para problemas onde $N \ll n$.

3.2.2 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 analoga 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 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 (Lorena & Carvalho, 2007). Este método é baseado na teoria de aprendizado estatístico, pioneiramente descrita por Vapnik (1998). 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:

$$w_{SVM} = \sum_{j=1}^{N} \alpha_j y_j \mathrm{x}_j, \tag{6}$$

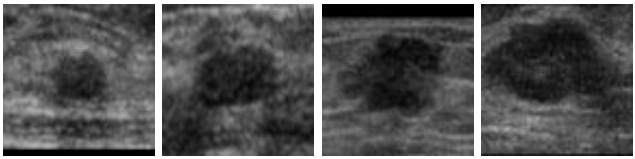
onde α_j são os coeficientes de Lagrange não-negativos obtidos pela solução de um problema de otimização quadrático com restrições de desigualdade

linear (Lorena & Carvalho, 2007). As observações de treinamento x_j, com α_j não-zero, ficam na fronteira da margem e são chamadas de vetores de suporte (Sato et al., 2009). 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.

4. Experimentos

4.1 Banco de imagens e características geométricas e de textura

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 2 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 2. 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 desta 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 (Rodrigues et al., 2006a). Os valores de pré-processamento das características foram atribuídos por radiologistas e calculados em estudos anteriores (Rodrigues et al., 2006b; Giraldi et al., 2008).

4.2 Pré-processamento das imagens

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. A intensidade de cada pixel da imagem é definida 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 maioria das imagens foram 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 destas 20 imagens, foi de 13 pixels. Nas demais imagens redimensionadas o maior ajuste foi de 5 pixels.

Após este pré-processamento inicial, as imagens foram submetidas a um processo de redução de dimensionalidade utilizando a técnica PCA (*Principal Component Analysis*) (Fukunaga, 1990), porém preservando todas as componentes com autovalores não-nulos (Thomaz et al., 2007). 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.

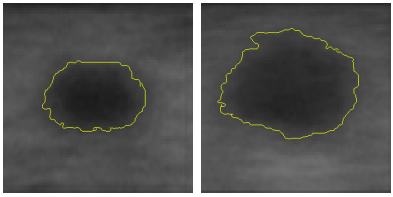
O cálculo da acurácia dos classificadores foi baseado na abordagem de validação cruzada (*cross-validation*). Nos experimentos aqui realizados, foi

adotada a forma extrema da abordagem de validação cruzada denominada *leave-one-out* (Fukunaga, 1990). Este método foi aplicado em decorrência do pequeno número de amostras rotuladas.

5. Resultados

5.1 Análise estatística massivamente univariada

A Figura 3 apresenta as imagens médias das amostras dos grupos benigno e maligno[2]. Claramente observa-se uma diferença no formato e tamanho dos tumores nas imagens. Na imagem média das lesões benignas, vista na Figura 3(a), o tumor é menor e mais concentrado. Já o tumor das imagens de lesões malignas, visto na Figura 3(b), parece ser maior e mais espalhado.



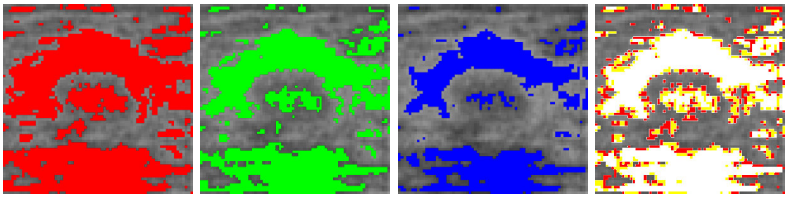
(a) (b)

Figura 3. (a) Imagem média de tumores benignos; (b) Imagem média de tumores malignos.

Em complemento às diferenças observadas entre as imagens médias de tumores benignos e malignos, foi implementado o teste de hipóteses baseado na distribuição de probabilidade de t de Student. Nesta análise, a finalidade foi obter as variações mais relevantes estatisticamente entre as imagens dos grupos de tumores malignos e benignos. Nas Figuras 4(a), 4(b) e 4(c) pode-se observar as regiões mais discriminantes selecionadas de acordo com níveis de significância de 5%, 1% e 0,1% da tabela t de Student. Do ponto de vista estatístico, estas são as regiões de maior discriminância entre os dois grupos de imagens de tumor mamário, avaliadas pixel a pixel. Estas regiões mais discriminantes foram projetadas sobre uma imagem de referência que representa aqui a subtração das imagens médias dos dois grupos. A Figura 4(d) demonstra a projeção simultânea das três imagens relacionadas nas Figuras 4(a), 4(b) e 4(c) representando os níveis de significância de 5%, 1% e 0,1% da tabela t de Student. A cor branca é resultante da intersecção das cores vermelho, verde e azul e representa a projeção simultânea dos três níveis de significância adotados (5%, 1% e 0,1%). A cor amarela é resultante da intersecção das cores vermelho e verde e representa a projeção

[2] Devido a baixa qualidade das imagens ultrassonográficas as bordas do tumores foram destacadas utilizando o filtro de Canny.

simultânea das regiões discriminantes com níveis de significância de 5% e 1% .



(a) (b) (c) (d)

Figura 4. Projeção das regiões discriminantes: (a) Nível de significância de 5%; (b) Nível de significância de 1%; (c) Nível de significância de 0,1%; (d) Projeção simultânea de regiões discriminantes representando todos os níveis de significância experimentados (5%, 1% e 0,1%).

5.2 Análise estatística multivariada

Os resultados da análise multivariada são apresentados em dois passos. Primeiramente as imagens e seus valores de pré-processamento (referentes às características circularidade, heterogeneidade e sombra acústica) foram submetidos aos classificadores lineares para obter os hiperplanos de separação e também avaliar o desempenho dos mesmos. O segundo passo fez uso de um mapa de pesos discriminantes, gerado no processo de classificação, para identificar e mostrar as regiões mais discriminantes das imagens de tumor mamário analisadas.

5.2.1 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 destes mesmos classificadores analisando as características extraídas das imagens. Estes resultados indicam que os fatores determinantes na classificação dos grupos vão além de características como circularidade, heterogeneidade e sombra acústica, principalmente para separações lineares.

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 por meio de testes de validação cruzada extrema (*leave-one-out*). 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.

Tabela 1. Taxas de classificação.

	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%

A Figura 5 apresenta, de maneira simultânea, os hiperplanos obtidos dos dois classificadores com o intuito de observar que a ordem de classificação das amostras, no entanto, ficou diferente entre os mesmos. 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 válido 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.

5.2.2 Transição visual dos tumores classificados

A Figura 6(a) apresenta uma padronização genérica 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 de forma gráfica. A diferença no formato dos nódulos é bastante expressiva. Em âmbito geral, trata-se de um bom guia visual para o estudo dos tumores mamográficos em imagens ultrassonográficas.

De forma análoga a Figura 6(b) ilustra, em modo negativo, a transição das imagens entre os grupos de tumores benignos e malignos obtida pelos classificadores MLDA e SVM. Esta transição visual foi construída com as imagens médias calculadas a partir de intervalos estabelecidos na ordenação

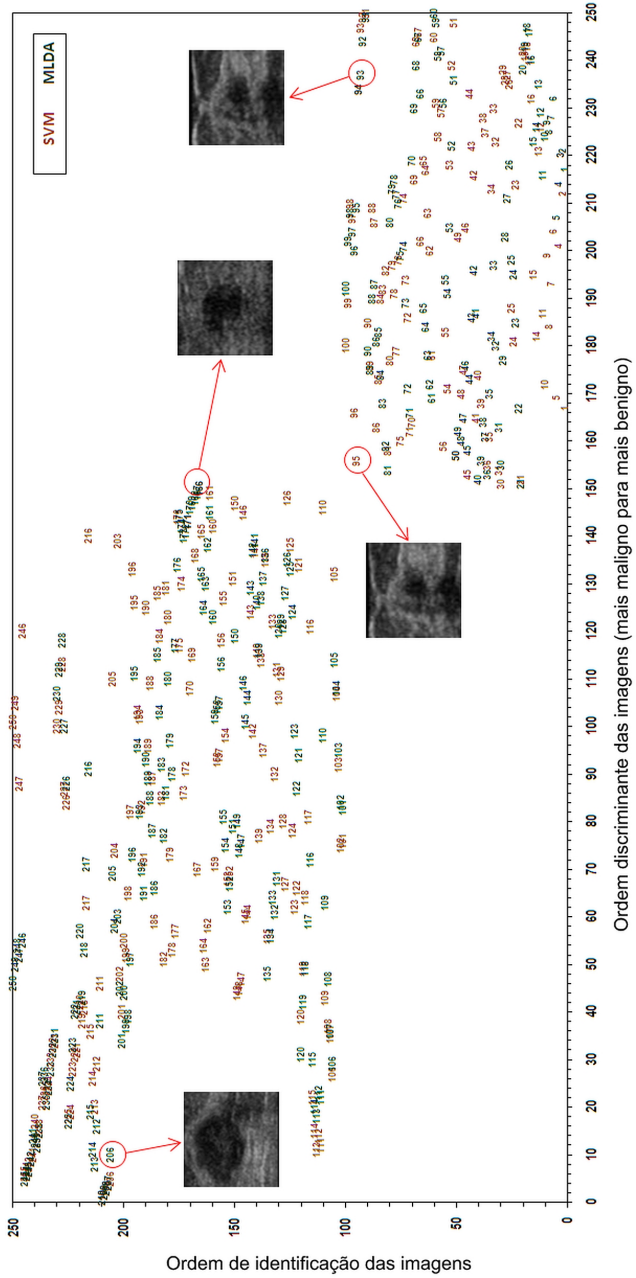
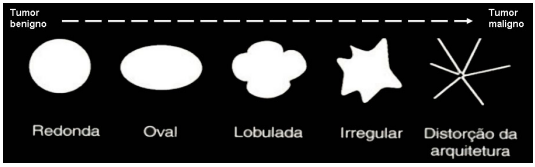


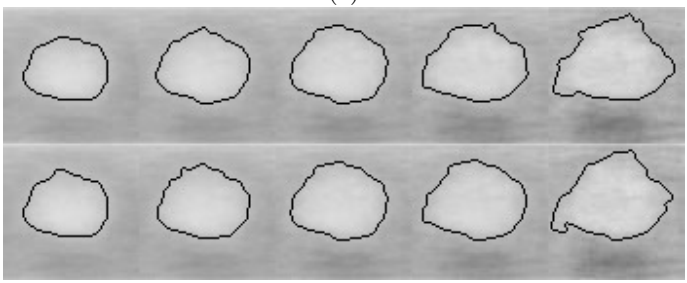
Figura 5. Hiperplanos MLDA e SVM de separação: maligno a esquerda e benigno a direita.

discriminante das imagens. O intervalo adotado foi de cinquenta imagens, ou seja, a cada 50 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". 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).

Nota-se ainda uma boa semelhança nas sequências de imagens geradas pelos classificadores SVM e MLDA. Apesar da ordem das imagens não ser igual, em ambos resultados, o tamanho do tumor aumenta na navegação em direção aos tumores malignos e a possível sombra acústica de cor preta, na parte inferior da imagem, diminui de tamanho na navegação em direção aos tumores benignos (esquerda).



(a)



(b)

Figura 6. (a) Formato BI-RADS dos tumores mamários; (b) Transição visual dos tumores ordenados pelos classificadores MLDA (linha superior) e SVM (linha inferior), partindo dos "mais benignos" (a esquerda) para os "mais malignos" (a direita).

6. Conclusão

Este trabalho realizou comparações quantitativas e qualitativas entre os resultados das aplicações dos métodos estatísticos univariado e multivariado na análise de imagens ultrassonográficas de mama. O sucesso da aplicação destes métodos estatísticos é fortemente dependente das fases de pré-processamento e segmentação das imagens. Os resultados obtidos na análise multivariada considerando a intensidade de todos os pixels da imagem simultaneamente, por meio dos classificadores SVM e MLDA, alcançaram desempenho superior perante a mesma análise utilizando características extraídas das imagens como circularidade e sombra acústica. Enquanto que a classificação linear utilizando a imagem como um todo alcançou a taxa de acerto absoluta em ambos classificadores, obteve-se a maior taxa de acerto de 78% analisando a característica circularidade pelo classificador MLDA.

É importante ressaltar, no entanto, que todas as imagens analisadas apresentavam tumores e que estes estavam posicionados na região central das mesmas. Em outras palavras, estas imagens foram normalizadas espacialmente com o intuito de evidenciar a região do tumor e eliminar regiões de fundo da imagem. A obtenção deste alto índice de acurácia foi certamente favorecida por estes ajustes feitos na etapa de pré-processamento das imagens. Na literatura o índice médio de acurácia dos classificadores é da ordem de 90%. Talvez a aplicação dos mesmos métodos em imagens de ultrassom sem pré-processamento possa apresentar taxas de acertos menores e mais próximas desta, mas a análise estatística da imagem como um todo normalizada espacialmente apresentou resultado de classificação extremamente promissor.

Adicionalmente, a análise univariada indicou as regiões central e inferior da imagem como mais diferentes estatisticamente. Na transição entre os grupos de tumores benignos e malignos, foi possível notar as diferenças entre os grupos justamente nestas regiões indicadas pela análise univariada. Muito provavelmente as características de circularidade e sombra acústica exerceram influência na discriminância 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. Na verdade, estas diferenças estão espalhadas por toda a área da imagem. Pode-se concluir que as diferenças entre os grupos não são limitadas à observação de determinadas regiões ou características exclusivas. As diferenças são difusas e complexas, e, portanto, outras partes da imagem devem ser observadas para alcançar um bom resultado na identificação correta de cada classe ou grupo de tumores.

Acredita-se que os resultados promissores alcançados na análise das imagens como um todo, tanto para classificadores não-paramétricos e paramétricos como o SVM e o MLDA, podem ser de grande valia no trei-

namento e suporte para radiologistas. No caso de radiologistas iniciantes, este tipo de análise possibilitaria a confirmação de determinada classificação do ponto de vista estatístico, principalmente em casos mais simples. Já no caso de radiologistas experientes, os resultados dos classificadores poderiam servir de apoio na tomada de decisões, principalmente nos casos mais complexos e duvidosos. Por exemplo, nos casos mais duvidosos, o radiologista poderia verificar o resultado do classificador antes de solicitar um novo tipo de exame ou mesmo determinar um tipo de tumor. A natureza não invasiva, o baixo custo e a formação da imagem em tempo real fazem da imaginologia ultrassonográfica, aliada a análise estatística de imagem, uma ferramenta extremamente útil no diagnóstico médico.

Referências

ACS, , Breast Cancer: Cancer Facts and Figures. American Cancer Society, 2010. Http://www.cancer.org/Research/CancerFactsFigures/BreastCancerFactsFigures/breast-cancer-facts–figures-2009-2010.

Cohn, W.F.; Ropka, M.E.; Jones, S.M. & Miesfeldt, S., Information needs about hereditary breast cancer among women with early-onset breast cancer. *Cancer Detection and Prevention*, 27(5):345–352, 2003.

Davatzikos, C., Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *Neuroimage*, 23(1):17–20, 2004.

Dupont, W.D. & Page, D.L., Risk factors for breast cancer in women with proliferative breast disease. *New England Journal of Medicine*, 312(3):146–151, 1985.

Fisher, R.A., The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(7):179–188, 1936.

Fukunaga, K., *Introduction to Statistical Pattern Recognition*. 2a edição. London, UK: Academic Press, 1990.

Giraldi, G.A.; Rodrigues, P.S.; Kitani, E.C.; Sato, J.R. & Thomaz, C.E., Statistical learning approaches for discriminant features selection. *Journal of the Brazilian Computer Society*, 14(2):7–22, 2008.

Hastie, T.; Tibshirani, R. & Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2a edição. New York, USA: Springer, 2009.

INCA, , Estimativa 2010: incidência de câncer no Brasil. Instituto Nacional de Câncer, 2009. Http://www.inca.gov.br/estimativa/2010/.

Jain, A.K. & Chandrasekaran, B., Dimensionality and sample size considerations in pattern recognition practice. In: Krishnaiah, P.R. & Kanal, L.N. (Eds.), *Handbook of Statistics*. Amsterdam, The Netherlands: North-Holland, v. 2, p. 835–855, 1982.

Kopans, D.B., *Imagem da Mama*. 2a edição. São Paulo, SP: Medsi, 2000.

Leão, R.D.; Sato, J.R. & Thomaz, C.E., *Comparação entre as análises estatísticas univariada e multivariada para extração de informação discriminante em imagens de ressonância magnética do cérebro humano*. Relatório Técnico 01/2009, Departamento de Engenharia Elétrica, FEI, São Bernardo do Campo, SP, 2009.

Lorena, A.C. & Carvalho, A.C.P.L.F., Uma introdução às *Support Vector Machines*. *Revista de Informática Teórica e Aplicada*, 14(2):43–67, 2007.

Magalhães, M.N. & Lima, A.C.P., *Noções de Probabilidade e Estatística*. 6a edição. São Paulo, SP: EDUSP, 2010.

Robson, M.E. & Offit, K., Considerations in genetic counseling for inherited breast cancer predisposition. *Seminars in Radiation Oncology*, 12(4):362–370, 2002.

Rodrigues, P.S.; Giraldi, G.A.; Chang, R.F. & Suri, J.S., Automatic classification of breast lesions in 3-D ultrasound images. In: Suri, J.S.; Kathuria, C.; Molinari, F. & Fenster, A. (Eds.), *Advances in Diagnostic and Therapeutic Ultrasound Imaging*. Norwood, USA: Artech House Publishers, p. 189–224, 2006a.

Rodrigues, P.S.; Giraldi, G.A.; Chang, R.F. & Suri, J.S., Non-extensive entropy for CAD systems of breast cancer images. In: *Proceedings of the 19th Brazilian Symposium on Computer Graphics and Image Processing*. Los Alamitos, USA: IEEE Computer Society, v. 3, p. 121–128, 2006b.

Sato, J.R.; Fujita, A.; Thomaz, C.E.; da Graça Morais Martin, M.; Mourão-Miranda, J.; Brammer, M.J. & Amaro Junior, E., Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. *NeuroImage*, 46(1):105–114, 2009.

Shah, B.A.; Fundaro, G.M. & Mandava, S., *Breast Imaging Review: A Quick Guide to Essential Diagnoses*. New York, USA: Springer, 2010.

Thomaz, C.E.; Boardman, J.P.; Counsell, S.; Hill, D.L.G.; Hajnal, J.V.; Edwards, A.D.; Rutherford, M.A.; Gillies, D.F. & Rueckert, D., A multivariate statistical analysis of the developing human brain in preterm infants. *Image and Vision Computing*, 25(6):981–994, 2007.

Thomaz, C.E.; Kitani, E.C. & Gillies, D.F., 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, 2006.

Vapnik, V.N., *Statistical Learning Theory*. New York, USA: John Wiley & Sons, 1998.

Notas Biográficas

Albert da Costa Xavier é graduado em Ciência da Computação (Universidade José do Rosário Vellano – UNIFENAS, 2001) e mestre em Engenharia Elétrica (Centro Universitário da FEI, 2011). Atualmente é professor em regime parcial da Faculdade de Informática e Administração Paulista – FIAP. Tem experiência na área de Ciência da Computação com ênfase em reconhecimento de padrões em estatística e arquitetura de sistemas.

João Ricardo Sato é graduado em Estatística, mestre e doutor (Universidade de São Paulo, 2002, 2004 e 2007, respectivamente). Atua em projetos de pesquisa multidisciplinares envolvendo os seguintes temas: modelagem estatística e computacional em neurociências, neuroimagem, mapeamento funcional do cérebro humano, análise de séries temporais, bioestatística, estatística não-paramétrica e modelos de regressão.

Gilson Antônio Giraldi é graduado em Matemática (Pontifícia Universidade Católica de Campinas, 1986), mestre em Matemática Aplicada (Universidade Estadual de Campinas, 1993) e doutor em Engenharia de Sistemas e Computação (Universidade Federal do Rio de Janeiro, 2000). Atualmente é pesquisador adjunto III do Laboratório Nacional de Computação Científica. Tem experiência na área de Ciência da Computação, com ênfase em Matemática da Computação, atuando principalmente nos seguintes temas: segmentação, modelos deformáveis, processamento de imagens, *snakes* e animação de fluidos.

Paulo Sérgio Rodrigues é graduado em Ciência da Computação (Universidade Federal do Pará, 1996), mestre e doutor em Ciência da Computação (Universidade Federal de Minas Gerais, 1997 e 2003, respectivamente) e tem pós-doutoramento (Laboratório Nacional de Computação Científica, 2003-2007). Desde 2007 é professor em tempo integral da Fundação Educacional Inaciana, ministrando as cadeiras de Computação Gráfica (graduação), Visão Computacional e Geometria Computacional (pós-graduação em Engenharia Elétrica). Suas principais áreas de interesse são: visão computacional e reconhecimento de padrões, tendo publicado diversos artigos internacionais nessas áreas.

Carlos Eduardo Thomaz é graduado em Engenharia Eletrônica e mestre em Engenharia Elétrica (Pontifícia Universidade Católica do Rio de Janeiro, 1993 e 1999, respectivamente), doutor em Ciência da Computação (Imperial College London, 2005). Atualmente é professor adjunto do Centro Universitário da FEI. Tem experiência na área de Ciência da Computação, com ênfase em reconhecimento de padrões em estatística, atuando principalmente nos seguintes temas: visão computacional, computação em imagens médicas e biometria.

Capítulo 5

Auxílio ao Diagnóstico do Glaucoma Utilizando Processamento de Imagens

Virginia Ortiz Andersson e Lucas Ferrari de Oliveira*

Resumo: Glaucoma é uma doença capaz de causar danos no nervo óptico com impacto no campo visual. O diagnóstico do glaucoma é baseado na análise da escavação patológica através da inspeção do nervo óptico pela oftalmoscopia. Este capítulo apresenta o desenvolvimento de um software que calcula a razão entre as áreas da escavação e disco óptico bem como a razão entre os diâmetros. Estes valores são de extrema importância no diagnóstico do glaucoma e indicam o quanto o nervo óptico está escavado. Utiliza-se como técnica de segmentação de imagens o algoritmo de crescimento de região. Embora ferramentas similares existam, os custos são inacessíveis aos usuários. O diferencial do sistema proposto se baseia em uma maior acessibilidade que o software terá em relação aos já existentes.

Palavras-chave: Segmentação de imagens, Glaucoma, Crescimento de região.

Abstract: *Glaucoma is a disease capable of causing an injury in the optic nerve with progressive impact on the visual field. The diagnosis of glaucoma is based on the analysis of pathological excavation by inspection of the optic nerve through ophthalmoscopy. This chapter shows the development of a software which calculates the area ratio between pathological excavation and optical disk and the diameter ratios as well. These values are extremely important in the diagnosis of glaucoma since it indicates how much the optic nerve is excavated. We use region growing as technique for image segmentation. Although similar tools do exist, the cost is unreachable to users. The differential of the proposed system is based on a greater accessibility of the software related to the existing ones.*

Keywords: *Image segmentation, Glaucoma, Region growing.*

*Autor para contato: lferrari@ufpr.br

Neves et al. (Eds.), Avanços em Visão Computacional (2012) DOI: 10.7436/2012.avc.5 ISBN 978-85-64619-09-8

1. Introdução

O glaucoma é uma doença grave capaz de causar uma lesão progressiva no nervo óptico com repercussão sobre o campo visual. Ela é assintomática e é causa irreversível de cegueira. O diagnóstico precoce desta doença, feito principalmente pela análise do nervo óptico, é de extrema importância e não está acessível a todas as pessoas devido ao alto custo que este exame pode ter. Por não apresentar sintomas, o glaucoma só é percebido pelo paciente quando o mesmo já começa a ter dificuldades de enxergar, o que significa o comprometimento de grande parte das fibras nervosas oculares. O tratamento nesse caso é limitado. Na maioria dos casos esse quadro grave pode ser evitado se o glaucoma for detectado e tratado em tempo hábil.

1.1 Fisiopatologia do glaucoma

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 (Machado, 2000).

O disco óptico (DO) designa a porção intra-ocular do nervo óptico visível à oftalmoscopia e é a elipse com eixo vertical maior. O disco óptico é maior em míopes e pessoas afro-descendentes. Nos homens pode chegar a uma área 3% maior do que nas mulheres (Almeida, 2004).

A escavação é a elipse com menor eixo vertical e ocupa uma porção variável no nervo óptico. Na oftalmoscopia é visível pela área de coloração esbranquiçada no centro do disco óptico. A artéria e veias centrais da retina atravessam o disco e se bifurcam na superfície e bordas da escavação. 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 (Almeida, 2004).

A Figura 1 mostra duas imagens de fundo de olho apresentando discos ópticos de diferentes configurações: Em (a) escavação normal ou moderada, ocupando no máximo 30% do disco e em (b) escavação no glaucoma avançado ocupando 80% a 90% do disco óptico. Na Figura 2 observam-se as regiões do disco óptico D e escavação E, bem como a razão entre elas, calculada por D/E. Esta razão é conhecida como *Cup-to-Disc ratio* (CDR) ou razão escavação por disco (E/DO) e representa a divisão entre os limites da escavação pelos limites do disco óptico (Almeida, 2004).

Inúmeras vezes especialistas da área utilizam ferramentas como Adobe Photoshop (Knoll & Knoll, 2011) e GIMP (Kimball & Mattis, 1996) para elaborarem seus diagnósticos. Estes 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

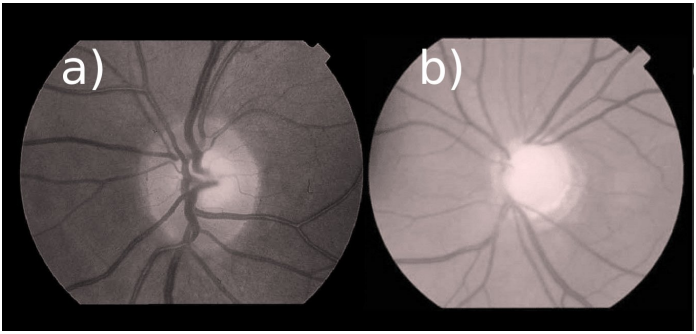


Figura 1. Em (a) uma retinografia apresentando DO com escavação considerada normal e em (b) escavação no glaucoma avançado (Bourne, 2006).

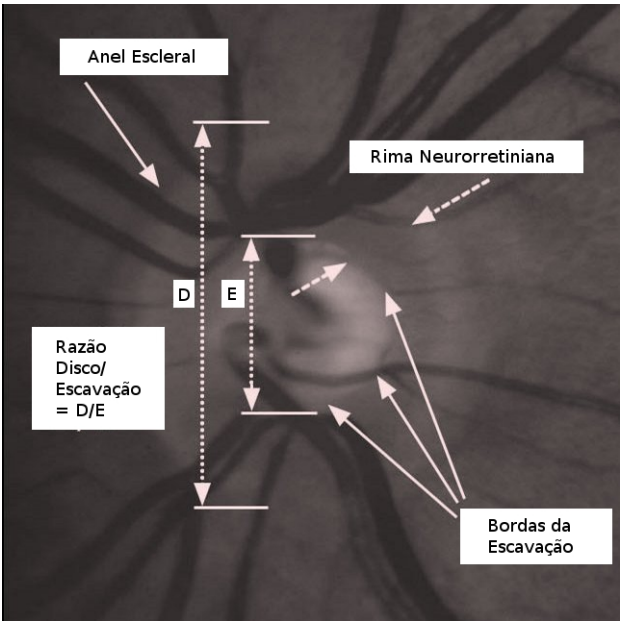


Figura 2. Marcação das regiões e razões entre elas (Tavares & Mello, 2005).

do tempo dedicado a elaboração do diagnóstico, pois os resultados provenientes dos softwares devem ser anotados manualmente para sua utilização.

O processamento de imagem é uma forma de processamento de dados no qual a entrada e a saída é uma imagem ou quadros de vídeos. Esse processo tem o objeto de otimizar a extração de informações e, portanto, ajudar na interpretação da imagem, que envolve a detecção e reconhecimento de elementos contidos na imagem. Basicamente o processamento de imagem tem como entrada uma imagem e saída um conjunto de valores numéricos, que podem ou não compor uma outra imagem (Marengoni & Stringhini, 2009). O processamento de imagens pode ser dividido em várias etapas, que vão desde a melhora da aquisição até a sua identificação. Uma das etapas é a de segmentação que consiste em subdividir a imagem em suas partes ou objetos constituintes. Utiliza-se para isto as propriedades básicas de descontinuidade dos níveis de cinza para segmentação através de bordas, fronteiras e linhas, ou de similaridade destes para separação de regiões que apresentem determinada característica em comum (Azevedo-Marques, 2001; Castleman, 1996).

O auxílio ao diagnóstico por computador (CAD - *Computer-Aided Diagnosis*) foi inicialmente utilizado na área de radiologia, porém hoje várias áreas se beneficiam das suas aplicações. Neste tipo de sistema o computador é utilizado como uma ferramenta para se obter uma informação extra e o diagnóstico é sempre feito pelo especialista humano. A finalidade dos sistemas CAD é aumentar a acurácia do diagnóstico, pois a avaliação do especialista, dependendo do tipo de exame e da técnica empregada, pode ser subjetiva e estar sujeita a variações inter e intrapessoais (Azevedo-Marques, 2001).

Com o intuito de auxiliar o diagnóstico precoce do glaucoma e prover o acompanhamento da neuropatia óptica glaucomatosa (danos no nervo óptico causados pelo glaucoma) foi desenvolvido um software que calcula a razão entre a área da escavação e a área do disco de um nervo óptico a partir de uma imagem de retinografia. Intitulado OnScope (*Optical Nerve Scope*), o software fornece ao especialista informações relevantes para o diagnóstico da doença. Além do cálculo de áreas, é possível calcular a razão entre os diâmetros das duas regiões e armazenar os resultados obtidos para o histórico e acompanhamento da doença. Essas informações periódicas são de grande valor pois mostram ao oftalmologista a evolução ou regressão da neuropatia óptica no paciente.

2. Materiais e Métodos

As imagens utilizadas neste trabalho foram adquiridas no formato JPEG e são provenientes dos equipamentos (i) retinógrafo TOPCON modelo TRC-

50DX[1] e (ii) OPTO ADS sistema de angiografia digital[2], que possui tecnologia CCD ADS 1.5 (resolução 1392 x 1040 pixels) e ADS 4.0 (resolução 2048 x 2048 pixels). Os equipamentos pertencem a duas clínicas de oftalmologia da cidade de Porto Alegre e as imagens foram cedidas e avaliadas por oftalmologistas.

A proposta do software OnScope é calcular duas áreas e diâmetros em imagens de retinografia e apresentar uma razão entre elas para utilização no diagnóstico clínico do glaucoma.

Devido à natureza do problema optou-se por utilizar um algoritmo de crescimento de região para segmentação de imagens, pois o mesmo funciona para o dado problema e é similar a algoritmos encontrados em funções de seleção de regiões nos softwares utilizados pelos especialistas. Com isto, a adaptação do usuário pode ser mais fácil.

As técnicas de crescimento de regiõe normalmente são utilizadas em conjunto com outras técnicas de segmentação, como no trabalho de Tang (2010). Ele propôs em seu trabalho o uso do algoritmo de *Watershed* na base do tradicional algoritmo de crescimento de região: onde antes as sementes eram escolhidas manualmente, na solução de Tang elas são geradas pelas regiões resultantes do algoritmo de *Watershed*. O uso combinado destes métodos resulta em um processo cuja complexidade computacional diminui em comparação com método de crescimento de região original. Originalmente o algoritmo depende do número de pixels da imagem e, na solução de Tang, o algoritmo de crescimento de região depende do número de regiões geradas pelo algoritmo de *Watershed*, como mostra a Figura 3. Além disto, esta proposta automatiza a seleção das sementes (Tang, 2010).

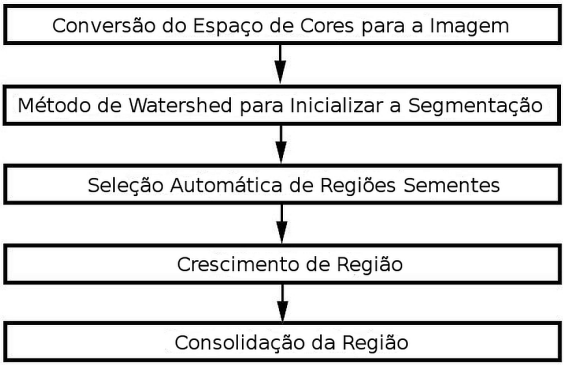


Figura 3. Fluxograma do algoritmo de Tang.

[1] http://www.topconmedical.com/products/trc50dx.htm
[2] http://www.opto.com.br/Produto.php?lingua_id=1&divisao_id=1&produto_id=49

2.1 Crescimento de região (*region growing*)

O crescimento de região é um procedimento contextual em segmentação de imagens, cuja forma mais simples é por agregação de pixels, onde o algoritmo começa com um conjunto de sementes e, a estas, agregam-se outros pixels que possuem propriedades similares (como níveis de cinza, textura ou cor) indicadas por algum predicado de uniformidade (Gonzalez & Woods, 2002).

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. Neste algoritmo é assumido 8 pixels como vizinhos ao redor do pixel atualmente analisado (*8-connectivity*) (Efford, 2000).

A agregação de um pixel candidato em uma região é feita se, e apenas se: a) O pixel não foi assimilado a nenhuma outra região b) O pixel é vizinho desta região c) A nova região, criada após sua agregação, continua uniforme

Cada pixel, ao ser adicionado à região, passa por um teste de uniformidade dado pela Equação 1 (Efford, 2000)

$$P(R) = \begin{cases} Verdadeiro & \text{se } |f(x,y) - \mu_R| \leq \Delta \\ Falso & \text{caso contrário} \end{cases} \quad (1)$$

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 quando não encontra mais pixels que satisfaçam os critérios de adição para cada região (Efford, 2000).

O algoritmo da metodologia de crescimento de região é:

Algorithm 1 Algoritmo de Crescimento de Região:

```
while R ← receber pixels do
  for i = 1 →n do
    for p = vizinhos da borda → borda do
      if (vizinho nao esta marcado  E  f(x,y) − μ_R ≤ Δ) then
        R_i ← f(x,y)
        μ_i = μ_i + 1
      end if
    end for
  end for
end while
```

A Figura 4 mostra como é aplicado o algoritmo em uma imagem simples (4 linhas e 5 colunas), considerando 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. Na Figura 4 (a) a semente inicial está marcarda em cinza na Figura 4 (b) mostra uma das iterações do algoritmo. Como resultado, apenas pixels com valores 0, 1 ou 2 são adicionados, pois eles satifazem a regra com limite igual a 3 para a semente escolhida.

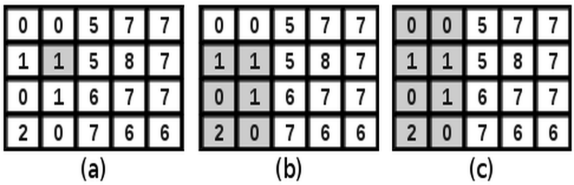


Figura 4. (a) Pixel semente; (b) primeira iteração;. (c) iteração final. (Efford, 2000)

No software, o especialista escolhe alguns pontos 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 software desenvolvido, o software Adobe Photoshop e um diagnóstico clínico realizado por um especialista. O comparativo serviu para mostrar a semelhança entre os resultados. Além disto, mostrou que ferramenta desenvolvida 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. Primeiramente, a verificação da capacidade do software de calcular a razão entre duas áreas usando objetos com medidas e resultados conhecidos. Em seguida, a verificação da capacidade do software ser usado no diagnóstico clínico. Para isto, os resultados foram comparados com resultados fornecidos pelo especialista.

3.1 Validação da metodologia

Cada paciente possui um tipo diferente de nervo óptico. Cada disco e escavação possuem diferentes medidas. Devido à esta falta de padrão nas medidas encontradas em indivíduos quaisquer, tornou-se impossível verificar se o software OnScope calculava corretamente a razão entre duas áreas em questão.

Para contornar este 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 escolhida uma moeda brasileira de 1 Real, mostrada na Figura 5.

Figura 5. Teste e validação utilizando moeda de 1 Real.

Os valores das áreas da moeda de 1 Real foram calculados utilizando uma régua milimetrada e $\pi = 3{,}14159265$. Segundo o Banco Central do Brasil (BACEN, 2011), 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 podem ser vistos 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 visando avaliar a variação entre as várias análises que foram feitas, pois assim tem-se uma estimativa da variação intra-imagem. Os valores mostram que a variação é pequena (desvio padrão igual a 0,01) e que a metodologia é robusta (Tabela 2).

Tabela 1. Valores reais da moeda de 1 Real.

Área Maior (cm^2)	**Área Menor** (cm^2)	**Razão das Áreas**	**Razão dos Diâmetros**
5,7	2,5	0,4	0,6

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 isto, foram submetidas ao teste algumas imagens de fundo de olho com os limites de nervo óptico e escavação bem definidos.

Tabela 2. Média e desvio padrão das razões do diâmetro e área da moeda calculados com o uso dos softwares.

	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

As imagens foram agrupadas por equipamento de origem: TOPCON e sistema de Angiografia OPTO ADS, e para cada equipamento um oftalmologista foi responsável pela análise das imagens. Os oftalmologistas calcularam os CDR's dos diâmetros utilizando o software PhotoShop e também a razão entre as áreas de maneira qualitativa, ou seja, atribuíram valores com base em suas experiências 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 pelos especialistas no software Photoshop utilizando a ferramenta de seleção por gama de cores. Essa etapa serviu como uma forma de comparação entre os resultados obtidos na segmentação realizada pelo software Onscope com outra ferramenta que é utilizada em casos clínicos, mas não é específica para isto.

4. Resultados

As Figuras 6 e 7 mostram o resultado de um caso segmentado e armazenado pelo software OnScope. As cores auxiliam na diferenciação entre as regiões de estudo. A Figura 7 é o resultado final do software a imagem é gravada com as informações (canto superior esquerdo) do cálculo entre regiões marcadas. Facilitando a recuperação posterior da informação pelo especialista, pois a própria imagem já possui o resultado. 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.

A comparação entre as razões das áreas das imagens de retinografia obtidas pelo OnScope, PhotoShop e oftalmologista pode ser vista na Tabela 3 e na Tabela 4. 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 e tem um valor

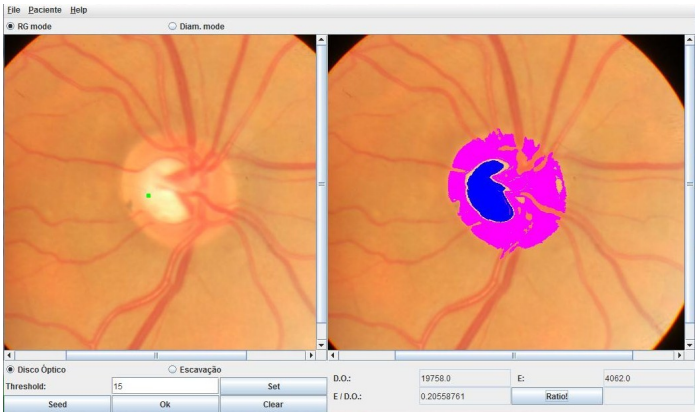


Figura 6. OnScope na modalidade de cálculo de áreas.

Figura 7. Dados do exame salvos no OnScope.

mais qualitativo, pois não é uma medida extraída da imagem e, sim, da experiência em diagnóstico.

Os valores das razões calculadas variam bastante entre o OnScope e o Photoshop, o que pode ser considerado uma diferença entre as metodologias utilizadas. Por outro lado, em comparação com o especialista também houve uma grande variação e a acurácia do sistema não pôde ser medida de forma direta.

Para a análise dos resultados foram calculados os coeficientes de correlação de Pearson entre os valores obtidos com os softwares OnScope, Photoshop e com os resultados do diagnóstico clínico, como mostram as Tabelas 5 e 6. Os valores superiores a 0,8 indicam uma alta correlação entre as variáveis comparadas e também um alto grau de semelhança nas respostas dadas pelos sistemas utilizados.

Tabela 3. Razões entre áreas obtidas nos softwares e no diagnóstico clínico, para imagens de retinografia obtidas com o retinógrafo TOPCON.

Teste	OnScope	Photoshop	Oftalmologista
1	0,20	0,32	0,20
2	0,35	0,24	0,30
3	0,42	0,47	0,50
4	0,23	0,20	0,30
5	0,92	0,45	0,80
6	0,12	0,10	0,30
7	0,42	0,36	0,40
8	0,18	0,31	0,10
9	0,94	1,17	0,90

Tabela 4. Razões entre áreas obtidas nos softwares e no diagnóstico clínico, para imagens de retinografia obtidas com o sistema de angiografia OPTO ADS.

Teste	OnScope	Photoshop	Oftalmologista
1	0,46	0,29	0,40
2	0,33	0,23	0,40
3	0,02	0,08	0,10
4	0,04	0,01	0,10
5	0,19	0,38	0,30
6	0,30	0,23	0,30
7	0,39	0,35	0,40
8	0,32	0,23	0,40
9	0,23	0,23	0,40
10	0,34	0,14	0,40
11	0,23	0,25	0,40
12	0,39	0,22	0,40
13	0,29	0,17	0,30
14	0,41	0,25	0,40
15	0,37	0,32	0,50
16	0,53	0,51	0,60
17	0,26	0,25	0,30
18	0,11	0,06	0,10
19	0,52	0,33	0,50
20	0,38	0,25	0,50

Tabela 5. Correlação de Pearson entre valores obtidos com OnScope, Photoshop e no diagnóstico clínico, para imagens de retinografia obtidas com o retinógrafo TOPCON.

OnScope e Photoshop	0,78
Photoshop e Oftalmologista	0,78
OnScope e Oftalmologista	0,95

Tabela 6. Correlação de Pearson entre valores obtidos com OnScope, Photoshop e no diagnóstico clínico, para imagens de retinografia obtidas com o sistema de angiografia OPTO ADS.

OnScope e Photoshop	0,78
Photoshop e Oftalmologista	0,77
OnScope e Oftalmologista	0,93

5. Discussão

Através dos coeficientes de correlação de Pearson é possível observar que existe uma forte correlação entre os resultados do OnScope e Phostoshop, bem como de ambos os softwares com os resultados fornecidos pelos oftalmologistas. Considerando que a correlação perfeita positiva entre duas variáveis é igual a 1, o software OnScope obteve 0,95 e 0,93 no ρ de Pearson mostrando a forte correlação existente entre as duas variáveis consideradas.

Comparando-se os resultados paralelamente, nas Tabelas 3 e 4 notamos uma diferença importante entre as razões de alguns testes. Isso se deve ao fato que o algoritmo de crescimento de região usado no software OnScope possibilita ao usuário uma maior liberdade 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 Photoshop forneceram resultados mais satisfatórios para determinados tipos de imagens 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.

6. Conclusão

O auxílio ao diagnóstico vem sendo amplamente implementado e testado por vários centros de pesquisa, pois aumenta o percentual de acertos, bem como auxilia nos casos mais complexos. O glaucoma é uma doença que pode ser tratada em seu estágio inicial e uma metodologia computacional que quantifique de forma precisa e eficaz se faz necessária.

Neste trabalho foi mostrado um sistema de auxílio ao diagnóstico do glaucoma e não foi encontrado na literatura científica nenhum trabalho onde se tenha desenvolvido ferramenta semelhante. Um trabalho que pode ser implementado como uma melhoria do OnScope é o trabalho de Tang (2010). Como o especialista gasta muito tempo tabelando as informações nos softwares que não são próprios para isto e precisam realizar cálculos com os valores, a grande contribuição do sistema proposto é o cálculo da CDR otimizando o trabalho e a tabulação dos dados clínicos.

O software OnScope funciona melhor para segmentações não-contínuas, 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 nestes casos.

7. Agradecimentos

Os autores agradecem aos oftalmologistas Dr. Manuel Augusto Pereira Vilela e Dr. Tadeu Antônio Di Francesco Pocai, pela ideia de desenvolvimento, bem como o fornecimento das imagens e suas avaliações clínicas. Os autores são gratos também ao CNPq (projeto 567035/2008-5) pelo apoio financeiro.

Referências

Almeida, G.V., *Manual de Semiologia do Glaucoma*. São Paulo: Phoenix, 2004.

Azevedo-Marques, P.M., Diagnóstico auxiliado por computador na radiologia. *Radiologia Brasileira*, 34(5):285–293, 2001.

BACEN, , Banco central do Brasil. http://www.bcb.gov.br/?MOEDAFAM2, 2011.

Bourne, R.R.A., The optic nerve head in glaucoma. *Community Eye Health Journal*, 19(59):44–45, 2006.

Castleman, K.R., *Digital Image Processing*. Upper Saddle River, USA: Prentice-Hall, 1996.

Efford, N., *Digital Image Processing – A Pratical Introduction Using Java*. Harlow, UK: Pearson Education, 2000.

Gonzalez, R.C. & Woods, R.E., *Digital Image Processing*. 2a edição. Upper Saddle River, USA: Prentice-Hall, 2002.

Kimball, S. & Mattis, P., GIMP – GNU image manipulation program. http://www.gimp.org/, 1996.

Knoll, T. & Knoll, J., Adobe Photoshop. http://www.photoshop.com/, 2011.

Machado, A.B.M., *Neuroanatomia Funcional*. 2a edição. São Paulo: Livraria Atheneu, 2000.

Marengoni, M. & Stringhini, S., Visão computacional usando OpenCV. *Revista Brasileira de Informática Teórica e Aplicada*, 16(1):125–160, 2009.

Tang, J., A color image segmentation algorithm based on region growing. In: *Proceedings of 2nd International Conference on Computer Engineering and Technology*. Piscataway, USA: IEEE Press, v. 6, p. 634–637, 2010.

Tavares, I.M. & Mello, P.A.A., Glaucoma de pressão normal. *Arquivos Brasileiros de Oftalmologia*, 64(8):565–575, 2005.

Notas Biográficas

Virginia Ortiz Andersson é graduada em Ciência da Computação (Universidade Federal de Pelotas – UFPel, 2008). Atualmente é Técnica de Tecnologia da Informação (CGIC - UFPel) e mestranda no Programa de Pós-Graduação em Ciência da Computação da UFPel.

Lucas Ferrari de Oliveira é graduado em Ciência da Computação (Universidade de Marília – UNIMAR, 1997), mestre em Engenharia Elétrica (Escola de Engenharia de São Carlos da Universidade de São Paulo – EESC/USP, 2000), doutor em Ciências Médicas (Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo – FMRP/USP, 2005) e fez pós-doutorado na FMRP/USP em 2006. Trabalha com processamento de imagens médicas, principalmente com imagens de Medicina Nuclear. Atualmente é professor e coordenador do curso de Tecnologia em Análise e Desenvolvimento de Sistemas (TADS) e professor colaborador do Programa de Pós-Graduação em Engenharia Elétrica, ambos na Universidade Federal do Paraná (UFPR).

Capítulo 6

Método para a Obtenção de Imagens Coloridas com o Uso de Sensores Monocromáticos

Flávio Pascoal Vieira e Evandro Luis Linhari Rodrigues*

Resumo: Este capítulo apresenta uma comparação entre sensores de imagens monocromáticas e sensores coloridos do tipo filtro de Bayer. Uma simulação ilustra a influência do processo de interpolação na resolução da imagem. É apresentada uma proposta para a obtenção de imagens coloridas a partir de capturas monocromáticas com a cena iluminada em diferentes regiões do espectro visível. O sistema foi implementado em um retinógrafo digital e imagens do fundo do olho são mostradas como exemplo do funcionamento do sistema proposto.

Palavras-chave: Retinografia, Filtro de Bayer, Imagens multiespectrais, Resolução.

Abstract: *This chapter presents a comparison between monochrome image sensors and Bayer filter color sensors. A simulation illustrates the influence of the interpolation process in the image resolution. A proposal is presented for obtaining color images from monochrome shots of the scene, illuminated in different regions of the visible spectrum. The system was implemented in a digital fundus camera and images of the eye are shown as an example of the proposed system.*

Keywords: *Fundus photography, Bayer filter, Multispectral images, Resolution.*

*Autor para contato: evandro@sc.usp.br

1. Introdução

As modernas técnicas de diagnóstico são ferramentas poderosas e seu uso está cada vez mais difundido entre médicos e outros profissionais da área da saúde, nas mais variadas especialidades. Dentre estas técnicas destacam-se as baseadas em imagens, que se desenvolveram de forma considerável nas últimas décadas. O surgimento de novos métodos computacionais e o aumento da capacidade dos *hardwares* contribuiram para o aparecimento de novas técnicas e melhorias nas já existentes.

As imagens geradas estão cada vez mais detalhadas, promovendo um aumento na segurança e precisão do diagnóstico. Esta tendência torna a área atrativa para a pesquisa acadêmica, pois a demanda tecnológica segue o crescimento de suas aplicações.

O estudo do princípio de funcionamento dos diversos métodos utilizados nesta área é particularmente importante quando se deseja promover uma melhoria ou a criação de uma nova técnica. Pesquisas recentes indicam uma tendência de exploração de imagens multiespectrais e hiperespectrais. Estas técnicas, que são comuns em sensoriamento remoto, mineralogia, agricultura e física, entre outras, está se mostrando uma grande aliada nas aplicações no campo da saúde e estética (Zawada, 2002; Pratavieira, 2010).

Têm-se destacado, também, técnicas de fluorescência e autofluorescência, principalmente na análise de tecidos biológicos. A combinação adequada entre as parcelas espectrais que iluminam a cena e a região do espectro em que ocorre a captura da imagem consegue evidenciar detalhes e revelar estruturas de acordo com o interesse do exame (Modugno, 2009; Oliveira et al., 2009).

Valendo destes princípios e seguindo a mesma tendência, encontram-se as técnicas para observação e registro de imagens da retina. A demanda por este tipo de exame tem crescido em função do aumento da expectativa de vida da população. A degeneração macular relacionada à idade (DMRI), a retinopatia diabética e o glaucoma apresentam-se no Brasil como principais doenças oculares dentre os idosos (Modugno, 2009).

O desenvolvimento de retinógrafos digitais é uma atividade constante nas empresas detentoras das tecnologias necessárias para a fabricação destes equipamentos. Um desafio permanente é o aumento da resolução das imagens captadas, que está relacionado principalmente à qualidade óptica e ao tipo de sensor utilizado. Outra busca é a adequação do sistema para a realização de exames multiespectrais que podem viabilizar novos métodos de diagnóstico, como mostrado recentemente nos estudos de oximetria de fundo de olho (Ramella-Roman & Mathews, 2007; Everdell et al., 2010).

A proposta apresentada neste trabalho consiste em uma modificação no método de captura de imagens, utilizando um sensor monocromático e a captura sequencial de três quadros para a composição da imagem colorida.

Cada uma das capturas sequenciais deve corresponder a uma parcela do espectro visível.

Para exemplificação do funcionamento do sistema proposto é apresentado sua utilização em um retinógrafo digital, embora esta técnica possa ser utilizada no desenvolvimento de outros equipamentos.

Na primeira parte, o principal objetivo é avaliar, por meio de simulação matemática, a degradação inerente ao processo de interpolação em um sensor de imagens em cores utilizado na captura de uma cena monocromática. A avaliação pode ser estendida ao universo da retinografia, já que alguns equipamentos utilizam sensores de imagens em cores nos exames que resultam em imagens monocromáticas.

Na sequência, apresenta-se um sistema com o qual é possível a obtenção de uma imagem colorida a partir do uso de um sensor de imagens monocromáticas. A explanação também serve de suporte para melhorias no sistema com o qual futuramente se pretende compor uma configuração multiespectral completa.

2. Fundamentação Teórica

2.1 Funcionamento do retinógrafo

Um retinógrafo digital é um equipamento concebido para realizar a captura de imagens do segmento anterior do olho, possibilitando a observação de estruturas como, por exemplo, a cabeça do nervo óptico, a retina e, em especial, a mácula e a fóvea.

Da mesma maneira que em uma 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 (Modugno, 2009).

Os processos de iluminação e captação são dificultados devido à grande quantidade de interfaces do olho, sendo necessário transpor as barreiras impostas pela córnea, cristalino, íris e humor vítreo. A consequência direta é que o sinal proveniente da retina é de baixa intensidade. Por este motivo quase a totalidade dos retinógrafos digitais utiliza sensores de imagem com tecnologia CCD (*Charge-Coupled Device*). Quando comparados a sensores de tecnologia CMOS (*Complementary Metal-Oxide Semiconductor*), os sensores CCD conferem ao equipamento uma relação sinal-ruído melhor, 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.

2.1.1 Tipos de exame

As imagens do fundo do olho podem ser divididas em dois grupos: as retinografias e as angiografias. Os principais tipos de retinografias são a colorida e a anerítra, também conhecida como *red free*, com exemplos mostrados na Figura 1.

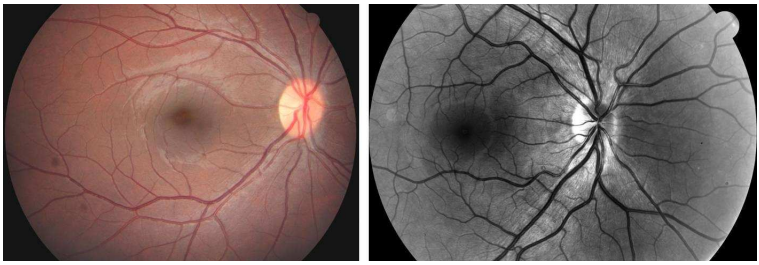


Figura 1. Exemplos de imagens de retinografia colorida (esquerda) e anerítra (direita) de pacientes distintos.

As angiografias são técnicas de fluorescência e envolvem a administração de contraste na corrente sanguínea do paciente para destaque da circulação. Os tipos mais comuns, mostrados na Figura 2, são com Fluoreceína (esquerda) e com Indocianina verde (direita).

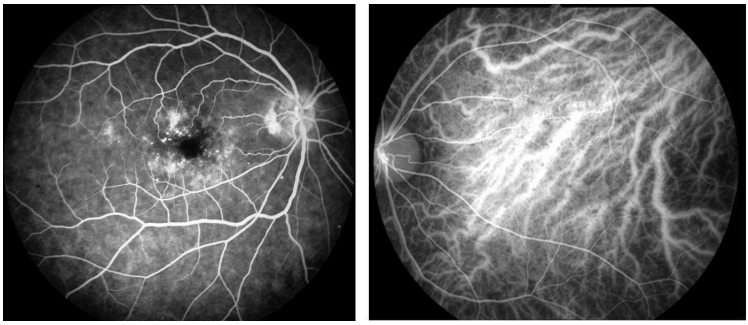


Figura 2. Angiografias do fundo do olho.

Os princípios de funcionamento destes exames são descritos com detalhes por Modugno (2009) e Carvalho (2006).

2.1.2 Arquitetura do sistema de captação

Alguns retinógrafos do mercado possuem apenas um sensor de imagem, geralmente do tipo com filtro de Bayer (Modugno, 2009). Neste tipo de

sensor, cada unidade do conjunto de elementos sensíveis possui um filtro espectral, que é depositado geralmente por processo de evaporação, de modo que apenas uma estreita faixa do espectro visível atinge o elemento sensível (Martins & Rodrigues, 2010). No modelo de Bayer, são utilizados três tipos de filtros espectrais R, G e B cuja faixa de transmissão está centrada respectivamente na região do vermelho (*red*), verde (*green*) e azul (*blue*). Os três tipos de filtros são dispostos em um padrão por toda a área sensível do sensor de modo que, para a composição de um pixel, é necessário informações referentes aos três elementos. Desta forma, para compor a imagem final, seja ela colorida ou monocromática, é necessário um processo matemático de interpolação (Rajeev et al., 2002). Para a avaliação do desempenho do sistema foi escolhida como métrica a avaliação da função de transferência (MTF). A resolução da imagem final está ligada diretamente às funções de transferência de cada conjunto que compõe o sistema de captura (Scaduto, 2008). Ou seja:

$$MTF_{total} = MTF_{optica} \cdot MTF_{sensor} \cdot MTF_{eletronica} \tag{1}$$

Explorando a Equação 1 pode-se expressar a influência do sensor como:

$$MTF_{sensor} = MTF_{geometrica} \cdot MTF_{difusaodecargas} \cdot MTF_{CTE} \tag{2}$$

A $MTF_{difusaodecargas}$ depende de parâmetros construtivos do componente, e a MTF_{CTE} está relacionada à interferência entre pixels vizinhos. Ambos os valores, dependem da tecnologia de fabricação do sensor.

Tanto a cena a ser capturada quanto a imagem projetada pelo sistema óptico no plano focal do sensor são funções contínuas no espaço. No momento em que é feita a digitalização pelo sensor de imagens, a informação passa a ser descrita por um número finito de elementos, no qual dentro de um mesmo elemento espacial não há variação de intensidade, ou seja, uma função descontínua (Pratt, 2007). O processo é descrito pelo teorema da amostragem e o tamanho do elemento sensível, determina a $MTF_{geometrica}$, que provoca na imagem uma degradação fixa no sinal, expressa pela função *sinc*:

$$MTF_{geometrica}(f_s) = \left| \frac{sen(\pi \cdot f_s \cdot \Delta S)}{(\pi \cdot f_s \cdot \Delta S)} \right| = |sinc(f_s \cdot \Delta S)| \tag{3}$$

onde: f_s é a frequência espacial e ΔS é o tamanho do pixel.

A frequência de Nyquist f_N, que limita a amostragem e em consequência a resolução espacial, é expressa por:

$$f_N = \frac{1}{2 \cdot \Delta S} \tag{4}$$

Esta análise é aplicável diretamente a sensores de imagens monocromáticas. No caso de sensores de imagens em cores é necessário a consideração de outros elementos (Elor et al., 2007). 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.

Situação semelhante acontece em outros equipamentos que possuem como princípio de funcionamento algum fenômeno de fluorescência ou quando a iluminação é limitada a uma fração do espectro visível. Busca-se, neste caso, avaliar a influência de um sensor de imagens em cores em comparação com um sensor de imagens monocromáticas. Ambos os sistemas foram modelados conforme a Figura 3. Considera-se que a óptica é perfeita e o tamanho do elemento sensível é igual para os dois sistemas, visto que neste trabalho é avaliada apenas a influência da arquitetura dos sensores.

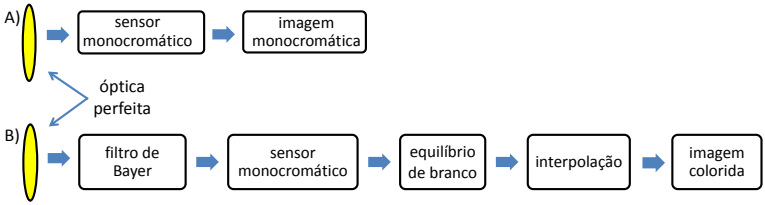


Figura 3. Modelo para o sistema de aquisição de imagens (A) monocromáticas e (B) coloridas.

O sistema B na Figura 3, além do filtro de Bayer, possui dois elementos adicionais (Elor et al., 2007). O sistema de equilíbrio 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 a seguir.

2.2 Filtro de Bayer e interpolação

Considerando um filtro de Bayer clássico com configuração BGGR, mostrado esquematicamente na Figura 4, é possível observar que cada elemento sensível é responsável por apenas um canal de cor (Bayer, 1976). 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 (Rajeev et al., 2002).

Neste trabalho será considerado o processo de interpolação linear. Este processo foi escolhido pois permite evidenciar com facilidade a influência do processo matemático necessário para a formação da imagem, um dos objetos de estudo deste trabalho. Além disto, o método é amplamente utilizado devido à sua facilidade de implementação. Neste processo o valor

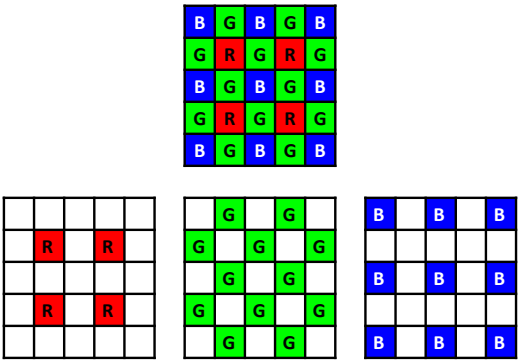


Figura 4. Representação de um sensor com filtro de Bayer. Cada elemento sensível recebe apenas um canal de cor.

de cada pixel após a interpolação é uma função linear de sua vizinhança na imagem original (Elor et al., 2007). O algoritmo de interpolação linear é expresso por:

$$Vout_c(x,y) = \sum_{x'} \sum_{y'} Vin_c(x',y') \cdot f_c(x,y,x',y') \tag{5}$$

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:

$$Vin_c = \begin{cases} V(x,y), & \text{caso pixel (x,y)} \in \text{canal de cor c} \\ 0, & \text{caso contrário} \end{cases} \tag{6}$$

onde: $V(x,y)$ é o valor amostrado do pixel de coordenadas (x,y).

A exemplo de outros algoritmos de interpolação linear (Elor et al., 2007), utilizaram-se como funções peso as f_c^k mostradas na Figura 5 onde c representa o canal de cor da imagem interpolada (RGB) e k representa o plano do filtro de Bayer (BGGR).

Com base nos resultados obtidos por Elor et al. (2007), 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 frequências espúrias no sistema. Ainda, seguindo os resultados de Elor et al. (2007), tem-se na forma analítica:

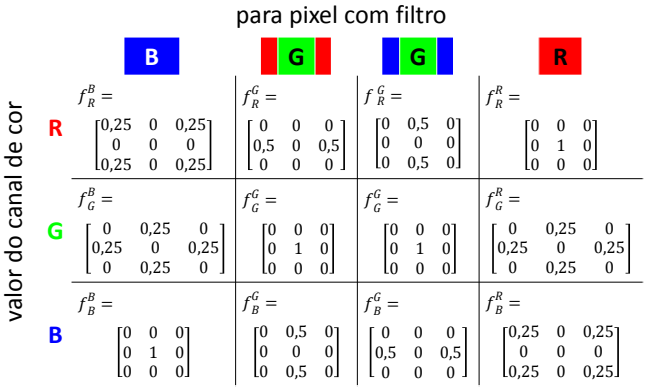


Figura 5. Funções peso para interpolação linear.

$$MTF_{interpolacao} = 0,65 + 0,35 \cdot cos(2 \cdot \pi \cdot f) \tag{7}$$

$$\text{Efeito Espúrio} = 0,1 - 0,1 \cdot cos(2 \cdot \pi \cdot f) \tag{8}$$

Como se trata de um processo de discretização, o sistema é limitado pelo teorema da amostragem, ou seja, a resposta em termos da frequência espacial é limitada em função das características geométricas de cada elemento sensor. Busca-se, neste momento, explorar apenas os efeitos dos processos matemáticos sobre o processamento da imagem. Desta maneira, aplicando-se uma imagem da forma $\phi(f) = \sqrt{sinc(f)}$ como entrada do sistema, tem-se o valor máximo possível para cada frequência, e considerando:

$$MTF_{optica} = MTF_{eletronica} = MTF_{ideal} \tag{9}$$

$$MTF_{ideal} = \begin{cases} 1, & \text{para} \quad x < f_N, \\ 0, & \text{cc} \end{cases} \tag{10}$$

Pode-se traçar o gráfico apresentado na Figura 6, que mostra a função de transferência do sensor de imagem em função da frequência espacial, para o sistema monocromático e colorido, além da curva da resposta espúria do processo de interpolação.

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.

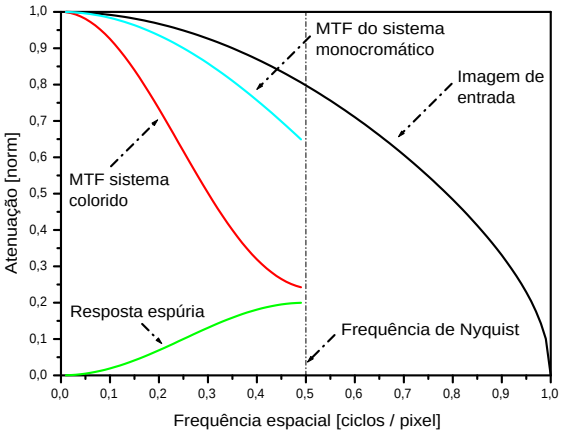


Figura 6. Resultados analíticos da MTF para um sistema colorido com filtro de Bayer e processo de interpolação linear, e um sistema monocromático.

2.2.1 Resultados da simulação numérica

A simulação utiliza um princípio comum na estimação da função de transferência, o algoritmo de *Knife Edge*, fundamentado na resposta ao degrau (Modugno, 2009). Nesta técnica, e com as considerações feitas acerca da idealidade da óptica e eletrônica, supõe-se no CCD uma imagem composta por apenas dois tons. De um lado a energia mínima (preto) e do outro a máxima (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 7, que representa uma parcela de 20 pixels do CCD colorido, na qual a transição ocorre entre a segunda e a terceira coluna. A Figura 8 identifica os itens na representação de cada pixel.

Para a imagem capturada com o sensor de imagens em cores é necessário a conversão para a escala de cinza, já que a visualização e reprodução,

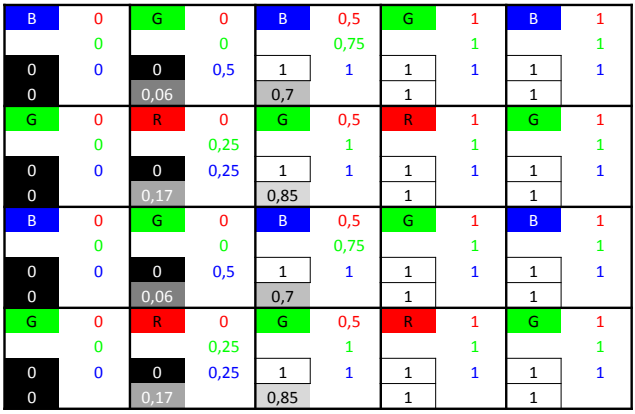


Figura 7. Representação simulada da imagem de um degrau atingindo um sensor com filtro de Bayer.

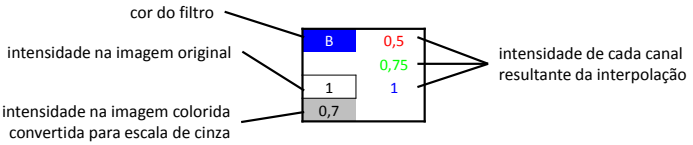


Figura 8. Identificação dos valores para cada pixel apresentados na simulação.

assim como a cena capturada, são monocromáticas. Esta conversão pode ser realizada por meio da equação de luminância recomendada pela CIE (Smith & Guild, 1931):

$$V_{PB} = \begin{bmatrix} 0,2989 & 0,5879 & 0,1140 \end{bmatrix} \cdot \begin{bmatrix} V_R \\ V_G \\ V_B \end{bmatrix} \tag{11}$$

Observando-se a Figura 9 (B), na segunda coluna, os pixels que originalmente possuíam valor 0 (preto) passaram para os valores 0,06 e 0,17 (conforme Figura 7). De forma semelhante, na terceira coluna, onde originalmente os valores eram 1 (branco), após o processo de interpolação os pixels assumiram valores 0,85 ou 0,7.

O resultado se traduz no surgimento de valores espúrios na proximidade da borda de transição, conforme representado na Figura 9. 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.

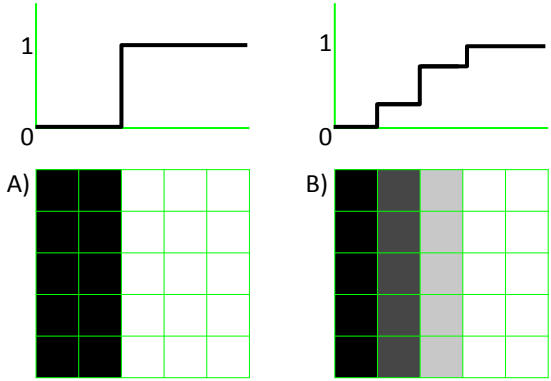


Figura 9. Resultado do imageamento de uma cena monocromática com um sensor de imagens monocromático (A) e um sensor de imagens em cores (B). Apresenta-se também um gráfico com o perfil da transição em cada caso.

Por meio da transformada de Fourier destas duas imagens resultantes, é possível determinar a função apresentada na Figura 10. Para tal, foram consideradas especificações do tamanho do elemento sensor igual a 9μm, o que implica no valor da frequência de Nyquist próximo de 55 ciclos/mm.

Com a observação do resultado ilustrado no gráfico é possível notar que, para a quase totalidade da faixa considerada, a função de transferência do sistema monocromático provoca uma atenuação menor no sinal. Este fato tem impacto direto no contraste da imagem, além de aumentar a relação sinal ruído do sistema monocromático quando comparado com o sistema colorido. Outra informação possível de ser extraída pela observação do mesmo gráfico é que o sistema monocromático preserva melhor as componentes de alta frequência da imagem, contribuindo positivamente com a definição de detalhes e a capacidade de reprodução de estruturas diminutas.

Supondo que um algoritmo de detecção de padrões seja capaz de identificar estruturas cuja a atenuação máxima devido à função de transferência seja de 75% (apenas na MTF_{sensor}), pode-se calcular o limite inferior do tamanho do elemento que se consegue resolver. Com base no gráfico mostrado, tem-se em 75% do contraste uma resolução de 10 ciclos/mm para o sistema colorido e 18 ciclos/mm no sistema monocromático. Aplicando os resultados a um retinógrafo comercial que possui uma ampliação de

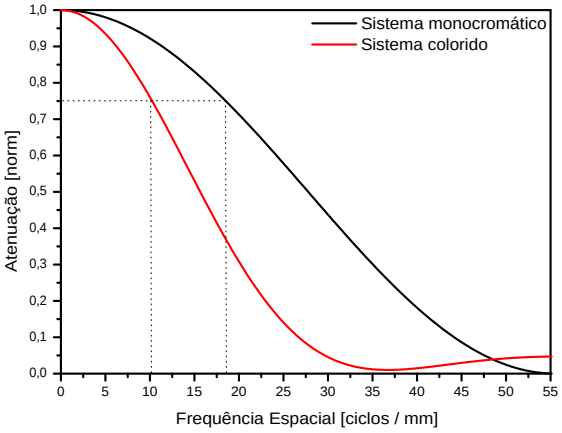


Figura 10. Resultado da simulação para os dois sistemas propostos.

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.

Com base nos resultados obtidos, fica evidente um melhor desempenho de sistemas com sensores monocromáticos quando utilizados no imageamento de cenas monocromáticas. Este resultado serve de suporte para a modificação proposta no sistema.

3. Metodologia

A obtenção de uma imagem colorida pode ser realizada de diversas maneiras. Na literatura, as primeiras ocorrências relatando a obtenção de uma imagem em cores descrevem o uso de um sensor monocromático e uma roda de filtros inserida no caminho óptico de formação da imagem. Na época, a tecnologia disponível ainda não permitia a fabricação do filtro de Bayer. Há relatos do uso destes sistemas nas primeiras missões espaciais com elementos imageadores.

Mesmo parecendo arcaico, este sistema ainda é utilizado em alguns instrumentos ópticos como telescópios e microscópios. O uso de filtros em diversas bandas criam imagens com diferentes formações espectrais e permitem a exploração de diferentes composições no espaço de reprodução de cores. Na área de interesse as perdas inerentes ao processo são minimizadas.

Outra forma de se obter a imagem colorida é o uso de três sensores e um conjunto óptico que promove a separação de cores. Com uma montagem semelhante à da Figura 11, a energia da imagem incidente é separada e as parcelas correspondentes às cores primárias são direcionadas para os respectivos sensores.

Esta tecnologia pode ser encontrada em câmeras de vídeo profissionais. Sua grande vantagem em relação ao método anterior é que a captura para cada canal de cor é realizada simultaneamente. Suas principais desvantagens são o custo dos três sensores e do elemento óptico, bem como a maior complexidade do sistema eletrônico.

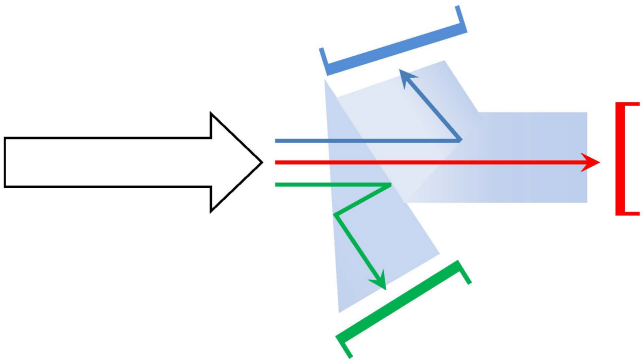


Figura 11. Montagem com 3 sensores para imageamento colorido.

Ambos os métodos apresentados possuem a vantagem de serem livres dos processos matemáticos de interpolação que, como demonstrado, degradam o sinal.

A proposta consiste em manter apenas um elemento sensor e modificar a iluminação da cena. A cena será iluminada com o uso de três LEDs (*Light Emitting Diodes*) monocromáticos, vermelho, verde e azul, acionados sequencialmente e em sincronismo com a captura da imagem.

A cada conjunto de três imagens consegue-se toda a informação necessária para a composição de uma imagem colorida utilizando todos os pixels do sensor.

Para a implementação do sistema de iluminação descrito é possível o uso de uma arquitetura conforme a mostrada na Figura 12. Com o uso de dois filtros dicróicos, (um para azul e outro para vermelho) é possível direcionar a energia luminosa dos três LEDs, permitindo que seja acoplada no sistema de iluminação do equipamento.

As principais vantagens são: o uso de apenas um sensor, sem componentes móveis e sem processos matemáticos de interpolação.

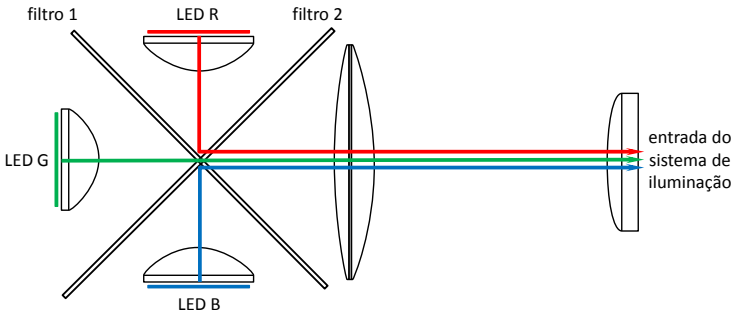


Figura 12. Sistema desenvolvido com a finalidade de direcionar a energia proveniente de três LEDs em um mesmo caminho óptico.

Outros detalhes específicos da metodologia variam de acordo com as limitações dos componentes utilizados e integração sistêmica das partes. Eles serão abordados na seção seguinte.

4. Resultados

Para a validação da metodologia e comprovação prática dos resultados obtidos teoricamente, um protótipo de retinógrafo digital foi montado com o sistema proposto. O desempenho superior para captura de imagens geradas a partir de cenas monocromáticas pôde ser notado já nas primeiras fotos.

O imageamento com a iluminação nos três canais e posterior composição da foto colorida para cenas estáticas é feita de maneira direta. A partir do imageamento de um alvo branco padrão é feito um ajuste na corrente de cada LED, de modo a obter uma imagem final com uma reprodução de cores mais próxima possível do objeto original.

Para imagens coloridas do olho, a obtenção da foto final demanda um processamento adicional. O olho humano pode apresentar pequenas movimentações entres as três fotos, o que causa o desalinhamento dos canais e consequente prejuízo da qualidade da imagem colorida. Tal fato está relacionado com a baixa frequência de aquisição da câmera utilizada (aproximadamente 7 fps).

Neste caso, para alinhar os três canais um algoritmo de *image registration* foi desenvolvido. O alinhamento é feito por meio do método de correlação de fase, na qual há o processamento da imagem no domínio da frequência. Todo o processo é descrito por (Pratt, 2007) e se inicia com um par de imagens (F_1 e F_2) transladada de $(x_0, y_0,)$ conforme a Equação 12.

$$F_2(x, y) = F_1(x - x_0, y - y_0) \tag{12}$$

Aplicando a transformada de Fourier e a propriedade do deslocamento tem-se:

$$F_2(\omega_x, \omega_y) = F_1(\omega_x, \omega_y) e^{\{-i(\omega_x x_0 + \omega_y y_0)\}} \tag{13}$$

O fator exponencial de deslocamento de fase pode ser calculado com o produto espectral cruzado e é dado pela Equação 14:

$$G(\omega_x, \omega_y) \equiv \frac{F_1(\omega_x, \omega_y) F_2^*(\omega_x, \omega_y)}{\|F_1(\omega_x, \omega_y) F_2(\omega_x, \omega_y\|} = e^{\{i(\omega_x x_0 + \omega_y y_0)\}} \tag{14}$$

A informação do deslocamento é tomada a partir da obtenção do ponto máximo da transformada inversa deste produto. Aplicando-se a transformada de Fourier inversa, obtém-se o deslocamento espacial (Equação 15) que é utilizado para realizar uma translação em X e Y para alinhar as imagens. Como um deslocamento no domínio do espaço gera um deslocamento de fase no domínio da frequência, apenas a fase da imagem é analisada.

$$G(x, y) = \delta(x - x_0, y - y_0) \tag{15}$$

O canal verde é fixado como canal base por ser o que normalmente apresenta melhor definição e contraste. Os outros dois canais são alinhados em relação a ele. Há a possibilidade de se utilizar processamentos semelhantes para ajuste de escala e rotação. Entretanto, até o momento, para as imagens de retina processadas, tais correções não foram necessárias.

As imagens originais para cada canal de iluminação e o resultado final da montagem colorida são mostradas na Figura 13.

A comparação visual das imagens pode ser feita na Figura 14, que mostra à esquerda uma imagem capturada com o uso de um sensor de imagens em cores com filtro de Bayer e à direita uma imagem formada a partir de três capturas conforme o método proposto. A região fotografada é a cabeça do nervo óptico. Observa-se uma maior definição de detalhes, permitindo a identificação de estruturas diminutas.

Outra metodologia para avaliar a influência do filtro de Bayer e processos de interpolação envolvidos na captura de imagens coloridas é a subtração pixel a pixel de duas imagens da mesma cena, uma obtida pelo método proposto e outra que sofreu os efeitos já mencionados. Como os processos influenciam principalmente as regiões de transição, o resultado deve se aproximar ao efeito da aplicação de um processo de detecção de bordas, isto é, evidenciando as diferenças existentes nos detalhes.

Como há uma grande dificuldade técnica de se obter duas imagens por métodos diferentes e mantendo as características tomadas como premissas no início do desenvolvimento matemático, tais como óptica perfeita, e

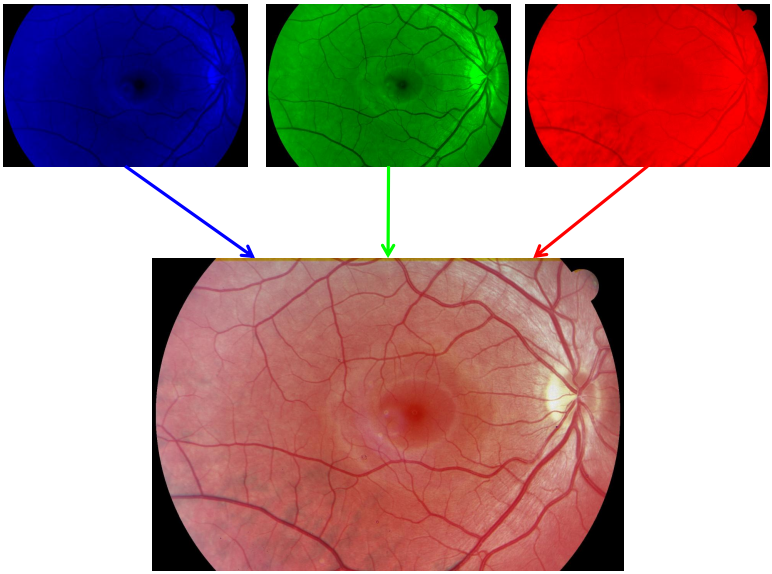


Figura 13. Imagem colorida obtida a partir da captura sequencial dos três canais.

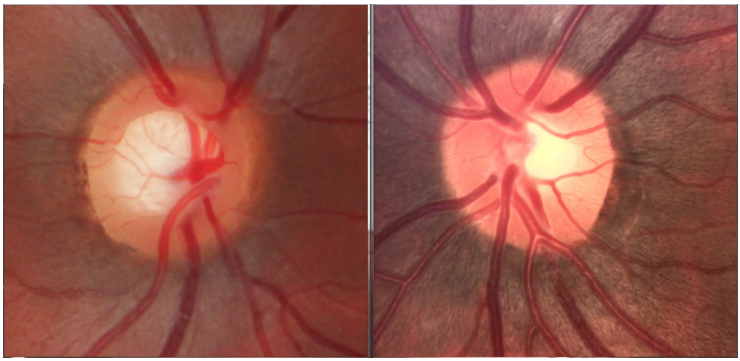


Figura 14. Imagem obtida com um sensor de imagens em cores (esquerda) e com o método proposto (direita).

mesmo tamanho do elemento sensor, o par de imagens foi obtido da seguinte maneira: capturou-se uma imagem em cores da retina mostrada na Figura 15 (A) com o método proposto neste trabalho. O efeito do filtro de Bayer que impede que cada elemento sensível seja atingido pelos três canais

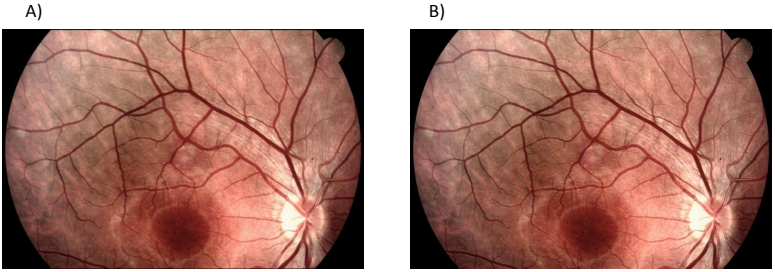


Figura 15. (A) Imagem obtida com o método proposto. (B) Imagem gerada simulando os efeitos do filtro de Bayer e processos de interpolação.

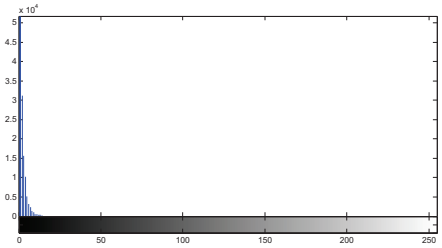


Figura 16. Histograma da diferença das imagens em análise.

espectrais e os processos de interpolação linear descritos foram aplicados na imagem original, gerando a imagem mostrada na Figura 15 (B).

Realizou-se a subtração das imagens, entretanto o resultado obtido não carregava informações suficientes para justificar sua exibição. Com a análise por meio do histograma, mostrado na Figura 16, pode-se observar que as diferenças são de baixa intensidade. Mesmo assim a presença desta diferença se distribui por toda a imagem e pode ser avaliada espacialmente por meio da Figura 17 que mostra a imagem diferença após um processo de equalização de histograma.

5. Discussão e Conclusões

Com os resultados teóricos apresentados conclui-se que para cenas monocromáticas, que é o caso das angiografias e retinografia anerítra, o sistema com sensores de imagens em cores apresenta um desempenho inferior quando comparado ao sistema monocromático. O efeito espúrio decorrente do processo de interpolação acarreta uma diminuição significativa da resolução do sistema.

O sistema montado usando apenas um sensor monocromático possibilita a obtenção de imagens coloridas sem a interferência de processos

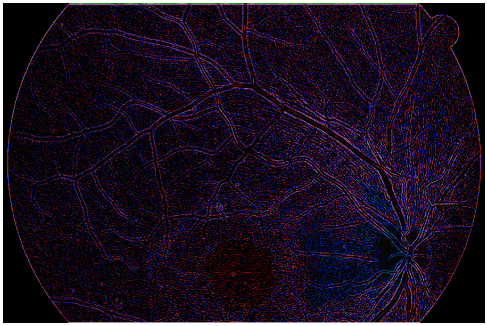


Figura 17. Diferença entre as imagens após processo de equalização de histograma.

matemáticos. O sistema pode ser aproveitado para o imageamento em alta resolução em diversas ocasiões, apresentando uma maior facilidade de implementação quando se trata de cenas estáticas.

Em retinografia digital esta proposta é particularmente interessante, pois a maioria dos exames realizados são cenas monocromáticas e o sistema apresentou ainda um excelente desempenho quando utilizado para a imagem colorida da retina.

O sistema montado pode ser modificado facilmente de modo a permitir o estudo em outras regiões espectrais. Com o uso de LEDs em diferentes comprimentos de onda é possível obter um conjunto de imagens multiespectrais. A partir deste conjunto, podem ser utilizados algoritmos de mineração de dados, tanto baseado em informações radiométricas, quanto espectrais, que poderão gerar resultados que auxiliem no diagnóstico.

Agradecimentos

À Opto Eletrônica SA, pelo suporte, incentivo à produção acadêmica e pela cessão de alguns resultados apresentados. À Edenilda Aparecida da Silva, pela colaboração com o texto e imagens. Aos colegas, Diego Lencione e José Augusto Stuchi, pelos inestimáveis ensinamentos.

Referências

Bayer, B.E., *Color Imaging Array.* U.S. Patent No 3971065, 1976.

Carvalho, E.G., *Desenvolvimento de um Sistema Óptico para Retinografia e Angiografia Digital.* Dissertação de mestrado em física aplicada, Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, 2006.

Elor, Y.; Pinsky, E. & Yaacobi, A., MTF for Bayer pattern color detector. In: Kadar, I. (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition XVI*. Bellingham, USA: Society of Photo Optical, v. 6567 de *Proceedings of SPIE*, p. 65671M, 2007.

Everdell, N.L.; Styles, I.B.; Calcagni, A.; Gibson, J.; Hebden, J. & Claridge, E., Multispectral imaging of the ocular fundus using light emitting diode illumination. *Review of Scientific Instruments*, 81(9):093706, 2010.

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. In: *Anais do VI Workshop de Visão Computacional*. Presidente Prudente, SP, p. 168–174, 2010.

Modugno, R.G., *Uma Contribuição ao Projeto de Retinógrafos Digitais*. Dissertação de mestrado em engenharia elétrica, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, SP, 2009.

Oliveira, T.B.; Trevelin, L.C.; Moreira, F.M.A.; Bagnato, V.S.; Schor, P. & Carvalho, L.A.V., Desenvolvimento e resultados preliminares de um sistema cromático de iluminação para oftalmoscópios indiretos. *Arquivos Brasileiros de Oftalmologia*, 72(2):146–151, 2009.

Pratavieira, S., *Desenvolvimento e Avaliação de um Sistema de Imagem Multiespectral para o Diagnóstico Óptico de Lesões Neoplásticas*. Dissertação de mestrado em física aplicada, Instituto de Física de São Carlos, Universidade de São Paul, São Carlos, SP, 2010.

Pratt, W.K., *Digital Image Processing: PIKS Scientific Inside*. 4a edição. J. Wiley, 2007.

Rajeev, R.; Snyder, W.E.; Bilbro, G.L. & Sander III, W.A., Demosaicking methods for Bayer color arrays. *Journal of Electronic Imaging*, 11(3):306–315, 2002.

Ramella-Roman, J.C. & Mathews, S.A., Spectroscopic measurements of oxygen saturation in the retina. *IEEE Journal of Selected Topics in Quantum Electronics*, 13(6):1697–1703, 2007.

Scaduto, L.C.N., *Desenvolvimento e Avaliação do Desempenho de Sistema Óptico Aplicado a Sensoriamento Remoto Orbital*. Dissertação de mestrado em física aplicada, Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, 2008.

Smith, T. & Guild, J., The C.I.E. colorimetric standards and their use. *Transactions of the Optical Society*, 33(3):73–134, 1931.

Zawada, D.G., *The Application of a Novel Multispectral Imaging System to the in vivo Study of Fluorescent Compounds in Selected Marine Organisms*. Phd thesis in oceanography, University of California at San Diego, San Diego, USA, 2002.

Notas Biográficas

Flávio Pascoal Vieira é graduado em Engenharia Elétrica com ênfase em Eletrônica (Escola de Engenharia de São Carlos da Universidade de São Paulo – USP, 2008). Atualmente é mestrando na área de instrumentação e processamento de sinais (USP, São Carlos) e trabalha na empresa Opto Eletrônica S/A, em pesquisa e desenvolvimento atuando em projeto de *hardware* e *software* embarcado para equipamentos médicos.

Evandro Luis Linhari Rodrigues possui graduação em Engenharia Elétrica (Escola de Engenharia de Lins, 1983), mestrado em Engenharia Elétrica (Universidade de São Paulo – USP, 1992) e doutorado em Física (USP, 1998). Atualmente é professor da USP e atua principalmente nos seguintes temas: processamento de imagens, microprocessadores/microcontroladores, visão computacional, análise carpal e automação.

Capítulo 7

Sistema para Classificação Automática de Café em Grãos por Cor e Forma Através de Imagens Digitais

Pedro Ivo de Castro Oyama [*], Lúcio André de Castro Jorge, Evandro Luis Linhari Rodrigues e Carlos Cesar Gomes

Resumo: Este capítulo apresenta um sistema de visão computacional desenvolvido para classificação de pequenas amostras de café em grãos, sendo utilizado em laboratórios de análise. Ele tem o intuito de substituir o atual processo de inspeção visual feito por especialista humano, que é lento e muitas vezes incapaz de satisfazer a demanda da indústria de café. O sistema usa algoritmos de processamento de imagem para extrair atributos de cor e forma de imagens de amostra de grãos. Redes neurais artificiais do tipo *Multilayer Perceptron* foram usadas para reconhecer padrões de cor e forma. Bons resultados preliminares foram obtidos, mas melhorias ainda são necessárias.

Palavras-chave: Visão computacional, Classificação de café, Redes neurais, Reconhecimento de padrões.

Abstract: *This chapter presents a vision system for the classification of small samples of coffee beans to be used in the laboratory of coffee quality analysis. It is aimed at replacing the current visual inspection made by a human specialist, which is slow and often unable to fulfil the demands of the coffee industry. The system uses image processing algorithms to extract color and shape attributes from images of samples of beans. Multilayer Perceptron neural networks were used to recognize color and shape patterns. Good preliminary results were achieved, although improvements are still necessary.*

Keywords: *Computer vision, Coffee beans classification, Neural networks, Pattern recognition.*

[*]Autor para contato: pedro.oyama@gmail.com

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 (Bewley & Black, 1994).

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, etc (Brasil, 1992). 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 Krzyzanowski et al. (1991), Marcos Filho (1995) e Vieira (1994) 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 na literatura muitos trabalhos envolvendo classificação de sementes e grãos. Pode-se citar Khatchatourian & Padilha (2008), que utiliza técnicas de visão computacional e redes neurais artificiais para a classificação de sementes de soja baseada na forma. Em MacDougall (2002) é descrito como a análise de cor é incorporada em máquinas utilizadas pela indústria alimentícia para a separação de grãos.

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. Neste 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.

Na Seção 2 são explicados os conceitos teóricos envolvendo cor e forma, e também as redes neurais *Multilayer Perceptron*. A Seção 3 apresenta como as imagens são adquiridas e explica os procedimentos adotados para se chegar à classificação. Os resultados obtidos e comentários sobre estes são mostrados na Seção 4. Finalmente, a Seção 5 apresenta as conclusões e possibilidades futuras para o trabalho.

2. Fundamentação Teórica

2.1 Espaços de cor

Os seres humanos percebem cores através de células sensíveis a luz presentes na retina do olho. Estas células são denominadas cones e bastonetes, sendo os primeiros muito mais eficazes em proporcionar distinção de cores que os últimos, que são mais relacionados à visão noturna por serem mais sensíveis a baixos níveis de iluminação. Existem três tipos de cones, diferenciados pelo comprimento de onda que são capazes de estimulá-los. Os cones do tipo L são sensíveis a comprimentos de onda longos, referentes ao espectro de cores avermelhadas. Os cones de tipo M são sensíveis a médios comprimentos de onda, referentes ao espectro de cores esverdeadas. Finalmente os cones de tipo S são sensíveis a comprimentos de onda curtos, referentes ao espectro de cores azuladas. Assim, as cores que percebemos estão relacionadas ao número de cones sensibilizados de cada tipo. Por exemplo, quando há muitos cones S sensibilizados, e poucos dos outros tipos, vemos uma cor de tonalidade azul.

Esta é a forma como o olho humano capta as cores. Para lidarmos matematicamente ou computacionalmente com elas, algum sistema semelhante deve ser utilizado, e estes sistemas são chamados de espaços de cor. Existem diversos deles na literatura, sendo que suas eficiências são dependentes da aplicação em que estão sendo empregadas. Uma comparação entre vários espaços de cor, especificamente avaliados para a detecção de pele em reconhecimento de face, foi feita por Chaves-González et al. (2010). Neste trabalho foi adotado como primeira alternativa o espaço de cor RGB, por ser um sistema simples, cujas componentes podem ser obtidas diretamente da imagem a ser processada, sem necessidade de conversões.

O modelo RGB trata as cores da mesma forma como o olho humano as capta. As cores são representadas pelas componentes R, G e B, que definem respectivamente intensidade com a qual as cores primárias vermelho, verde e azul estão presentes. Todas as componentes são números inteiros que variam entre 0 e 255.

2.2 Descritores de forma

Um dos desafios da visão computacional é conseguir definir formas de objetos através de representações matemáticas de maneira concisa, e que ainda assim consigam expor informações intrínsecas da forma, as quais as pessoas conseguem perceber naturalmente, como circularidade e excentricidade, para citar alguns exemplos. Essas representações podem ser denominadas de descritores de forma. Inúmeras técnicas já foram apresentadas e suas eficiências dependem muito da aplicação e de quais características da forma se deseja exaltar. Outro critério muito relevante envolvendo os descritores é o custo computacional exigido nos cálculos, que sendo muito alto pode comprometer algumas aplicações.

Neste projeto os descritores utilizados para representar os grãos necessitam atender a duas exigências. A primeira é que eles devem ser invariantes à rotação, isto é, os valores dos descritores de um objeto e os da sua versão rotacionada devem ser os mesmos. A segunda é não ter um custo computacional elevado, visto que a aplicação deve ter a resposta mais rápida possível, por ser tratar de uma solução a ser empregada na indústria. Portanto, inicialmente foram avaliadas as variâncias das assinaturas e as magnitudes dos descritores de Fourier, em conjunto com descritores mais triviais: área, perímetro, comprimento e largura.

2.2.1 Assinaturas

Assinaturas são uma função 1D calculada a partir dos pontos do contorno de uma forma. Geralmente elas são normalizadas para se obter invariância quanto a escala. Existem várias abordagens para calculá-la, como, por exemplo, a tangente do ângulo, a área, comprimento da corda e a distância do centróide (Zhang & Lu, 2004). Esta última foi a utilizada neste trabalho. Vale ressaltar que as assinaturas não são invariantes à rotação, mas como o descritor adotado foi a variância (σ^2) destes valores, a invariância é obtida.

As coordenadas $(\bar{x}, \bar{y})$ do centróide de uma figura representada pela região $\mathcal{R}$ contendo N *pixels* são dadas pelas Equações 1 e 2:

$$\bar{x} = \frac{1}{N} \sum_{(x,y)\in\mathcal{R}} \sum x \tag{1}$$

$$\bar{y} = \frac{1}{N} \sum_{(x,y)\in\mathcal{R}} \sum y \tag{2}$$

Portanto, para um objeto com contorno $\mathcal{C}$, as assinaturas $z(i)$ são calculadas pela Equação 3, com $(x(i), y(i)) \in \mathcal{C}$.

$$z(i) = \sqrt{(x(i) - \bar{x})^2 + (y(i) - \bar{y})^2} \tag{3}$$

2.2.2 Comprimento e largura

Os passos envolvidos no cálculo do comprimento e da largura de uma forma são descritos em Jain (1989).

Os momentos centrais de ordem (p, q) de uma figura representada pela região $\mathcal{R}$ contendo N *pixels* são calculados pela Equação 4.

$$\mu_{p,q} = \sum_{(x,y)\in\mathcal{R}} \sum (x - \bar{x})^p (y - \bar{y})^q \tag{4}$$

A partir dos momentos é possível se calcular a orientação de um objeto, ou seja, o ângulo θ (Figura 1) com o qual ele possui o menor momento de inércia. θ é dado pela Equação 5.

$$\theta = \frac{1}{2} \arctan\left(\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}}\right) \quad (5)$$

Pode-se adotar um sistema alternativo de coordenadas, eixos α e β, que sejam respectivamente paralelo e perpendicular à orientação do objeto, como ilustrado na Figura 1.

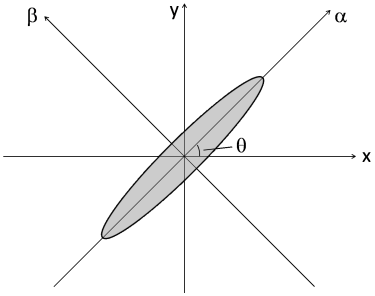


Figura 1. Imagem adaptada de Jain (1989), p.393.

Assim, as coordenadas α e β podem ser calculadas a partir das coordenadas x e y utilizando as seguintes Equações 6 e 7:

$$\alpha = x \cos\theta + y \sin\theta \quad (6)$$

$$\beta = -x \sin\theta + y \cos\theta \quad (7)$$

Encontrando o menor e maior valores de α (α_{min} e α_{max}, respectivamente) e o menor e maior valores de β (β_{min} e β_{max}, respectivamente) dos pontos da região $\mathcal{R}$, pode-se definir o comprimento e a largura da forma:

$$largura = MIN(\alpha_{max} - \alpha_{min}, \beta_{max} - \beta_{min}) \quad (8)$$

$$comprimento = MAX(\alpha_{max} - \alpha_{min}, \beta_{max} - \beta_{min}) \quad (9)$$

2.2.3 Descritores de Fourier

Em 1822 foi publicada por Fourier uma teoria que afirma que qualquer sinal periódico pode ser representado como uma soma de senóides multiplicadas por diferentes coeficientes. Mais tarde esta definição foi estendida para sinais não-periódicos, e estes coeficientes podem ser calculados pela chamada transformada de Fourier (Gonzalez & Woods, 2001). Esta transformada é amplamente utilizada na área de análise de sinais por permitir uma visualização do sinal no domínio da frequência, fornecendo possibilidades não

alcançadas no domínio do tempo. Nas últimas décadas ela também vem sendo muito explorada no processamento de imagens. A versão discreta da transformada (transformada discreta de Fourier) pode ser utilizada na representação de forma em imagens digitais, bastando tratar um conjunto de assinaturas como sendo um sinal e aplicar a transformada. Os coeficientes das senóides obtidas passam então a ser chamados de descritores de Fourier. Características típicas dos descritores de Fourier são que geralmente a forma geral da figura é relativamente bem definida a partir de alguns dos termos de menor ordem da expansão, e as magnitudes dos termos são invariantes à rotação.

Os descritores de Fourier $F(u)$ de um conjunto de N assinaturas $z(i)$ são obtidos pela Equação 10, para $u = 0, 1, 2, ..., N-1$:

$$F(u) = \frac{1}{N} \sum_{i=0}^{N-1} z(i) e^{-j2\pi ui/N} \tag{10}$$

Na Figura 2 foram comparados exemplos de assinatura e magnitudes de descritores de Fourier para contornos de grãos normais e quebrados, mostrando como eles podem ser utilizados para diferenciar os dois tipos de grãos.

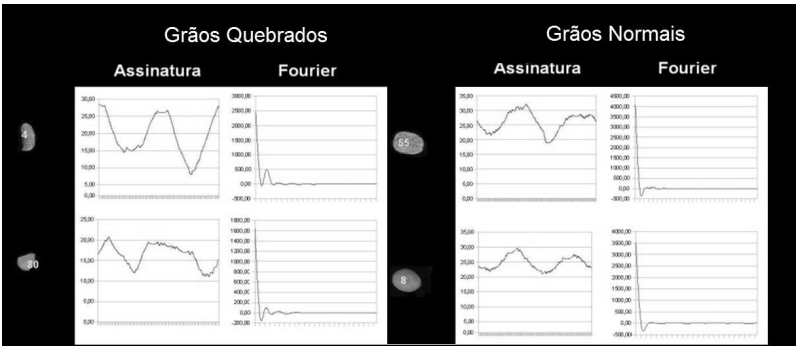


Figura 2. Comparação entre as magnitudes dos descritores de Fourier e as assinaturas de grãos quebrados e grãos normais.

2.3 *Multilayer Perceptron*

Redes neurais artificiais do tipo *Multilayer Perceptron* (MLP) são modelos computacionais inspirados no modelo biológico de interligação entre neurônios, e visam promover o aprendizado de máquina. Elas são largamente utilizadas na classificação de padrões, ou seja, indicar a que classe pertencem os dados de entrada. As MLPs são redes neurais cujo aprendizado é supervisionado. Portanto, antes de realizar qualquer classificação é necessário um processo de treinamento, no qual vários exemplos de dados

pré-classificados são apresentados. As MLPs são constituídas de várias camadas de neurônios artificiais, sendo uma de entrada, por onde dados são inseridos, uma de saída, por onde os resultados são fornecidos, e pelo menos uma camada escondida (Figura 3). O número de camadas escondidas e o número de neurônios em cada uma devem ser definidos de acordo com a aplicação e influenciam na taxa de sucesso da classificação (Silva et al., 2010).

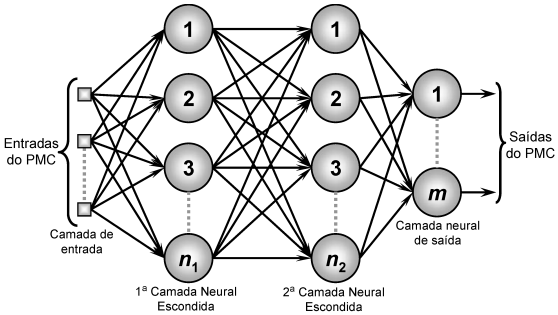


Figura 3. Topologia de uma MLP (Silva et al., 2010).

3. Materiais e Métodos

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.

3.1 Amostragem e classificação

As classes de grãos e impurezas a serem identificadas pelo *software* foram definidas de acordo com o modelo de classificação utilizado pela Cooperativa Regional de Cafeicultores de Guaxupé Ltda (Cooxupé), que forneceu amostras de grãos defeituosos, impurezas e grãos saudáveis de diferentes peneiras. Esse foi o conjunto de amostras utilizado no processo de treinamento das redes neurais, como será detalhado mais adiante. Todas as amostras foram previamente separadas e classificadas por especialistas da cooperativa através do processo manual. As classes de defeitos e de grãos sadios são apresentadas na Figura 4. As impurezas de uma amostra são paus e pedras. O grão do tipo brocado não foi tratado nesta fase do trabalho, por ser caracterizado pela presença de pequenas perfurações, sendo que o processo necessário para identificá-las não coincide com a metodologia empregada no reconhecimento das outras classes.

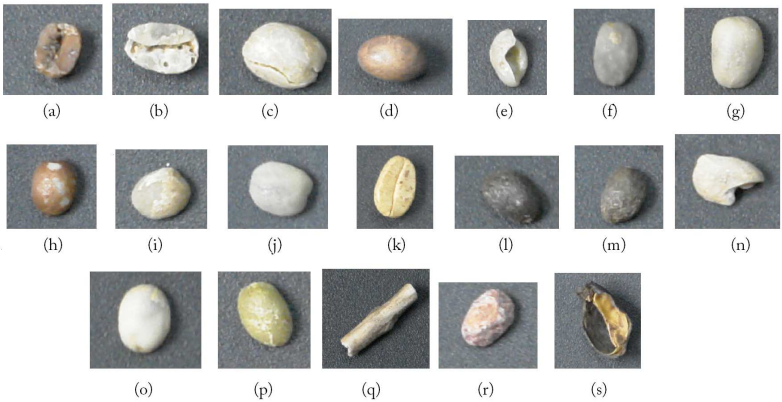


Figura 4. Classes de grãos e impurezas. (a) ardido, (b) brocado, (c) cabeça, (d) coco, (e) concha, (f) fava escura, (g) marago, (h) melado, (i) película, (j) perfeito, (k) pergaminho, (l) preto, (m) preto-verde, (n) quebrado, (o) secador, (p) verde, (q) paus, (r) pedras e (s) cascas.

3.2 Aquisição de imagens

Para se poder capturar imagens de amostras de café por um método sistemático, com o qual todas as condições que afetam a foto pudessem ser reproduzidas exatamente da mesma forma a cada processo de aquisição, uma máquina foi desenvolvida (Figura 5). Ela é composta por uma bandeja horizontal de vidro, onde os grãos são depositados, sob condições controladas de luz. Ambos os lados da bandeja são iluminadas por lâmpadas de LED e duas câmeras de vídeo são posicionadas cerca de 60 cm acima e abaixo dela, de forma que possam ser capturadas imagens de cada um dos lados da amostra, podendo-se analisar quase toda a superfície dos grãos. As imagens superior e inferior são capturadas em momentos distintos, tornando necessária a presença de um dispositivo manual para alternar as duas placas de metal revestidas com a cor azul, utilizadas como fundo das fotos. As câmeras de vídeo utilizadas são do tipo IP e são conectadas ao computador que executa o software através de uma interface *Gigabit Ethernet*. As imagens são capturadas diretamente pelo software e são salvas no formato *bitmap* com resolução de 1024 × 768 *pixels*.

3.3 Processamento

O software desenvolvido foi dividido em módulos: pré-processamento, classificação por cor, por forma e por tipo de grão, conforme descrito no diagrama da Figura 6.

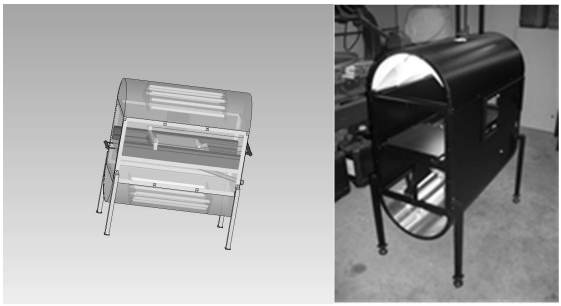


Figura 5. (a) Representação gráfica da máquina de classificação de café, (b) Foto da primeira versão construída.

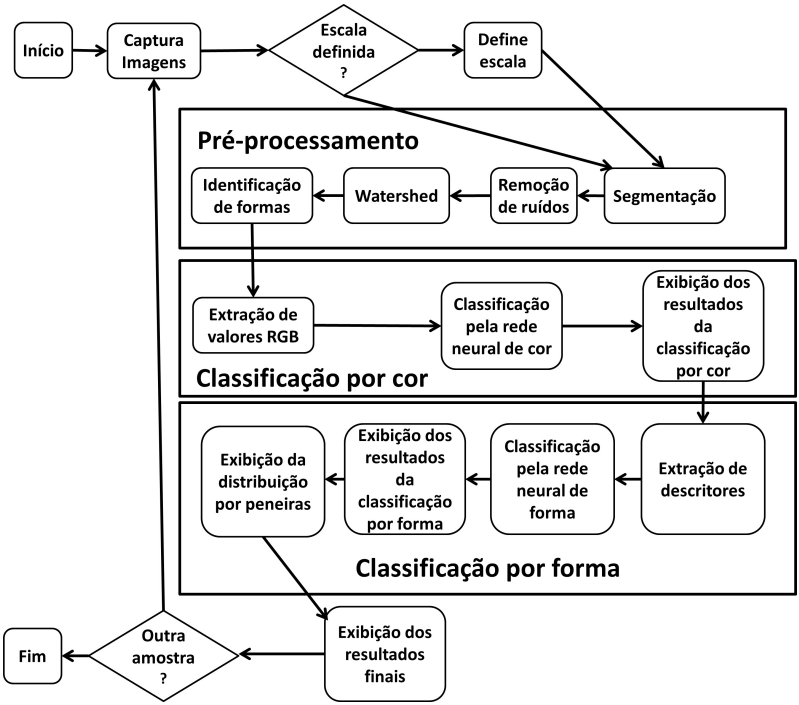


Figura 6. Diagrama de blocos do Qualicafé.

Todos os processos de classificação são feitos por redes neurais artificais do tipo *Multilayer Perceptron* (MLP), utilizando o algoritmo *backpropagation*, com taxa de aprendizado de 0,1, momento de 0,7 e 5000 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. As MLPs continham apenas uma camada escondida, com n neurônios cada, sendo n definido pela Equação 11 (arredondado para baixo).

$$n = \frac{numero\,de\,atributos + numero\,de\,classes}{2} \tag{11}$$

3.4 Pré-processamento

A primeira etapa de processamento é a identificação de grãos e impurezas presentes na amostra, que é feita pela aplicação de tradicionais algoritmos de processamento de imagem.

O fundo de cor azul foi escolhido devido ao fato de nenhuma das classes apresentarem tonalidades próximas a esta cor, de forma a facilitar a segmentação da imagem. A segmentação foi feita comparando-se *pixel* a *pixel* as componentes de cor do modelo RGB. Naturalmente, a cor azul apresenta valores mais expressivos de B e valores mais baixos de R e G. Imagens contendo todos os tipos de grãos e impurezas foram submetidas a algoritmos de segmentação que separam os *pixels* utilizando diferentes expressões que comparam valores de B aos valores de R e G. Os resultados foram comparados visualmente para se eleger a expressão que apresenta a melhor segmentação. Assim sendo, definiu-se que um *pixel* seria considerado como fundo se para dadas suas componentes RGB, B $\geq$ 3,3R E B $\geq$ 3,3G. Caso contrário o *pixel* é considerado objeto. Tendo os objetos sido separados do fundo, a imagem é binarizada para que os outros algoritmos possam ser aplicados. Utiliza-se os processos morfológicos de erosão e dilatação, para suavizar os objetos, eliminar possíveis ruídos e imperfeições da segmentação.

Visando separar grãos que estejam encostados uns nos outros, podendo gerar problemas na contagem de grãos e objetos com formas anormais, um algoritmo de *Watershed* é executado. Para cada *pixel* "objeto" da imagem, sua distância ao *pixel* "fundo" mais próximo é calculada e armazenada. Então são definidos os pontos com as distâncias máximas locais e cada um destes pontos é dilatado até que se chegue à borda do objeto ou à borda da região de outro ponto sendo dilatado. Desta forma, a região de encontro entre os pontos crescentes é definida como "fundo", separando os objetos (Ferreira & Rasband, 2011).

A partir da imagem binarizada é finalmente executado um algoritmo para rotular as diferentes formas. A rotulação é feita pelo método das componentes conexas, o qual agrupa os *pixels* de acordo com as suas conecti-

vidades. Existem duas variantes que definem a conectividade dos *pixels*: a vizinhança-de-4 e a vizinhança-de-8. A primeira considera que dois *pixels* são vizinhos somente se estiverem conectados pelos eixos vertical ou horizontal, enquanto a segunda, adotada neste trabalho, também considera os diagonais. Uma componente conexa é um grupo de *pixels* que compartilham da propriedade de que existe um caminho de ligação (uma sucessão de vizinhos) entre quaisquer dois deles (Costa & Cesar Jr., 2000). Uma vez definidas as componentes conexas, cada uma delas e seus respectivos *pixels* são rotulados, representando um objeto diferente. Depois de rotuladas todas as formas presentes na imagem, é trivial se extrair os *pixels* de seus contornos.

Finalmente, são descartados os objetos compostos por menos que 40 *pixels*, por eles serem muito pequenos para representar um objeto de interesse, provavelmente sendo resultado de ruído da imagem ou uma falha no algoritmo de identificação.

Na Figura 7 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.

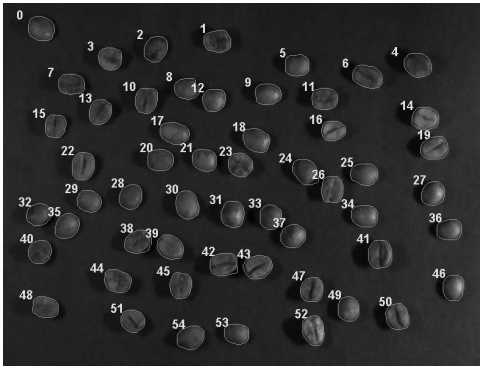


Figura 7. Objetos identificados e rotulados.

3.5 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. Esta 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, casca/coco, fava escura, melado, pau, pedra, película, perfeito, pergaminho, preto, preto-verde, secador e verde (todas ilustradas na Figura 4). A MLP foi definida com três atributos de entrada: as componentes de cor R, G e B e 13 neurônios na camada de saída. Pela aplicação da Equação 11 obteve-se uma camada escondida com oito neurônios.

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 8), ou um padrão de saudável. Para cada um dos *pixels* que compõem estas áreas foi determinado o valor médio de R,G e B, com base nos *pixels* imediatamente vizinhos. Estes valores foram então utilizados como entrada para a MLP no seu treinamento. Para cada classe foram coletados cerca de 1700 *pixels*.

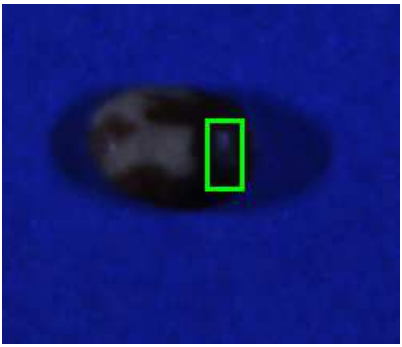


Figura 8. Seleção de amostra em um grão para treinamento da MLP.

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 9. 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.

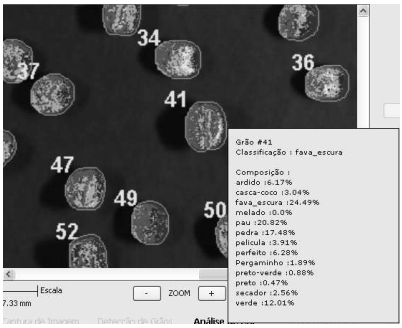


Figura 9. Resultado de uma classificação por cor.

3.6 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. Os paus apresentam um contorno fino e comprido, e as pedras 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 estas premissas foram testadas duas configurações de MLP para reconhecer os padrões de forma dos objetos presentes na imagem de uma amostra e indicar de qual classe de grão ou impureza estes padrões são característicos. As MLPs se diferenciam pelos descritores de forma utilizados como entrada. As classes definidas foram nove, a saber: cabeça, casca, coco, concha, marago, pau, pedra, perfeito, e quebrado (apresentadas da Figura 4). Nesta etapa do projeto somente uma das duas imagens (superior e inferior) foi utilizada na análise.

A primeira configuração de MLP foi definida com 44 atributos, dos quais 40 são magnitudes dos descritores de Fourier, os restantes são largura, comprimento e perímetro (em milímetros), e o último é área em milímetros quadrados. Sendo assim, pela Equação 11 a camada escondida foi definida com 26 neurônios.

A segunda MLP foi configurada com cinco atributos: variância(σ^2) das assinaturas, largura, comprimento e perímetro em milímetros e área em milímetros quadrados. A variância das assinaturas foi adota com o intuito de se identificar as irregularidades do contorno do grão. Visto que

um grão arredondado e com poucas imperfeições no contorno apresenta menores variâncias em relação a grãos quebrados e irregulaes. O número de neurônios na camada escondida calculada pela Equação 11 foi de sete.

Analogamente ao treinamento da MLP de cor, o das redes 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 dos descritores. Foi coletado um conjunto de cerca de 160 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 (Brasil, 2003).

4. Resultados e Discussão

4.1 Identificação dos objetos

Constatou-se que o processo de reconhecimento de formas na imagem apresenta algumas falhas. O problema mais recorrente ocorre na execução do algoritmo de *Watershed*, que algumas vezes divide a forma de um único objeto em várias (o que ocorre com frequência com os paus), como ilustrado na Figura 10 e em certas ocasiões traça erroneamente a fronteira entre dois objetos que se encostam.

Na identificação e extração de contorno de 2766 objetos, incluindo grãos e impurezas de todas as classes, constatou-se apenas 48 falhas, ou seja, houve uma precisão de 98,45%. Dessas 43 falhas, notou-se que 18 (41,86%) são referentes à segmentação do fundo azul, com sombras sendo identificadas como partes dos objetos, e 25 (58,14%) são resultantes da execução do *Watershed*. Destas, 20 (80%) deram-se em objetos da classe pau. Outras pequenas falhas na extração de contorno puderam ser notadas em algumas ocorrências, mas estas foram consideradas insignificantes e foram ignoradas. Deve-se salientar que procurou-se sempre posicionar os objetos para a captura de imagem de modo que eles não se encostassem, deixando o algoritmo de *Watershed* somente para eventuais ocorrências.

A taxa de acerto de 98,45% indica que o algoritmo de reconhecimento de objetos se mostrou muito eficiente, apresentando precisão adequada para a aplicação.

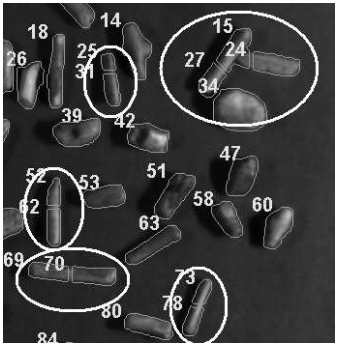


Figura 10. Deficiências na identificação de paus.

4.2 Classificação por cor

Com a rede neural de cor devidamente treinada, alguns testes foram conduzidos para se avaliar a eficiência da classificação. Amostras de grãos e impurezas de todas as classes foram submetidas ao processo de classificação por cor. A matriz de confusão resultante deste conjunto de experimentos é apresentada na Tabela 1.

Tabela 1. Matriz de confusão da classificação por cor, onde: a=ardido, b=casca/coco, c=fava escura, d=melado, e=pau, f=pedra, g=pelicula, h=perfeito, i=pergaminho, j=preto-verde, k=preto, l=secador, m=verde, T.A.=taxa de acerto.

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**T.A.**
a	31	44	0	0	100	5	0	0	0	1	28	0	8	14,29%
b	5	131	0	8	25	2	0	0	0	3	16	0	0	68,95%
c	4	0	83	0	35	36	0	0	0	0	0	0	0	52,53%
d	0	23	5	69	110	14	0	0	1	0	0	0	1	30,94%
e	0	1	0	0	166	39	0	0	0	0	0	0	0	80,58%
f	0	1	0	0	14	135	0	0	0	0	0	0	0	90,00%
g	0	0	18	0	88	63	7	0	0	0	0	1	1	3,93%
h	1	0	65	0	60	24	3	0	0	0	0	0	42	0,00%
i	0	4	0	4	13	0	1	0	151	0	0	0	0	87,28%
j	22	93	0	0	18	3	0	0	0	8	57	0	0	3,98%
k	18	43	0	0	7	2	0	0	0	6	151	0	0	66,52%
l	0	0	26	1	93	38	1	0	2	0	0	4	7	2,33%
m	9	2	13	0	25	9	0	0	0	0	0	0	102	63,75%

A taxa de acerto global foi de 43,47%, a qual não pode ser considerada um bom resultado. Avaliando as taxas de acerto por classe, percebe-se que existem resultados muito bons para casca/coco, pau, pedra, pergaminho e

preto, resultados medianos para fava escura, melado e verde e resultados muito ruins para ardido, película, perfeito, preto-verde e secador.

Espera-se que ao se cruzar os dados dos resultados das duas classificações – por cor e por forma – muitas das baixas taxas de acerto possam ser melhoradas. Como exemplo, pode se perceber que há um grande número de ocorrências de falsos positivos das classes pau e pedra, mas elas apresentam um alto número de verdadeiros positivos, e, como será mostrado na Seção 4.3, na análise de forma estas classes apresentam boas taxas de acerto e poucos falsos positivos. Portanto, um objeto só seria classicado como pau ou pedra, se ambas as análises indicassem tal, o que diminuiria a ocorrência de falsos positivos.

Mesmo com as prováveis melhorias resultantes do cruzamento das classificações, o resultado obtido encoraja a exploração de outros espaços de cor, tendo em vista a possibilidade de eles exaltarem informações que o sistema RGB não é capaz, ou ainda, de serem mais tolerantes às variações de luminosidade presentes na bandeja onde as amostras são depositadas, problema este que pode ter afetado negativamente o resultado.

4.3 Classificação por forma

As duas MLPs propostas para a classificação de forma foram avaliadas utilizando o método da validação cruzada com 10 *folds*, sendo destinadas 80% das amostras para o treinamento e 20% para a validação.

A matriz de confusão originada pela rede neural que utiliza os descritores de Fourier é mostrada na Tabela 2 e a matriz obtida pela rede neural que utiliza a variância das assinaturas é apresentada na Tabela 3. Constata-se que a segunda abordagem apresentou melhores resultados para todas as classes, conseguindo uma taxa de acerto global de 54,5%, a qual é consideravelmente maior que os 40,8% conseguidos pela primeira. Além disto, a primeira MLP conta com muito menos neurônios, tanto na camada de entrada como na camada escondida, proporcionando um custo computacional na geração das classificações muito menores.

Analisando a Tabela 3 percebe-se que as principais classes responsáveis por limitar a eficiência da classificação são cabeça, coco e quebrado. A classe cabeça apresenta uma confusão muito grande com a classe perfeito, por elas apresentarem uma semelhança muito grande no contorno. A classe quebrado foi muito confundida com a classe concha, classes estas que apresentam visualmente contornos muito semelhantes, por se tratarem grãos lascados. Assim, conclui-se que o conjunto de descritores não são muito eficientes em distinguir pequenas diferenças, sendo necessário, ou substituí-los ou adicionar mais alguns que sejam capazes de enaltecer outras particularidades que definem a forma dos grãos.

Tabela 2. Matriz de confusão da classificação por forma utilizando os descritores 40 coeficientes de Fourier, perímetro, área, comprimento e largura, onde: a=cabeça, b=casca, c=coco, d=concha, e=marago, f=pau, g=pedra, h=perfeito, i=quebrado, T.A.=taxa de acerto.

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**T.A.**
a	45	8	18	8	12	0	0	65	3	28,3%
b	3	77	18	18	5	15	6	11	12	46,7%
c	8	13	31	42	11	0	19	13	14	20,5%
d	2	21	16	76	1	6	0	10	30	46,6%
e	13	4	3	2	88	0	0	21	0	67,2%
f	0	14	0	6	3	119	1	0	13	76,3%
g	0	5	5	7	0	0	115	0	16	77,7%
h	11	5	6	0	7	0	0	166	0	85,1%
i	0	8	20	43	0	7	17	1	55	36,4%

Tabela 3. Matriz de confusão da classificação por forma utilizando os descritores variância das assinaturas, perímetro, área, comprimento e largura, onde: a=cabeça, b=casca, c=coco, d=concha, e=marago, f=pau, g=pedra, h=perfeito, i=quebrado, T.A.=taxa de acerto.

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**T.A.**
a	45	8	18	8	12	0	0	65	3	28,3%
b	3	77	18	18	5	15	6	11	12	46,7%
c	8	13	31	42	11	0	19	13	14	20,5%
d	2	21	16	76	1	6	0	10	30	46,6%
e	13	4	3	2	88	0	0	21	0	67,2%
f	0	14	0	6	3	119	1	0	13	76,3%
g	0	5	5	7	0	0	115	0	16	77,7%
h	11	5	6	0	7	0	0	166	0	85,1%
i	0	8	20	43	0	7	17	1	55	36,4%

4.4 Classificação por 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 11 mostra o histograma obtido da amostra da peneira 17.

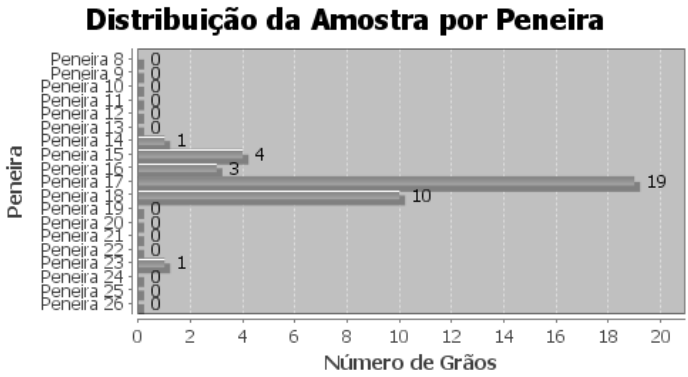


Figura 11. Distribuição por peneira de uma amostra de peneira 17.

4.5 Distinção entre grãos e impurezas

Comparando-se as classificações geradas pelo software 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, esta característica indica um importante resultado. Na Figura 12 é mostrado um exemplo de classificação, na qual esta distinção é notada. Em branco estão os objetos corretamente classificados (como paus, pedras ou grãos) em pelo menos umas das duas análises, de cor e de forma, e em preto aqueles cujas classificações não foram corretas em nenhuma delas.

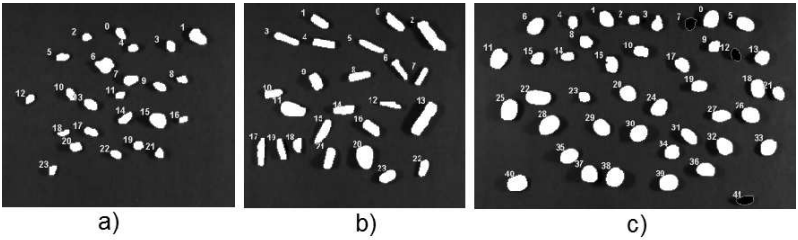


Figura 12. Acertos de classificação de grãos e impurezas a) pedras b) paus c) grãos.

Uma tela da interface do software é mostrada na Figura 13. Ela apresenta os resultados da classificação de uma amostra.

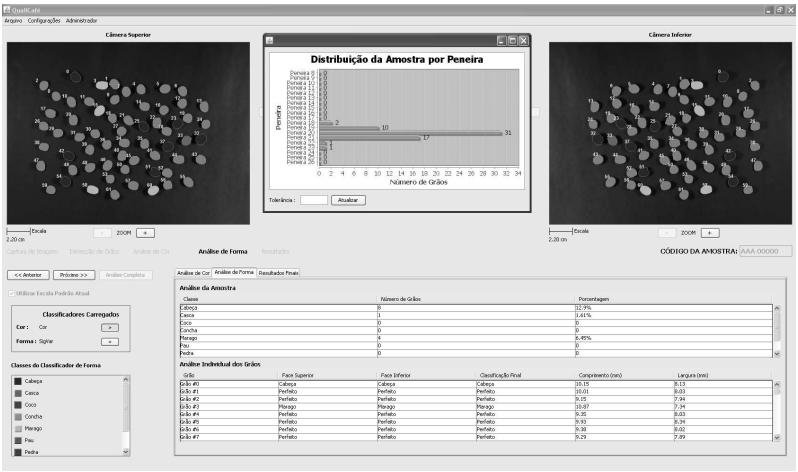


Figura 13. Tela de resultados.

5. 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.

A taxa de acerto das classificações ainda é insatisfatória, mas pode-se considerar que os resultados preliminares são bons – distinguir grãos de impurezas e identificar certas classes com um acerto significativo – visto que as técnicas utilizadas são triviais na área de processamento de imagem e visão computacional, sugerindo que melhores resultados podem ser conseguidos com técnicas mais sofisticadas.

Algumas possibilidades futuras para o trabalho descrito são explorar outros espaços de cor e descritores de forma e se adicionar uma etapa de análise por textura, com o objetivo de se capturar informações que não podem ser avaliadas por cor ou forma. Também se faz necessário se desenvolver uma técnica para cruzar as informações das duas classificações de modo a otimizar o resultado da classificação final.

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.

Referências

Bewley, J.D. & Black, M., *Seeds: physiology of development and germination*. 2a edição. New York, USA: Plenun Press, 1994.

Brasil, , *Regras para análise de sementes*. Brasília, DF: Ministério da Agricultura e Reforma Agrária, SNDA/DNDV/CLAV, 1992.

Brasil, , Ministério da Agricultura, Pecuária e Abastecimento. *Instrução Normativa N° 8*, de 11 de junho de 2003. Diário Oficial da União. Sessão I. 2003. 20 ago.

Chaves-González, J.M.; Vega-Rodríguez, M.A.; Gómez-Pullido, J.A. & Sánchez-Pérez, J.M., Detecting skin in face recognition systems: a colour space study. *Digital Signal Processing*, 20(3):806–823, 2010.

Costa, L.F. & Cesar Jr., R.M., *Shape Analysis and Classification: Theory and Practice*. Boca Raton, USA: CRC Press, 2000.

Ferreira, T. & Rasband, W.. *The ImageJ User Guide* 1.44, 2011. p. 111–112. Http://imagej.nih.gov/ij/docs/user-guide.pdf.

Gonzalez, R.C. & Woods, R.E., *Digital Image Processing*. 2a edição. Upper Saddle River, USA: Prentice Hall, 2001.

Jain, A.K., Fundamentals of Digital Image Processing, Englewood Cliffs, USA: Prentice Hall. p. 392–394.

Khatchatourian, O. & Padilha, F.R.R., Reconhecimento de variedades de soja por meio do processamento de imagens digitais usando redes neurais artificiais. *Engenharia Agrícola*, 28(4):759–769, 2008.

Krzyzanowski, F.C.; França Neto, J.B. & Henning, A.A., Relato de testes de vigor disponíveis para as grandes culturas. *Informativo ABRATES*, 1(2):15–50, 1991.

MacDougall, D.B. (Ed.), *Colour in food: Improving quality*. Boca Raton, USA: CRC Press, 2002.

Marcos Filho, J., Utilização de testes de vigor em programas de controle de qualidade de sementes. *Informativo ABRATES*, 4(2):59–61, 1995.

Silva, I.N.; Spatti, D.H. & Flauzino, R.A., *Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas: Um Curso Prático*. Porto Alegre, RS: Artliber, 2010.

Vieira, R.D., *Testes de vigor em sementes*. Jaboticabal, SP: FUNEP/UNESP, 1994.

Zhang, D. & Lu, G., Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.

Notas Biográficas

Pedro Ivo de Castro Oyama é graduado em Engenharia de Computação (UFSCar, 2010) e atualmente é mestrando em Engenharia da Elétrica na USP São Carlos.

Lúcio André de Castro Jorge é graduado em Engenharia Elétrica (Faculdade de Engenharia de Barretos, 1987), mestre em Ciência da Computação (USP São Carlos, 2001) e doutor em Engenharia Elétrica na área de processamento digital de imagens (USP São Carlos, 2011). Atualmente é pesquisador da EMBRAPA Instrumentação em São Carlos.

Evandro Luis Linhari Rodrigues é graduado e mestre em Engenharia Elétrica (Escola de Engenharia de Lins, 1983 e Universidade de São Paulo, 1992), e doutor em Física (Universidade de São Paulo, 1998). Atualmente é professor da Universidade de São Paulo no Departamento de Engenharia Elétrica.

Carlos Cesar Gomes é analista de qualidade na Cooperativa Regional de Cafeicultores de Guaxupé Ltda. (Cooxupé)

Capítulo 8

Diferenciação do *Greening* de Outras Doenças Foliares em Citros Utilizando Técnicas de Processamento de Imagens

Patricia Pedroso Estevam Ribeiro*, Lúcio André de Castro Jorge e Maria Stela Veludo de Paiva

Resumo: O *greening* ou *huanglongbing* (HLB) é considerada atualmente uma das mais graves doenças dos citros no Brasil. Não possuindo cura ou tratamento, o controle da doença é realizado atualmente por meio de análise de PCR e por método visual por especialistas. Este trabalho utiliza técnicas de processamento de imagens para ajudar a diferenciar o *greening* de outras doenças. Isto é feito por segmentação da cor da imagem de folhas e classificação através de uma rede neural artificial (RNA) do tipo *Perceptron* Multicamada (PMC) com descritores de forma. Os resultados mostram que apenas a mancha amarela não é um diferencial forte desta doença.

Palavras-chave: greening, escala diagramática, métodos de segmentação por cor.

Abstract: *The greening or huanglongbing (HLB) is now considered one of the most serious diseases of citrus in Brazil. It has no cure or treatment and disease control is currently carried out by PCR analysis and visual method by experts. This paper uses image processing techniques to differentiate the greening from other diseases. This is done by color image segmentation of the leaves and further classification by means of a Multilayer Perceptron (MLP) artificial neural network (ANN) using shape descriptors. Results show that only the yellow spot is not a strong spread of this disease.*

Keywords: *greening, scale, segmentation methods by color.*

*Autor para contato: patriciapedrosoestevam@hotmail.com

1. Introdução

No Brasil, a produção de laranjas tanto para suco como para o consumo *in natura* vem crescendo em todo o país, sendo um dos maiores exportadores de suco de laranja, destacando o estado de São Paulo como sendo o responsável por 80% da produção de laranjas em 565 mil hectares de área cultivada (Agrianual, 2008). Para os anos 2009 a 2010, a produção estimada foi de 318,6 milhões de caixas de 40,8 kg, sendo que deste montante 83,4% foi destinada à indústria e 16,6% para o consumo[1].

Apesar de sua eficiência e capacidade de produção, 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, ocasionando uma perda de 10% da produção média nos últimos anos[2].

Dentre estas doenças, o *greening*, também conhecido como *huanglongbing* (HLB), é considerado atualmente a mais grave doença dos citros no mundo (Bové, 2006). Ela é causada pela bactéria *Candidatus* Liberibacter spp., e é transmitida pelo inseto psilídeo *Diaphorina citri*, que adquire e transmite a bactéria às demais plantas ao se alimentar de uma planta já contaminada (FUNDECITRUS, 2009).

Por não possuir cura ou tratamento e por não existir variedade comercial de copa ou porta-enxerto resistente à doença, o controle do *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*) (Innis et al., 1990), 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 (Belasque et al., 2009).

O sintoma característico do *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 do *greening* (FUNDECITRUS, 2009)

[1] Divulgação da safra paulista de laranja 2009/2010: http://www.iea.sp.gov.br/out/LerTexto.php?codTexto=12002

[2] http://www.ripa.com.br/index.php?id=1823

(a) (b)

Figura 1. Inspeções visuais realizadas em pomares de citros, em a) realizada a pé, b) realizada por plataformas movidas por tratores (FUNDECITRUS, 2009).

como, por exemplo, as doenças CVC (Clorose Variegado dos Citros), Rubelose e as deficiências de Magnésio, Manganês e Zinco. Isto causa confusão nos inspetores na identificação destas doenças pelo método visual. Nos experimentos apresentados por Nutter Jr. & Schultz (1995), Martins et al. (2004) e Kowata et al. (2008) 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. (2000) utilizam técnicas de visão computacional para inspeção da qualidade de produtos. Sposito (2004) elaborou uma escala diagramática para quantificar área foliar lesionada por doença a partir de processamento e análise de imagens e Yonekawa et al. (1996) verificam que os fatores de forma são úteis para a identificação de plantas por meio de suas folhas.

Este trabalho tem como objetivo aplicar técnicas de processamento de imagens, permitindo analisar imagens de folhas sintomáticas digitalizadas, por descritores de cor e forma, quantificando a severidade das manchas para auxiliar na diferenciação do *greening* e outras doenças em citros, possibilitando um diagnóstico mais rápido e preciso.

O capítulo está organizado da seguinte maneira: a Seção 2 aborda a teoria relacionada ao trabalho, na Seção 3 são apresentados os materiais e métodos, na Seção 4 são mostrados os resultados e a discussão, e finalmente na Seção 5 apresenta-se a conclusão.

2. Fundamentação Teórica

2.1 Extração de atributos

2.1.1 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, encontra-se na literatura inúmeras aplicações com o uso de redes neurais artificiais (RNA) com bons resultados (Simões, 2000; Simões & Reali Costa, 2000). Em Cavani et al. (2006) e Simões et al. (2001) é utilizada segmentação por cor em imagens de frutas, utilizando-se uma RNA do tipo *Perceptron* multicamada (PMC) (Silva et al., 2010).

2.1.2 Descritores de forma

Em processamento de imagens, a aplicação de descritores de forma permite analisar e extrair características e parâmetros dos objetos da imagem. Para esta finalidade, nas Seções 2.1.2.1 até 2.1.2.5 foi utilizado o software Image Pro-Plus[3].

2.1.2.1 Razão de aspecto

Razão de aspecto é a razão entre o eixo maior e eixo menor do objeto selecionado como, por exemplo, 4:3 ou 16:9, podendo ser observado na Figura 2, e implementado pela Equação 1 (Russ, 1998).

Figura 2. Ilustração da região representada pela razão de aspecto.

$$RazaodeAspecto = \frac{DiametroMaximo}{DiametroMinino} \quad (1)$$

[3] http://www.mediacy.com/

2.1.2.2 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.

Figura 3. Ilustração da região representada pelo diâmetro médio.

2.1.2.3 Razão do raio (*radius ratio*)

É a razão entre o raio máximo e o raio mínimo, em relação aos pontos centrais do objeto, como mostrado na Figura 4.

Figura 4. Ilustração da região representada pela razão do raio.

2.1.2.4 Aspecto arredondado (*roundness*)

Redondeza mede o a espessura média do objeto selecionado, como mostrado na Figura 5, Equação 2.

Figura 5. Ilustração da região representada redondeza.

$$AspectoArredondado = \frac{Perimetro^2}{4.\pi Area} \quad (2)$$

2.1.2.5 *Feret* médio

O diâmetro *feret* é a medida que caracteriza o tamanho do objeto selecionado (Russ, 1998), sendo a média dos *ferets* em várias direções. Pode ser observado na Figura 6.

Figura 6. Ilustração da região representada pelo diâmetro *feret*.

2.2 Quantificação de doenças

A escala diagramática é um método utilizado para quantificar uma doença, permitindo avaliar o grau de severidade e intensidade da lesão, mancha ou doença em uma folha. Desta maneira, pode-se providenciar a melhor medida de ação para o controle e combater a doença avaliada. Para o desenvolvimento das escalas diagramáticas devem ser observados os aspectos do limite superior e inferior da escala, que devem ser iguais à intensidade real de doença no campo. As subdivisões das escalas devem ser proporcionais ao logaritmo da intensidade do estímulo, respeitando as limitações da acuidade visual humana e baseando-se na lei de *Weber-Fechner* (Amorim, 1995), como se pode observar na Figura 7.

3. Materiais e Metodologia

A Figura 8 apresenta as etapas de processamento aplicadas às imagens foliares, sendo estas: aquisição, segmentação por cor, extração dos descritores de forma, divisão em quadrantes e classificação. A classificação visa a obtenção de um classificador para diferenciar as manchas amarelas das amostras das doenças e deficiências cujos sintomas mais se assemelham ao *greening*.

3.1 Aquisição dos dados

Foram fornecidas pela empresa Fischer[4] seis tipos de folhas de citros com sintomas de doenças e deficiência nutricional, conforme a Figura 9. Es-

[4] http://www.citrosuco.com.br/fischer/fischer/sites/fischer/citrosuco/home/home.html

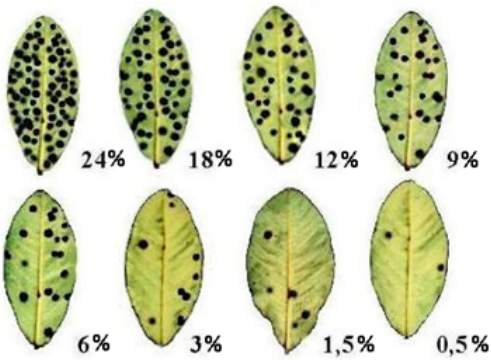


Figura 7. Exemplo de uma escala diagramática para mancha preta do amendoim. Fonte: (de Moraes, 2007).

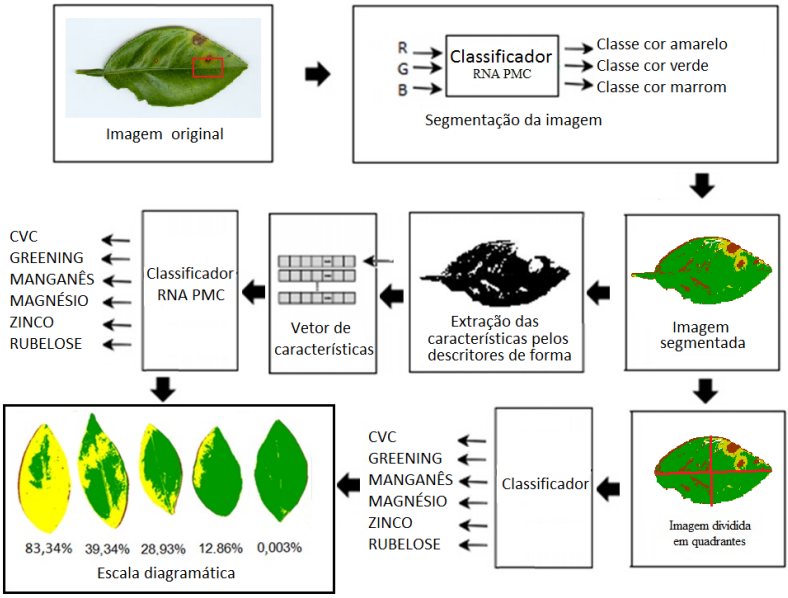


Figura 8. Etapas de processamento: aquisição, segmentação por cor, divisão por quadrante, classificação das doenças, extração de atributos e classificação das doenças com as informações extraidas do vetor de características e geração da escala diagramática.

tas amostras foram selecionadas por um técnico agrônomo denominado pragueiro que identificou os sintomas apenas por meio da inspeção visual, segundo instruções do manual técnico de *greening*, fornecido pela Fundecitrus (FUNDECITRUS, 2009).

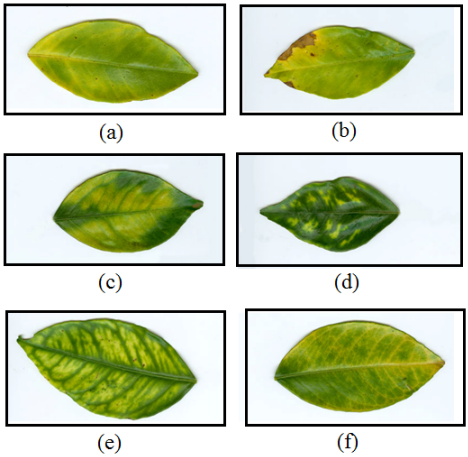


Figura 9. Amostras foliares com a) *Greening*, b) CVC, (c) Deficiência de Magnésio, (d) Deficiência de Manganês, (e) Deficiência de zinco, (f) Rubelose.

Conforme apresentado na Figura 9, 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 e 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 com 96,04% de acurácia 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 sigmoidal. Os parâmetros de entrada da RNA foram as 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 observado na Figura 10(a).

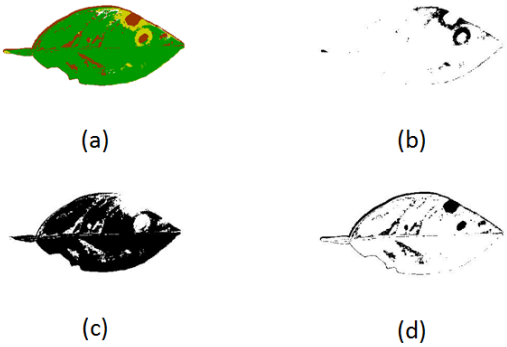


Figura 10. Apresenta as imagens das folhas segmentada por cor, em três classes de cores: a) imagem segmentada, b) somente amarelo, c) somente verde, d) somente marrom.

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 Figura 10(b), 10(c), 10(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 2.1.2. Os valores médio, máximo e mínimo são demonstrados pelas Equações 3, 4 e 5, como se segue:

$$Media = \frac{1}{n}\sum_{i=0}^{n} f(x_i) \tag{3}$$

$$Maximo = Maxf(x_i) \tag{4}$$

$$Minimo = Minf(x_i) \tag{5}$$

Na Equação 3, n corresponde ao número de objetos ou manchas da imagem binarizada $f(x)$. Na Equação 4, o valor Máximo representa o maior valor dentre os objetos ou manchas encontrados na imagem binarizada $f(x)$. 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. Descrição de cada vetor.

Vetor	Cor	Atributos	Instância
1º	Amarelo	15	60
2°	Verde	15	60
3º	Marrom	15	60
4º	Amarelo+Verde	30	60
5º	Amarelo+Verde+Marrom	45	60
6º	Amarelo+Verde+Marrom	45	50
7º	Amarelo+Verde+Marrom	45	40
8º	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 Divisão em quadrantes

Devido às manchas de *greening* apresentarem a cor amarela assimetricamente, conforme mencionado na Seção 3.1, foi proposta a divisão da folha em quadrantes para se determinar a porcentagem de cada cor utilizada em cada quadrante. Para a divisão da folha em quadrantes, foi determinado o o centro de massa nas coordenadas (xCM, yCM) e os pontos delimitadores do momento central (α_{min}, α_{max}, β_{min} e β_{max}), conforme apresentado na Figura 11.

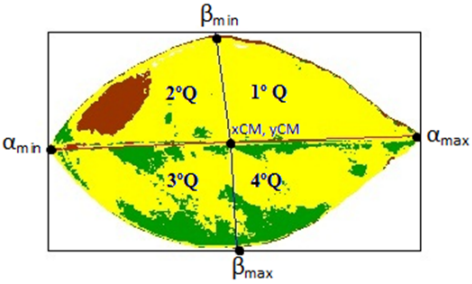


Figura 11. Imagem segmentada em quadrantes.

O cálculo do centro de massa em relação a uma região R é dado pelas Equações 6 e 7 (Jain, 1989).

$$xCM = \frac{1}{N_t} \sum_{i=0 \in R}^{x-1} X_i \tag{6}$$

$$yCM = \frac{1}{N_t} \sum_{i=0 \in R}^{y-1} Y_i \tag{7}$$

O parâmetro N_t é o número total de *pixels* dentro da região de interesse da coordenada (x, y). Para o cálculo dos pontos delimitadores do momento central (α_{min}, α_{max}, β_{min} e β_{max}), inicialmente determina-se os valores de $\mu_{p,q}$ (momentos centrais) pela Equação 8, variando-se *(p,q)* entre [0, 2], e o valor de θ_{CM} (ângulo do momento central) dado pela Equação 9 (Jain, 1989).

$$\mu_{p,q} = \sum_{i=0}^{x-1} \sum_{i=0}^{y-1} (X_i - xCM)^p (Y_i - yCM)^q \tag{8}$$

$$\theta_{CM} = \frac{1}{2} tan^{-1} \left[\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \right] \tag{9}$$

$$\alpha_b = X \cdot cos \cdot \theta_{CM} + Y \cdot sin \cdot \theta_{CM} \tag{10}$$

$$\beta_b = -X \cdot sin \cdot \theta_{CM} + Y \cdot cos \cdot \theta_{CM} \tag{11}$$

Determinado o θ_{CM}, usam-se as Equações 10 e 11 para determinar os mínimos e máximos pontos das coordenadas (X, Y) da imagem, ficando assim definidos os quatro pontos delimitadores do momento central (α_{min}, α_{max}, β_{min} e β_{max}), permitindo, desta maneira, dividir a imagem em quadrantes.

3.5 Classificação

Para a classificação de padrões foi utilizada a API (*Application Programming Interface*) do *toolbox* WEKA[5], para construção do classificador foi utilizado uma RNA do tipo *Perceptron* multicamada (PMC) com o algoritmo *backpropagation* (Witten & Frank, 2005).

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 do *greening* que apresenta as manchas amarelas assimétricas nas folhas, foi aplicada uma RNA. Os parâmetros da mesma foram ajustados com base nos testes realizados previamente. Desta forma, para a RNA foi utilizado uma camada de entrada, três camadas de neurônios escondidos 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, sendo que deste total o classificador conseguiu classificar corretamente trinta e duas instâncias para o 1° vetor, representando um percentual de 53,33%, vinte e seis para o 2° vetor (43,33%), dezesseis para o 3° vetor (26,67%), trinta e três para o 4° vetor (55%) e trinta e oito para o 5° vetor(63,33%). Desta forma é possível observar que apenas a distribuição de uma única cor na folha não é o suficiente para diferenciar o *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.

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 o *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

[5] http://www.cs.waikato.ac.nz/ml/weka

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%
2°	Verde	37,87 %	20,46 %	43,33%
3º	Marrom	42,47%	24,70 %	26,67%
4º	Amarelo+Verde	33,74 %	15,36 %	55,00%
5º	Amarelo+Verde+Marrom	31,59 %	14,84 %	63,33%

Tabela 3. Valores de precisão obtidos na classificação por RNA para cada classe de cor em relação às doenças do 1° ao 5° vetor, onde A=CVC, B=Magnésio, C=Manganês, D=Zinco, E=*Greening*, F=Rubelose.

		Precisão					
Vetor	Cor	A	B	C	D	E	F
1º	Amarelo	50%	70%	58,30%	50%	41,70%	50%
2°	Verde	25%	44,4%	41,7%	60%	36,4%	50%
3º	Marrom	37,5%	0%	30%	30%	31,3%	28,6%
4º	Amarelo+Verde	71,4%	42,9%	60%	58,3%	57,10%	50%
5º	Amarelo+Verde+Marrom	63,6%	57,10%	63,60%	69,2%	50%	70%

o *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.

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 30 instâncias analisadas do 8º vetor, 25 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%.

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

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 (%)
6º	Amarelo+Verde+Marrom	32,15%	15,23 %	62%
7º	Amarelo+Verde+Marrom	35,21%	17,57 %	70%
8º	Amarelo+Verde+Marrom	31,75%	15,39 %	83,33%

Tabela 5. Valores de Precisão obtidos na classificação por RNA para cada classe de cor em relação às doenças do 6° ao 8° vetor, onde: A=CVC, B=Magnésio, C:Manganês, D=Zinco, E=*Greening*, F=Rubelose.

		Precisão					
Vetor	Cor	A	B	C	D	E	F
6º	Amarelo+Verde+Marrom	66,7%	60%	63,6%	57,1%	63,6%	–
7°	Amarelo+Verde+Marrom	75%	–	58,3%	87,50%	66,7%	–
8º	Amarelo+Verde+Marrom	–	100%	–	80%	72,7%	–

o sétimo vetor 87,50% e para o oitavo vetor foi de 80%. Para o *Greening* a precisão do sexto vetor foi de 63,60%, para o sétimo vetor 66,70% e para o oitavo vetor 72,70%. Desta forma, observa-se que o menor número de doenças melhora os resultados obtidos com o classificador.

4.3 Escala diagramática

Para calcular os níveis de severidade de uma escala diagramática serão considerados os valores de máximo e de mínimo de proporção de área foliar manchada com a cor amarela, como os limites da escala diagramática. Serão utilizados cinco níveis de severidade das doenças analisadas, conforme padrão adotado em campo pela empresa Fischer. Foi gerada manualmente uma escala diagramática para cada tipo de doença baseando-se nos resultados preliminares obtidos com a RNA PMC. Estes resultados são apresentados nas Figuras 12 a 17.

A Figura 12 apresenta a escala diagramática da doença CVC, com os seguintes níveis de severidade: 56,69%, 43,10%, 27,86%, 13,12% e 0,07%. Observa-se nas folhas a presença de pequenas manchas amareladas e irregulares, e lesões na cor marrom.

A Figura 13 apresenta a escala diagramática para a deficiência de Magnésio, com os seguintes níveis de severidade: 57,27%, 42,31%, 28,87%, 13,02% e 0,23%. Observa-se nas folhas o amarelecimento em forma de "V" invertido.

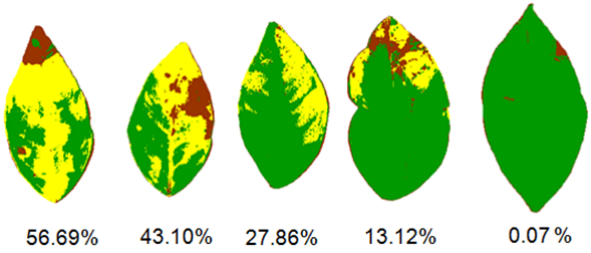


Figura 12. Escala diagramática do CVC.

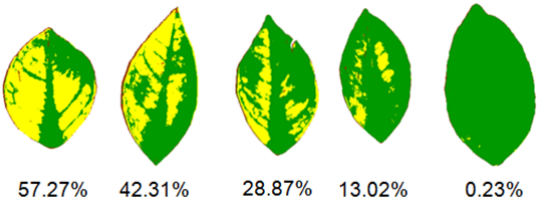


Figura 13. Escala diagramática da deficiência de Magnésio.

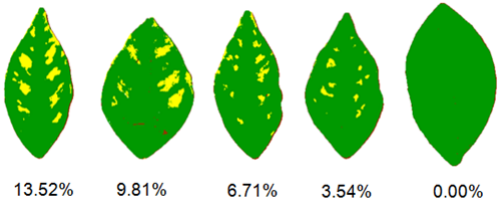


Figura 14. Escala diagramática da deficiência de Manganês.

A Figura 14 apresenta a escala diagramática para a deficiência de Manganês, com os seguintes níveis de severidade: 13,52%, 9,81%, 3,54%, 6,71% e 0,0%. Observa-se nas folhas que as manchas são menores entre as nervuras, sendo menos acentuadas que na deficiência de Magnésio, além de serem distribuídas de uma forma mais simétrica.

A Figura 15 apresenta a escala diagramática para a deficiência de Zinco, com os seguintes níveis de severidade: 74,07%, 54,24%, 35,43%, 17,56% e 3,05%. Observa-se que as folhas apresentam clorose acentuada do limbo entre as nervuras.

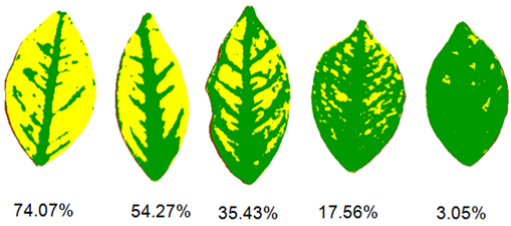


Figura 15. Escala diagramática da deficiência de Zinco.

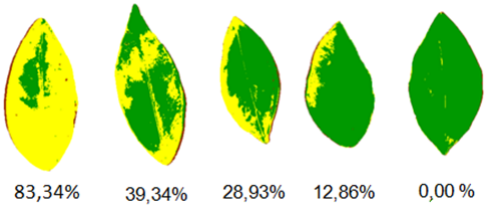


Figura 16. Escala diagramática *greening*.

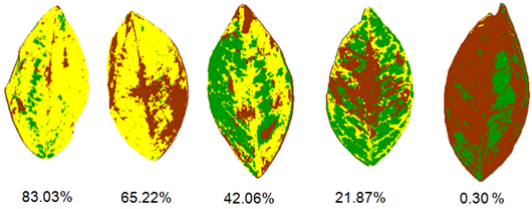


Figura 17. Escala diagramática da Rubelose.

A Figura 16 apresenta a escala diagramática para o *greening*, com os seguintes níveis de severidade: 83,34%, 39,34%, 28,93%, 12,86% e 0,00%. Observa-se nas folhas a clorose assimétrica.

A Figura 17 apresenta a escala diagramática para a Rubelose, com os seguintes níveis de severidade: 83,03%, 65,22%, 42,06%, 21,87% e 0,30%. Observa-se nas folhas manchas amarelas com lesões.

5. Conclusões

A proposta deste trabalho é aplicar técnicas de processamento de imagens em folhas de citros digitalizadas, para diferenciar o *greening* de outras pragas. A segmentação por cor usando RNA PMC, mostrou ser adequada, com acurácia de 96,04%. Em seguida, a imagem foi dividida em quadrantes para possibilitar uma análise inicial para diferenciar a variação das manchas em cada quadrante, para cada doença. A aplicação da RNA apresenta um melhor resultado na classificação das classes quando utilizada para as classes que mais se assemelham ao *greening*, isto é, 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, indicada pelos descritores de forma não consegue diferenciar totalmente o *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 avaliar os resultados obtidos na análise por quadrantes juntamente com os resultados obtidos pela extração de atributos, possibilitando diferenciar o *greening* de outras pragas, e posteriormente construir a escala diagramática.

Agradecimentos

Os autores agradecem o apoio financeiro do CNPq (processo 578627/2008-6) e o fornecimento das amostras de foliares pela empresa Fischer.

Referências

Agrianual, . *Anuário da Agricultura Brasileira*. São Paulo, SP, 2008. p. 520.

Amorim, L., Avaliação de doenças. In: Bergamin Filho, A.; Kimati, H. & Amorim, L. (Eds.), *Manual de Fitopatologia – princípios e conceitos*. São Paulo, SP: Agronômica Ceres, v. 1, p. 647–671, 1995.

Basset, O.; Buquet, B.; Abouelkaram, S.; Delachartre, P. & Culioli, J., Aplication of texture image analysis for the classification of bovine meat. *Food Chemistry*, 69(4):437–445, 2000.

Belasque, J.R.J.; Filho, A.B.; Bassanezi, R.B.; Barbosa, J.C.; Gimenes-Fernandes, N.; Yamamoto, P.; Lopes, A.S.; Machado, M.A.; Leite, J.R.P.; Ayres, A.J. & Massari, C.A.. *Greening*: A instrução normativa n. 53 e a necessidade de um controle efetivo no Brasil, 2009. p. 1–5. Http://www.fundecitrus.com.br/ImageBank/FCKEditor/file/pdf/artigo_controle_greening.pdf.

Bové, J.M., Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. *Journal of Plant Pathology*, 88(1):7–37, 2006.

Cavani, F.A.; de Sousa, R.V.; Porto, A.J.V. & Tronco, M.L., Segmentação e classificação de imagens de laranjeiras utilizando JSEG e perceptron multicamadas. *Revista Minerva*, 3(2):189–197, 2006.

FUNDECITRUS, . Manual técnico *Greening* 2009. Araraquara, SP, 2009. p. 1–12. Http://www.cda.sp.gov.br/greening/lnk_greening_ctr/ downloads/greening.pdf.

Innis, M.A.; Gelfand, D.H.; Sninsky, J.J. & White, T.J., *PCR Protocols: A guide to methods and applications*. San Diego, USA: Academic Pres, 1990.

Jain, A.K., *Fundamentals of Digital Image Processing*. Englewood Cliffs, USA: Prentice Hall, 1989.

Kowata, L.S.; May-De-Mio, L.L.; Dalla-Pria, M. & Santos, H.A.A., Escala diagrámatica para avaliar severidade de mildio na soja. *Scientia Agraria*, 9(1):105–110, 2008.

Martins, M.C.; Guerzoni, R.A.; Câmara, G.M.S.; Mattiazzi, P.; Lourenço, S.A. & Amorim, L., Escala diagramática para a quantificação do complexo de doenças foliares de final de ciclo em soja. *Fitopatologia Brasileira*, 29(2):179–184, 2004.

de Moraes, S.A., Quantificação de doenças de plantas. 2007. Http://www.infobibos.com/Artigos/2007_1/doencas/index.htm.

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*, 17(2):174–184, 1995.

Russ, J.C., *The Image Processing Handbook*. 3a edição. Boca Raton, USA: CRC Press, 1998.

Silva, I.N.; Spatti, D.H. & Flauzino, R.A., *Redes Neurais Artificiais Para Engenharia e Ciências Aplicadas: Um Curso Prático*. Porto Alegre, RS: Artliber, 2010.

Simões, A.S., *Segmentação de imagens por classificação de cores: uma abordagem neural*. Dissertação de mestrado em engenharia, Escola Politécnica da Universidade de São Paulo, São Paulo, SP, 2000.

Simões, A.S. & Reali Costa, A.H., Using neural color classification in robotic soccer domain. In: Barros, L.N.; Cesar Jr., R.M.; Cozman, F.G. & Reali Costa, A.H. (Eds.), *Proceedings of International Joint Conference IBERAMIA-SBIA – Workshop Meeting on Multi-Agent Collaborative and Adversarial Perception, Planning, Execution, and Learning*. Atibaia, SP: SBC, p. 208–213, 2000.

Simões, A.S.; Reali Costa, A.H. & Andrade, M.T.C., Utilizando um classificador *fuzzy* para seleção visual de laranjas. In: Ribeiro, C.H.C. & Sakude, M.T.S. (Eds.), *Anais do Workshop de Computação*. São José dos Campos, SP: ITA, p. 113–117, 2001.

Sposito, M.B., *Dinâmica temporal e espacial da mancha preta (*Guignardia citricarpa*) e quantificação dos danos causados à cultura dos citros*. Tese de doutorado em fisiopatologia, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, 2004.

Witten, I.H. & Frank, E., *Data mining: practical machine learning tools and techniques*. 2a edição. San Francisco, USA: Morgan Kaufmann, 2005.

Yonekawa, S.; Sakai, N. & Kitani, O., Identification of idealized leaf type using simple dimensionless shape factors by image analysis. *Transactions of the ASAE*, 39(4):1525–1533, 1996.

Notas Biográficas

Patricia Pedroso Estevam Ribeiro é graduado em Engenharia Elétrica com ênfase em computação (Faculdade de Engenharia de Barretos, 2006). Atualmente é mestrando em Engenharia Elétrica na USP São Carlos na área de processamento digital de imagens.

Maria Stela Veludo de Paiva possui graduação em Engenharia Elétrica/Eletrônica (Universidade de São Paulo – USP, 1979), mestrado e doutorado em Física Aplicada (USP/São Carlos, 1984 e 1990), tendo realizado Pós-Doutorado na University of Southampton (1992). Atualmente é docente no Departamento de Engenharia Elétrica da Escola de Engenharia de São Carlos (USP) e desenvolve pesquisas na área de Visão Computacional.

Lúcio André de Castro Jorge é graduado em Engenharia Elétrica (Faculdade de Engenharia de Barretos, 1987), mestre em Ciência da Computação (USP São Carlos, 2001) e doutor em Engenharia Elétrica na área de processamento digital de imagens (USP São Carlos, 2011). Atualmente é pesquisador da EMBRAPA Instrumentação em São Carlos.

Capítulo 9

Delimitação da Área de Impressões Digitais Utilizando Contornos Ativos

Marcos William da Silva Oliveira, Inês Aparecida Gasparotto Boaventura e Maurílio Boaventura*

Resumo: A segmentação de impressões digitais, etapa fundamental no pré-processamento de sistemas automatizados de identificação (AFIS, em inglês), é o foco do estudo apresentado neste capítulo. É descrito um novo método automático para delimitação da área de impressões digitais, obtidas em ambiente controlado, baseado na aplicação de contornos ativos sem detectores de bordas e de operações morfológicas. No caso de imagens de impressões digitais latentes, sugere-se uma extensão semiautomática do procedimento, em que a entrada de um contorno inicial simples é requerida. Resultados experimentais demonstram a exatidão da proposta e destacam sua contribuição devido à aplicabilidade em imagens de baixa qualidade.

Palavras-chave: Processamento de imagens, Impressões digitais, Segmentação, Contornos ativos.

Abstract: *Fingerprint segmentation, a fundamental pre-processing step of automatic fingerprint identification systems (AFIS), is the focus of this chapter. Based on active contours without edges and morphological operations, a new automatic method for area delimitation in fingerprint images is presented. In latent fingerprint images, a semi-automatic extension of this procedure is suggested, which requires a manual input of a simple initial contour. Experimental results show the accuracy of the model and highlight its contribution due to the applicability in low quality images.*

Keywords: *Image processing, Fingerprints, Segmentation, Active contours.*

*Autor para contato: maurilio@ibilce.unesp.br

1. Introdução

O interesse da humanidade pelas impressões digitais data da pré-história, como indicam desenhos arqueológicos que fazem alusão a seus padrões e placas de cerâmica antigas com impressões gravadas no sentido de selar acordos ou assinar autorias (Maltoni et al., 2009; Laufer, 1913). 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 (Maltoni et al., 2009).

O Brasil foi um dos primeiros países a adotar oficialmente tal processo de identificação datiloscópica através do Decreto 4.764 de 5 de Fevereiro de 1903 e da criação do então "Gabinete de Identificação e Estatística da Polícia do Distrito Federal", hoje "Instituto de Identificação Felix Pacheco" do Estado do Rio de Janeiro. Nesta mesma época, noticiava-se, em Nova York, a solução de um crime na Índia a partir das digitais do suspeito (Laufer, 1913) e, a partir daí, as linhas da ponta dos dedos começaram a ser utilizadas para identificação criminal em diversas localidades.

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) (Maltoni et al., 2009). De acordo com Yoon et al. (2010), com o surgimento do sistema AFIS aumentou significativamente a velocidade da identificação de impressões digitais e viabilizou a identificação de impressões digitais parciais, obtidas em locais de crimes, contra grandes bancos de dados. Atualmente, grandes avanços têm sido obtidos no sentido de aumentar o rendimento e a precisão destes sistemas.

Com esta motivação, este trabalho tem como objetivo apresentar uma fundamentação teórica acerca das imagens de impressão digital e descrever um processo automático de segmentação de impressões digitais roladas ou planas, baseado no modelo de contornos ativos (ou *snakes*). Os conceitos básicos aqui expostos foram baseados em Maltoni et al. (2009). O método adotado para segmentação é uma aplicação do modelo sem detectores de bordas de Chan & Vese (2001) e uma filtragem pela operação morfológica abertura. Sugere-se também uma extensão semiautomática do processo para impressões digitais latentes, originárias de ambiente não controlado.

Outros trabalhos publicados recentemente para detecção da área de impressões digitais utilizando *snakes* podem ser encontrados em Zheng et al. (2007) e Weixin et al. (2009). Zheng et al. (2007) aplicaram um método de contornos ativos sobre uma representação da área da impressão digital obtida através do descritor Fisher (Fukunaga, 1972), modificado pelo uso

da média dos níveis de cinza e da coerência do campo de orientação da impressão digital. De forma distinta, Weixin et al. (2009) apresentam um contorno ativo específico para impressões digitais, considerando a variância dos níveis de cinza e o histograma da imagem direcional para controlar o movimento da *snake*, diminuindo a sensibilidade do modelo à escolha do contorno inicial. No entanto, há limitações no uso de ambos os métodos, uma vez que o campo de orientação deve ser obtido antes da segmentação. Este fato inviabiliza, por exemplo, suas aplicações a imagens com pouca qualidade como as impressões digitais latentes. Neste trabalho, sugere-se uma extensão semiautomática do modelo proposto para o caso específico das imagens de impressões digitais latentes.

O capítulo é organizado da seguinte forma: na Seção 2, apresentam-se conceitos básicos do processamento de imagens de impressão digital; a Seção 3 trata das operações fundamentais da morfologia matemática binária; na Seção 4, destaca-se o modelo *snakes* de Chan & Vese (2001); na Seção 5 descreve-se a proposta de segmentação; por fim, na Seção 6, são mostrados alguns resultados obtidos e as avaliações quantitativas realizadas para a validação do processo.

2. Biometria de Impressões Digitais

"Como resultado da interação de fatores genéticos e condições embrionárias, o padrão de atrito das cristas papilares na ponta do dedo sobre uma superfície é único para cada dedo" (Feng & Jain, 2011). Este padrão não é aleatório devido à influência genética na sua formação. Desta maneira, sistemas automatizados de reconhecimento de impressões digitais desempenham papel importante em muitas situações em que se necessita identificar com confiabilidade, uma pessoa.

Conforme Maltoni et al. (2009), a mais evidente característica estrutural de uma digital é um padrão intercalado de cristas e vales dos sulcos papilares. Chamam-se "cristas" os picos dos sulcos, que formam as linhas escuras na impressão, e "vales" as 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 variadas qualidades. Assim, estas podem ser classificadas em:

Rolada: a digital é rolada sobre um papel, depois de embebida em tinta, ou sobre a superfície de captação do sensor óptico;

Plana: a digital é apenas pressionada sobre um papel, depois de embebida em tinta, ou sobre a superfície de captação do sensor óptico;

Latente: fragmento de impressão digital deixado por descuido do indivíduo em um local de crime. Este fragmento é revelado através de processos químicos e registrado, por exemplo, por fotografia.

Figura 1. Impressões digitais rolada, plana e latente do mesmo dedo (extraída de Yoon et al. (2010)).

Como é intuitivo, uma impressão digital rolada contém maior área do dedo que uma impressão digital plana, mas está sujeita a algumas distorções por causa do movimento realizado pelo indivíduo. Além disto, uma imagem de impressão digital plana ou rolada geralmente tem melhor qualidade que impressões digitais latentes, como pode-se ver na Figura 1.

De acordo com suas características, uma imagem de impressão digital pode ser analisada em três níveis, segundo Maltoni et al. (2009):

Nível 1 (Nível Global): referente às disposições ou formas das cristas. Pode-se observar as chamadas "singularidades" dos tipos presilha, delta e verticilo, parecidas respectivamente com os símbolos ⋒, Δ e ◎;

Nível 2 (Nível Local): referente à observação de características na vizinhança de um *pixel*, como as minúcias e a orientação local de crista;

Nível 3 (Nível Muito Local): caracterizado por detalhes adicionais finos que possam ser extraídos da imagem de impressão digital.

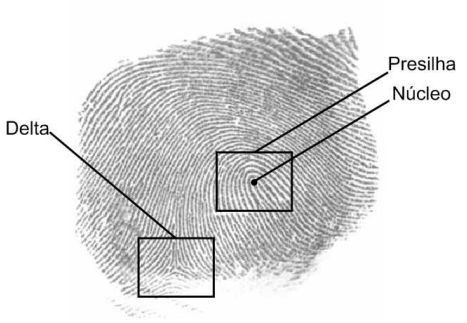
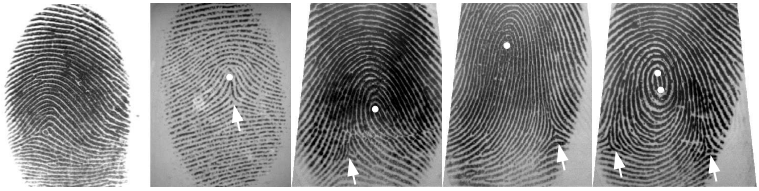


Figura 2. Posições do núcleo e delta em uma impressão digital.

No nível global, algoritmos que correlacionam impressões digitais podem pré-alinhar suas imagens de acordo com pontos de referência das singularidades. Um deles é o ponto central de uma presilha, quando esta ocorre, chamado "núcleo" (*core*). Henry (1900) define o núcleo como "o ponto mais ao norte da linha de crista mais interna" da estrutura. Outra singularidade que ocorre globalmente é chamada *delta*, por seu formato triangular (Figura 2). Quando estas singularidades não ocorrem, o referencial para o alinhamento deve ser obtido localmente.

A disposição das cristas e a ocorrência de singularidades permitem a classificação das impressões digitais em grandes classes, como apresenta a Figura 3, com imagens retiradas do banco de dados FVC2004 (Maio et al., 2004). Nesta figura, os núcleos e deltas são indicados por pontos e setas brancas, respectivamente.



(a) (b) (c) (d) (e)

Figura 3. Classes das impressões digitais no nível global: (a) Arco, (b) Arco Angular, (c) Presilha Direita, (d) Presilha Esquerda, (e) Verticilo.

Na classificação proposta por Henry (1900), digitais do tipo arco (*arch*) não possuem singularidades (Figura 3(a)); as do tipo arco angular (*tented arch*) (Figura 3(b)), e presilha (*loop*), à direita (Figura 3(c)) ou à esquerda (Figura 3(d)) possuem um núcleo e um delta; enquanto as do tipo verticilo (*whorl*) (Figura 3(e)) possuem duas de cada singularidade.

No nível local, estuda-se uma importante característica: a minúcia. Uma minúcia é uma descontinuidade, bifurcação ou variação abrupta no caminho esperado de uma linha de crista. Em sistemas automáticos o *Federal Bureau of Investigation* (FBI) considera apenas as de tipo terminação e bifurcação, devido à dificuldade de determinar os vários tipos de minúcias que podem ser lago, ponto, *spur*, dentre outros (Maltoni et al., 2009). Como os nomes sugerem, a terminação ocorre no ponto em que uma crista se encerra e a bifurcação ocorre onde uma crista se duplica, seguindo dois caminhos distintos (Figura 4).

No nível muito local, detalhes finos adicionais podem ser extraídos no padrão da impressão digital, mas para isto são necessárias imagens com resolução alta, a fim de destacar poros e largura de cristas. Porém, na prática nem sempre é possível encontrar imagens com alta resolução, o que afeta

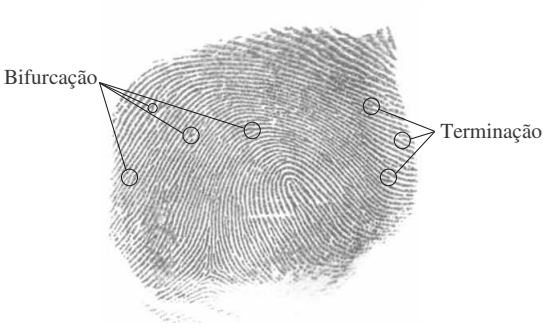


Figura 4. Destaque de algumas minúcias em uma impressão digital.

principalmente a confiabilidade de modelos de extração de características (Thai, 2003). Em geral, antes da aplicação destes métodos são necessários a segmentação e, em alguns casos, o melhoramento da imagem de entrada.

2.1 Segmentação de impressão digital

Maltoni et al. (2009) destacam que o "termo segmentação denota, geralmente, a separação da área da impressão digital (*foreground*) do fundo da imagem (*background*)", ou seja, destacar a região exata da imagem onde ocorre a impressão digital. Esta separação é útil para evitar a extração de características em áreas ruidosas que aparecem principalmente no fundo da imagem, tais como riscos manuscritos e caracteres impressos. Em Maltoni et al. (2009), encontram-se breves descrições de alguns destes métodos.

O resultado destes métodos é a máscara da área da impressão digital que pode ser usada em processos de melhoramento, de 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 se deseja processar. No caso deste trabalho, a abordagem baseada em contornos ativos propicia ainda a independência de pré-processamento para se obter a segmentação da impressão digital e, portanto, maior aplicabilidade e desempenho na extração de características, visto que, em geral, os demais métodos de segmentação baseiam-se em características pré-extraídas como campo de orientação.

Nas Seções 3 e 4 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 na Seção 5.

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. A 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. Estas técnicas extraem fronteiras, esqueletos, fecho convexo e também podem realizar pré ou pós-processamentos de filtragem, afinamento e poda (*pruning*) (Gonzalez & Woods, 2002). A seguir, descreve-se brevemente algumas operações de morfologia matemática para imagens binárias utilizadas no método de segmentação descrito na Seção 5.

As duas operações básicas da morfologia matemática binária são a dilatação e a erosão. 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. Na operação de erosão, o valor do *pixel* resultante é 1 se todos os *pixels* em sua vizinhança possuem 1 e nulo caso contrário.

Associando as duas operações, tem-se a operação chamada abertura, que é a dilatação da imagem erodida, considerando a mesma vizinhança. Esta vizinhança é definida pelo chamado elemento estruturante, que possui formas e tamanhos diversos.

Considere o seguinte exemplo para a imagem binária '`A`' e o elemento estruturante '`se`' apresentados na Figura 5. Ainda nesta figura, observa-se a dilatação '`dA`' e a erosão '`eA`' de '`A`' por '`se`'. Aplicar a operação abertura sobre '`A`' por '`se`' é, portanto, realizar a erosão de '`A`' seguida da dilatação do resultado, ambas por '`se`'. Este resultado é mostrado na Figura 6.

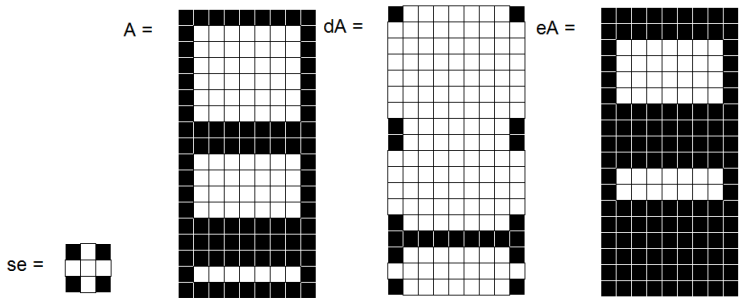


Figura 5. (a) Elemento estruturante '`se`', (b) imagem binária '`A`', (c) resultados da dilatação e (d) erosão.

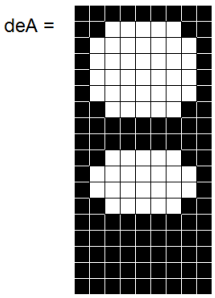


Figura 6. Resultado '`deA`' da dilatação de '`eA`' por '`se`'.

Como os nomes (e o exemplo) 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 são 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. Esta homogeneidade pode ser quanto às diferenças de tonalidade de cinza ou de texturas em cada região. Pode-se afirmar que estas técnicas são capazes de encontrar automaticamente objetos na imagem.

Neste sentido, o modelo de contornos ativos (ou *snakes*), primeiramente proposto por Kass et al. (1987), 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 a forças (ou energias) simulando propriedades físicas como comprimento, flexibilidade e rigidez. Em Kass et al. (1987), estas energias são divididas em energia interna, derivada da geometria da curva e das propriedades da imagem impondo restrições de suavidade; e energia externa, responsável por colocar o contorno sobre a borda do objeto.

Considere, para tanto, um conjunto aberto e limitado $\Omega \subset \mathbb{R}^2$, sendo $\partial\Omega$ sua fronteira e $\overline{\Omega}$ seu fecho. Ainda, sejam $u_0 : \overline{\Omega} \to \mathbb{R}$ uma dada imagem e $C(s) : [0, 1] \to \overline{\Omega}$ uma curva parametrizada. Desta forma, o funcional de energia da *snake* é escrito em Kass et al. (1987) como

$$E_{snake} = \int_0^1 E_{snake}(C(s))ds = \int_0^1 E_{int}(C(s)) + E_{im}(C(s)) + E_{ext}(C(s))ds, \tag{1}$$

onde E_{int} representa a energia interna, E_{im} determina as forças da imagem e E_{ext} é a energia externa. No desenvolvimento da Equação (1), obtém-se o modelo de contornos ativos baseado na minimização do funcional de energia $J(C)$ (Chan & Vese, 2001):

$$\inf_C J(C) = \inf_C \left[\alpha \int_0^1 |C'(s)|^2 ds + \beta \int_0^1 |C''(s)| ds - \lambda \int_0^1 |\nabla u_0(C(s))|^2 ds \right], \tag{2}$$

com α, β e λ sendo parâmetros positivos. Veja que buscar a curva C onde $J(C)$ seja ínfimo, equivale a buscar a curva C onde $|\nabla u_0|$ seja máximo. Desta forma, ∇u_0 atua em (2) como detector de bordas.

Outras funções de detecção de bordas poderiam ser utilizadas no funcional que define a energia externa, porém a necessidade de segmentar objetos sem borda definida pelo gradiente levou ao estudo de *snakes* baseadas em regiões, sem detectores de borda, como em Chan & Vese (2001).

Assim, 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

$$\inf_{c_1,c_2,C} F(c_1, c_2, C), \tag{3}$$

sendo o funcional F definido por

$$\begin{aligned} F(c_1, c_2, C) \quad = \quad & \mu.\text{Comprimento}(C) + \nu.\text{Área}(\omega) \\ & + \lambda_1 \int_\omega |u_0(x,y) - c_1|^2 dxdy \\ & + \lambda_2 \int_{\Omega \setminus \overline{\omega}} |u_0(x,y) - c_2|^2 dxdy, \end{aligned} \tag{4}$$

onde $\mu \geqslant 0$, $\nu \geqslant 0$, $\lambda_1, \lambda_2 > 0$ são parâmetros fixados.

Na Equação (4), os dois últimos termos atuam "empurrando" a *snake* até o equilíbrio das energias internas e externas "colocando" C sobre a borda do objeto e os dois primeiros são termos de regularização e suavidade.

A maneira apresentada para resolver a minimização da Equação (3) é através do funcional Mumford-Shah (Mumford & Shah, 1989), reduzido ao chamado problema de partição mínima, assim descrito:

$$\begin{aligned} F^{MS}(u, C) \quad = \quad & \mu.\text{Comprimento}(C) \\ & + \lambda \int_\omega |u_0(x,y) - u(x,y)|^2 dxdy \\ & + \int_{\Omega \setminus \overline{\omega}} |\nabla u(x,y)|^2 dxdy, \end{aligned} \tag{5}$$

onde $u_0 : \overline{\Omega} \to \mathbb{R}$ é a imagem dada, $\mu, \lambda > 0$ são parâmetros e $u(x, y)$ é constante em cada região:

$$u(x,y) = constante\ c_i = \begin{cases} c_1 = & \text{média}(u_0(x,y)), & (x,y) \in \omega \\ c_2 = & \text{média}(u_0(x,y)), & (x,y) \in \Omega \setminus \overline{\omega}. \end{cases}$$

Observa-se que, tomando na Equação (4) $\nu = 0$, $\lambda_1 = \lambda_2 = \lambda$ e c_1 e c_2 como acima, que o funcional F é um caso particular de F^{MS} em (5). Agora, para solucionar este caso particular do problema de partição mínima, pode-se utilizar o método *level set* (Osher & Sethian, 1988) em que $C \subset \Omega$ é representado pela função Lipschitziana $\phi : \Omega \times \mathbb{R}_+ \to \mathbb{R}$, tal que

$$\begin{cases} C = \partial\omega = \{(x,y) \in \Omega | \phi(x,y) = 0\}, \\ \omega = \{(x,y) \in \Omega | \phi(x,y) > 0\} \text{ e} \\ \Omega \setminus \overline{\omega} = \{(x,y) \in \Omega | \phi(x,y) < 0\}. \end{cases}$$

Pelo método *level set*, a função ϕ adiciona uma dimensão temporal à curva C e, daí, C pode ser recuperada como curva de nível de ϕ fazendo $t = 0$ ($\phi(x, y, 0) = C(x, y)$).

A partir disto e 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 (4) com respeito a ϕ, obtém-se que, para $\epsilon \to 0$, o problema de minimização (3) é solucionado resolvendo a equação diferencial de evolução:

$$\begin{cases} \dfrac{\partial \phi}{\partial t} = \delta_\epsilon(\phi)[\mu \text{div}\left(\dfrac{\nabla \phi}{|\nabla \phi|}\right) - \nu - \lambda_1 (u_0 - c_1(\phi))^2 \\ \qquad\qquad + \lambda_2 (u_0 - c_2(\phi))^2] = 0, (t,x,y) \in (0,\infty) \times \Omega, \\ \phi(0,x,y) = \phi_0(x,y), \quad (x,y) \in \Omega, \\ \dfrac{\delta_\epsilon(\phi)}{|\nabla \phi|} \dfrac{\partial \phi}{\partial \overrightarrow{n}} = 0, \quad (x,y) \in \partial\Omega, \end{cases} \tag{6}$$

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

$$\begin{cases} c_1(\phi) = & \text{média}(u_0(x,y)), & \phi(x,y) \geqslant 0, \\ c_2(\phi) = & \text{média}(u_0(x,y)), & \phi(x,y) < 0. \end{cases}$$

Para maiores detalhes desta última passagem da Equação (4) para a (6), bem como a aproximação numérica para solução da Equação (6) e algoritmo do modelo, indica-se a leitura de Chan & Vese (2001).

5. Modelo de Detecção da Área

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 da imagem.

Na abordagem de Oliveira (2011), considera-se a impressão digital como um objeto na imagem e, dado um contorno inicial, aplica-se o método de contornos ativos Chan-Vese (6). A partir de então, realiza-se uma filtragem para se 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. Para imagens de boa qualidade, esta é determinada automaticamente, tomando-se, para isto, um ponto qualquer no interior da impressão digital, geralmente o centro da imagem, e comparando-se a variância dos níveis de cinza 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 impressões digitais 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*. Isto é necessário para corrigir imperfeições nos casos em que os contornos são atraídos por ruídos específicos, como escritas ou desenhos manuais e caracteres impressos, presentes no fundo da imagem, como pode-se observar nos resultados apresentados na Seção 6.

5.1 Algoritmo

Primeiramente, a definição da máscara inicial automática é utilizada apenas para imagens de boa qualidade, como impressões digitais roladas e planas. Nesta implementação, é fornecida a imagem original `I` como entrada e a matriz `m`, que define a máscara, é inicializada como uma matriz nula com as dimensões de `I`. Toma-se um ponto `C`$= (c_1, c_2)$ pertencente à área da impressão digital como referencial da posição da impressão digital na imagem, em geral o ponto central da imagem, e no procedimento seguinte determina-se uma região retangular que contém a impressão digital. Para isto, faz-se uma busca dos blocos com variância dos níveis de cinza menor ou igual a um limite (`lim`) nas duas direções cartesianas e nos quatro sentidos a partir de `C`. Devido ao fato de que a variância em blocos com *pixels* no interior da impressão digital é maior do que em blocos com *pixels* no 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 a Figura 7. 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`.

A aplicação deste processo não é satisfatória para impressões digitais latentes, dado que estas imagens possuem um fundo complexo com ruído estruturado ou mesmo pouca variação dos níveis de cinza na região da impressão digital. Assim, a entrada manual de uma máscara inicial se faz necessária, tornando-se um processo semiautomático.

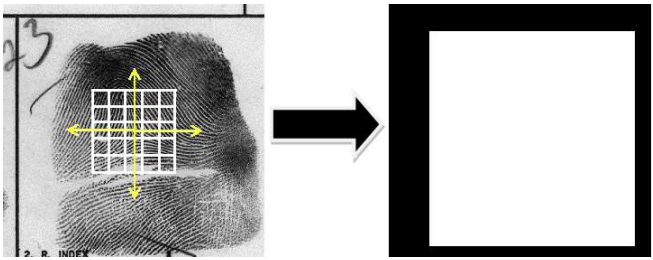


Figura 7. Esquema da obtenção do contorno inicial para uma dada imagem (NIST (2010)). À esquerda, as linhas brancas representam os 25 blocos centrais, os vetores indicam os sentidos da busca e `C` é a origem dos vetores. À direita, a máscara binária obtida.

Em suma, a proposta para segmentação da impressão digital, esquematizada a seguir como pseudo-código, é fornecer a imagem `I`, o número de iterações $\mathtt{n}_{it}$ e o parâmetro μ para o modelo *snakes*. A seguir, utiliza-se a máscara inicial `m` automática (para impressões digitais roladas e planas) ou manual (para impressões digitais latentes) como contorno inicial para aplicação do modelo Chan-Vese[1] (Chan & Vese, 2001), obtendo a matriz `area`, e realizar sua correção através da operação abertura com o elemento estruturante `se`.

No caso de impressões digitais latentes, utiliza-se uma imagem auxiliar, $\mathtt{I}_{aux}$ =`I(find(m))`, para ignorar as posições de `I` correspondentes aos elementos nulos em `m`, a fim de evitar incorreções no passo seguinte. A função `find` é uma função lógica padrão do MATLAB® que retorna as posições da matriz que são verdadeiras (sendo, em uma imagem binária, os 0 posições "falsas" e os 1 posições "verdadeiras"). No caso, $\mathtt{I}_{aux}$ contém apenas a imagem original `I` nas posições em que `m` possui valor 1.

A matriz resultante da aplicação do pseudo-código, `area`, bem como a matriz que define a máscara inicial, `m`, são matrizes binárias que possuem 1 nas posições que envolvem a impressão digital na imagem e 0 nas posições

[1] Disponível em http://www.mathworks.com/matlabcentral/fileexchange/23445

de fundo. Ou seja, a imagem produzida por `m` possui uma região retangular branca, que define o contorno inicial para o modelo *snakes*, e um fundo preto. Quanto à imagem produzida por `area`, também há uma região branca em um fundo preto, mas com a forma da área da impressão digital da imagem de entrada `I`, como se vê nas figuras apresentadas na Seção 6.

Pseudo-Código (Detecção da Área)

Início

1. **Entradas:**
   `I`: impressão digital em escala de cinza.
   `m`: máscara inicial não automática.
   $\mathtt{n}_{it}$: número de iterações do método Chan-Vese.
   μ: constante do método Chan-Vese.
   `dimblc`: tamanho do bloco.
2. **Impressões Digitais Roladas/Planas:**
   se (`I` é imagem rolada/plana)
   `m` = `mascara_inicial`(`I`, `dimblc`);
   fim se
3. **Impressões Digitais Latentes:**
   se (`I` é imagem latente)
   $\mathtt{I}_{aux}$ = `I`(`find`(`m`)); `I` = $\mathtt{I}_{aux}$;
   fim se
4. **Contorno Ativo (*Snakes*):**
   `area` = `chan-vese`(`I`, `m`, $\mathtt{n}_{it}$, μ);
5. **Abertura Binária:**
   `se` = elemento estruturante disco de raio 16;
   `area` = `abertura`(`area`, `se`);
6. **Saída:**
   `area`: máscara binária que define a área.

Fim

6. Resultados Experimentais

Nesta seção são apresentados alguns resultados obtidos com a aplicação da detecção da área descrita na Seção 5 em impressões digitais roladas e latentes do banco de dados NIST (2010) e impressões digitais planas do banco de dados FVC2000 (Maltoni et al., 2009). Alguns parâmetros foram definidos experimentalmente, como é o caso do valor μ, para o modelo Chan-Vese (6), em torno de 2×10^4, e a quantidade de iterações limitada em cada caso, com valores próximos a 200.

6.1 Segmentações obtidas

Na Figura 8, destaca-se o resultado obtido sobre uma impressão digital rolada, sendo que sua primeira linha traz, à esquerda, a imagem original, e à direita, a imagem segmentada. Ainda nesta figura, é apresentada, na segunda linha, da esquerda à direita, a máscara automática e os resultados da segmentação por *snakes* e da abertura binária. Na Figura 9, ocorre a mesma apresentação para uma impressão digital latente.

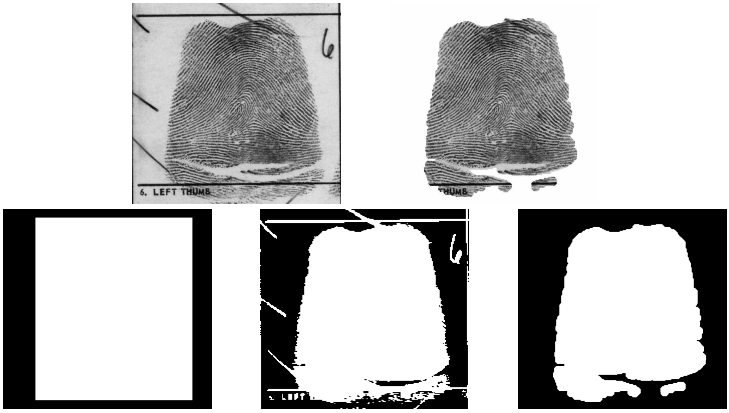


Figura 8. Primeira linha: imagem original e segmentada. Segunda linha: máscara automática, segmentação por *snakes* e resultado binário da abertura (NIST27).

As Figuras 10, 11 e 12 mostram outros resultados, com imagens de impressões digitais planas, roladas e latentes, respectivamente. Na primeira coluna, aparecem as imagens originais; na segunda, os resultados binários do procedimento e, na terceira, as segmentações finais. No caso de impressões digitais roladas e planas, o parâmetro $\mathtt{n}_{it}$ foi fixado em 200, enquanto para as imagens de impressão digital latentes, este parâmetro é variável (ver Figura 12). Em todos os casos, tem-se o parâmetro μ=20000.

6.2 Exatidão da segmentação

O objetivo desta avaliação é analisar a exatidão da segmentação pela comparação dos resultados da segmentação com máscaras manualmente marcadas (*ground truth*). A exatidão da segmentação foi avaliada baseada em quatro medidas: Taxa de Verdadeiros Positivos (TVP), Taxa de Falsos Positivos (TFP), Taxa de Verdadeiros Negativos (TVN) e Taxa de Falsos Negativos(TFN). TVP refere-se à porcentagem de áreas genuínas detectadas corretamente como genuínas; TFP é a porcentagem de áreas impostoras detectadas erroneamente como genuínas; TFN representa a porcenta-

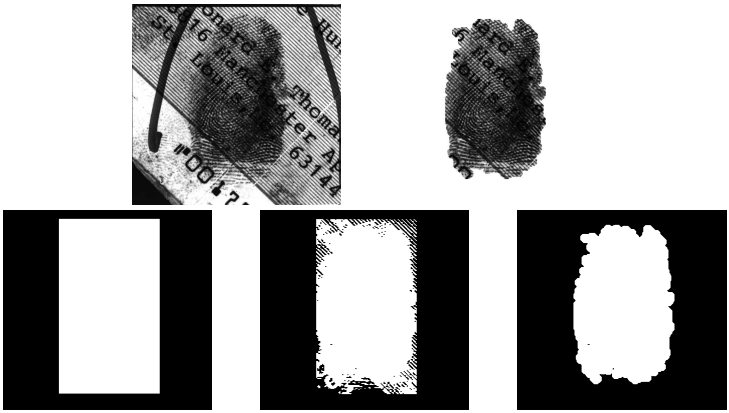


Figura 9. Primeira linha: imagem original e segmentada. Segunda linha: máscara automática, segmentação por *snakes* e resultado binário da abertura (NIST27) ($\mathbf{n}_{it} = 150$).

gem de áreas genuínas detectadas incorretamente como impostoras; e TVN refere-se à porcentagem de áreas impostoras detectadas corretamente como impostoras. A partir dos resultados anteriores, tem-se a TME, que diz respeito à taxa média de erros, a média entre os falsos positivos e os falsos negativos, e a TMA refere-se à taxa média de acertos, ou seja a média entre os verdadeiros positivos e os verdadeiros negativos. Os valores da avaliação são mostrados na Tabela 1 e, para as imagens de impressões digitais roladas, refletem o comportamento da aplicação do método aqui proposto a todas as imagens da base de dados NIST27, enquanto que para as imagens de impressões digitais planas e latentes, refletem o comportamento apenas das imagens mostradas neste trabalho.

Tabela 1. Comparação com a segmentação manual.

Impressões Digitais	**TVP**	**TVN**	**TFP**	**TFN**	**TME**	**TMA**
Planas	90,00%	98,35%	1,82%	10,00%	5,91%	94,17%
Roladas	89,02%	95,92%	4,32 %	10,97%	7,65%	92,47%
Latentes	92,03%	94,98%	14,63%	7,97%	11,30%	93,51%

Em segmentações adequadas, são esperados valores altos para TVP e TVN e baixos para TFP e TFN, comportamento que pode ser devidamente observado na Tabela 1.

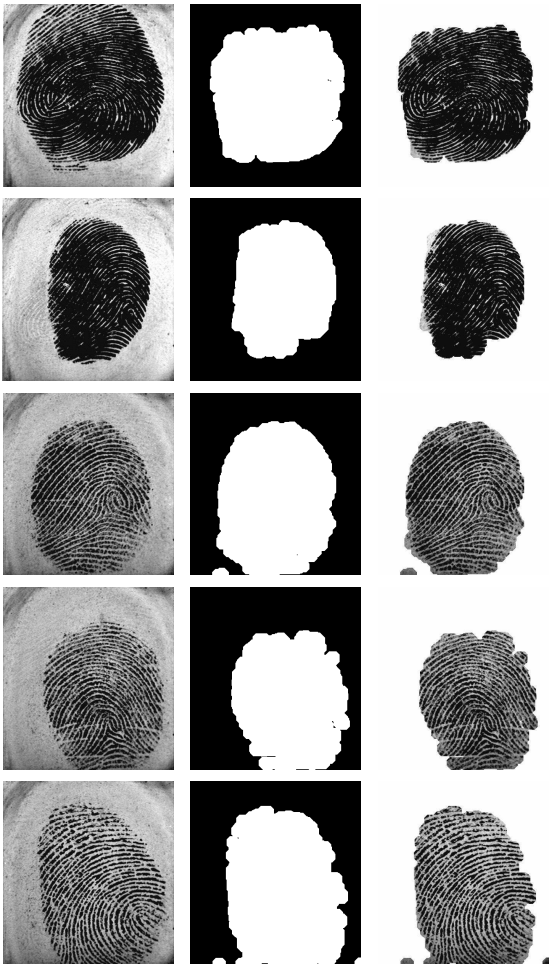


Figura 10. Segmentação de impressões digitais planas (FVC2000). Em cada linha, da esquerda para a direita, tem-se as imagens originais, as áreas obtidas e as segmentações finais.

7. Conclusões

O objetivo deste capítulo foi descrever 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 semiautomática para imagens de impressões digitais latentes, através da pré-definição manual de uma máscara inicial retangular simples.

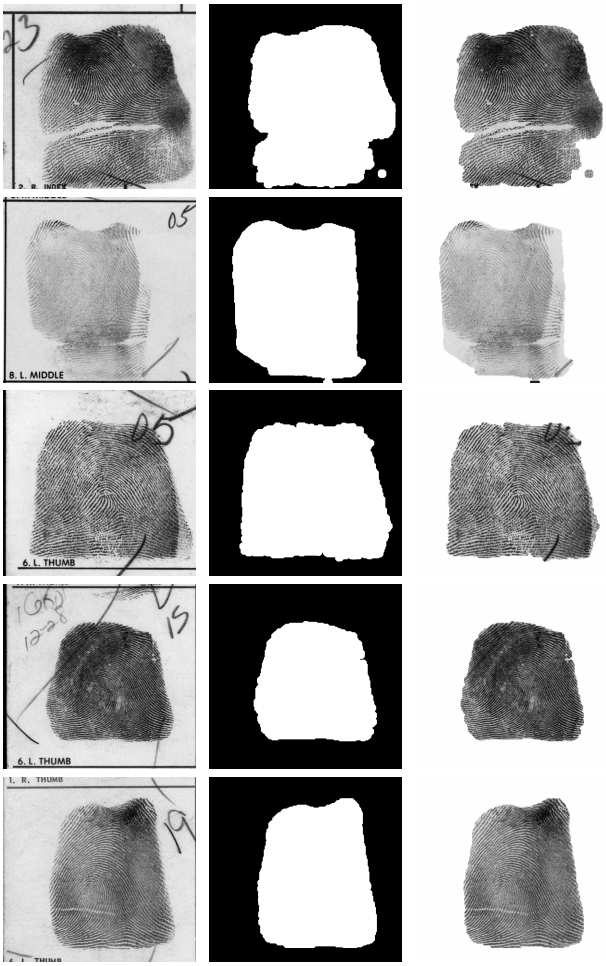


Figura 11. Segmentação de impressões digitais roladas (NIST27). Em cada linha, da esquerda para a direita, tem-se as imagens originais, as áreas obtidas e as segmentações finais.

Pelos resultados obtidos e avaliações realizadas, a segmentação de impressões digitais planas e roladas mostrou-se bastante satisfatória. 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 identificar a área da impressão digital preci-

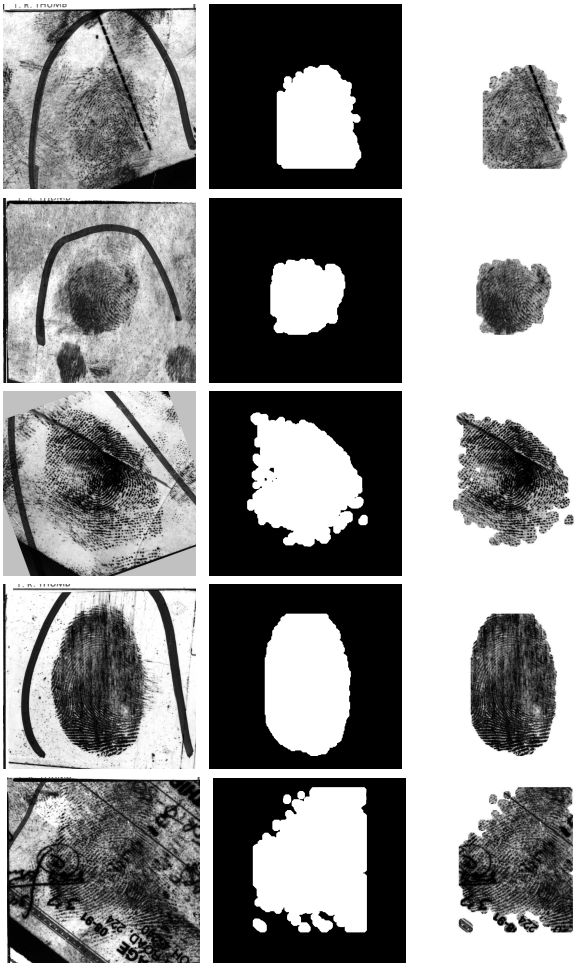


Figura 12. Segmentação de impressões digitais latentes (NIST27). Em cada linha, da esquerda para a direita, tem-se as imagens originais, as áreas obtidas com com $\mathbf{n}_{it}$ = 300, 130, 500, 200 e 300, respectivamente, e as segmentações finais.

samente. Trabalhos futuros devem ser realizados para evitar estes problemas, provavelmente focando em processos para homogeneizar o fundo das imagens.

Além disto, o método semiautomático para detectar a área de impressões digitais latentes também obteve resultados relativamente bons, sendo

a definição manual da máscara inicial retangular bastante simples. As dificuldades com este tipo de imagens são maiores principalmente pelo fato de serem mais frequentes as regiões com fundo bastante escuro. Trabalhos futuros deverão seguir no mesmo sentido que os de imagens de impressões digitais roladas, focando inclusive em comparações com outras técnicas para a validação da proposta.

8. Agradecimentos

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pelo auxílio financeiro à pesquisa realizada.

Referências

Chan, T.F. & Vese, L.A., Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

Feng, J. & Jain, A.K., Fingerprint reconstruction: From minutiae to phase. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):209–223, 2011.

Fukunaga, K., *Introdution to Statistical Pattern Recognition*. San Diego: Academic Press, 1972.

Gonzalez, R.C. & Woods, R.E., *Digital Image Processing*. 2a edição. Upper Saddle River, USA: Prentice-Hall, 2002.

Henry, E.R., *Classification and Uses of Finger Prints*. London, UK: George Routledge and Sons, 1900.

Kass, M.; Witkin, A. & Terzopoulos, D., Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1987.

Laufer, B., History of the fingerprint system. In: Smithsonian Report for 1912. Washington, USA: U.S. Government Printing Office, p. 631–652, 1913.

Maio, D.; Maltoni, D.; Cappelli, R.; Wayman, J.L. & Jain, A.K., FVC 2004: Third fingerprint verification competition. In: Zhang, D. & Jain, A.K. (Eds.), *Proceedings of the International Conference on Biometric Authentication*. Hong Kong, v. 3072 de *Lecture Notes in Computer Science*, p. 1–7, 2004.

Maltoni, D.; Maio, D.; Jain, A.K. & Prabhakar, S., *Handbook of Fingerprint Recognition*. 2a edição. New York, USA: Springer, 2009.

Mumford, D. & Shah, J., Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

NIST, , *Fingerprint Minutiae from Latent and Matching Tenprint Images*. NIST Special Database 27, 2010. http://www.nist.gov/srd/nistsd27.htm, Gaithersburg, USA.

Oliveira, M.W.S., *Detecção da Área e Extração do Campo de Orientação em Imagens de Impressão Digital*. Dissertação de mestrado, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, SP, 2011.

Osher, S. & Sethian, J.A., Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

Thai, R., *Fingerprint Image Enhancement and Minutiae Extraction*. Honours programme, School of Computer Science and Software Engineering, University of Western Australia, Perth, Australia, 2003.

Weixin, B.; Deqin, X. & Yi-Wei, Z., Fingerprint segmentation based on improved active contour. In: *Proceedings of the International Conference on Networking and Digital Society*. v. 2, p. 44–47, 2009.

Yoon, S.; Feng, J. & Jain, A.K., On latent fingerprint enhancement. In: Kumar, B.V.K.V.; Prabhakar, S. & Ross, A.A. (Eds.), *Biometric Technology for Human Identification VII*. Bellingham, USA: International Society for Optical Engineering, v. 7667 de *Proceedings of SPIE*, p. 766707–766707–10, 2010.

Zheng, X.; Wang, Y. & Zhao, X., Fingerprint image segmentation using active contour model. In: *Proceedings of the Fourth International Conference on Image and Graphics*. p. 437–441, 2007.

Notas Biográficas

Marcos William da Silva Oliveira é mestre em Matemática (UNESP/São José do Rio Preto, 2011) e atualmente Professor do Centro Universitário da Fundação Educacional de Barretos (UNIFEB). Tem interesse na área de matemática computacional, processamento de imagens e biometria.

Inês Aparecida Gasparotto Boaventura é mestre em Ciências da Computação (USP, 1992), doutor em Engenharia Elétrica (USP, 2010) e atualmente é Professor Assistente da Universidade Estadual Paulista (UNESP/São José do Rio Preto). Tem interesse na área de processamento de imagens, visão computacional, sistemas inteligentes e biometria.

Maurílio Boaventura é mestre em Ciências da Computação e Matemática Computacional (USP, 1989), doutor em Matemática Aplicada (UNICAMP, 1998) e livre-docente (UNESP, 2005). Atualmente é Professor Adjunto da Universidade Estadual Paulista (UNESP/São José do Rio Preto). Tem interesse na área de eliminação de ruídos, equações diferenciais parciais, processamento de imagens, retoque digital, diferenças finitas e métodos numéricos.

Capítulo 10

Verificação Facial em Vídeos Capturados por Dispositivos Móveis

Tiago de Freitas Pereira e Marcus de Assis Angeloni*

Resumo: Nos últimos anos, observa-se uma crescente adoção de dispositivos móveis, como *smartphones*, em tarefas antes executadas apenas por computadores. Com isto, uma maneira rápida e segura de acesso a informações pessoais e corporativas nestes aparelhos é um desafio. Este capítulo apresenta uma abordagem de verificação facial usando câmeras de *smartphones*, como alternativa para autenticação. O método utiliza *Local Binary Patterns* (LBP) para descrição da face e Modelo de Misturas Gaussianas (GMM) como classificador. Os experimentos realizados na base de dados MOBIO e a aplicação de teste criada para *smartphones* Android trazem resultados promissores e mostram a viabilidade da técnica neste cenário.

Palavras-chave: Biometria, Verificação facial, Autenticação baseada em vídeos, *Smartphones*.

Abstract: *In recent years, an increasing use of mobile devices, such as smartphones, has been observed in tasks previously performed only by computers. Consequently, a quick and secure access to personal and corporate information in these devices is a challenge. This chapter presents a face verification approach using smartphones cameras, as an alternative for authentication. The method uses Local Binary Patterns (LBP) for face description and Gaussian Mixture Models (GMM) as classifier. Experiments performed with the MOBIO database and the test application created for Android smartphones showed promising results and demonstrate the technical feasibility of the method in this scenario.*

Keywords: *Biometrics, Face verification, Video-based authentication, Smartphones.*

*Autor para contato: massis@cpqd.com.br

1. Introdução

Observa-se, nos últimos anos, 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*. *Smartphone* é um telefone celular dotado de funcionalidades e acessórios extras, que possui uma capacidade de processamento de dados superior aos telefones celulares convencionais, maior número de conexões disponíveis e um sistema operacional sobre o qual é possível criar novas aplicações. Segundo projeção divulgada no início de 2012 pela consultoria IDC, especializada no mercado de tecnologia e telecomunicações, serão vendidos no Brasil perto de 15,4 milhões de unidades destes aparelhos em 2012, o que corresponde a um crescimento de 73% no ano. O estudo aponta ainda que em 2011 o mercado de *smartphones* bateu recorde, e chegou a marca de aproximadamente 9 milhões de aparelhos vendidos, um aumento de 84% em relação ao ano de 2010 (IDC Brasil Releases, 2012). Um outro estudo, realizado pelo banco Credit Suisse, aponta que as vendas globais de *smartphones* alcançarão 1 bilhão de unidades em 2014 (FOLHA.com, 2012).

Com a popularização dos *smartphones*, eles passam a ser mais visados por criminosos e fraudadores e, por apresentarem dimensões e peso reduzidos, o risco de perda e furto destes aparelhos é maior do que de computadores pessoais, podendo resultar em prejuízo financeiro e violação de dados pessoais e corporativos. Diante deste cenário, proporcionar acesso rápido, fácil e principalmente seguro é um desafio nesta plataforma móvel.

Embora as senhas tradicionais sejam onipresentes como mecanismos de autenticação, elas apresentam uma vulnerabilidade intrínseca, associada ao fato de lidar com dois princípios antagônicos: elas devem ser fáceis de lembrar e difíceis de adivinhar. É um comportamento comum do 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.

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 ganhado 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 (Jain et al., 2004). No entanto, pouco foi desenvolvido para dispositivos móveis, em parte pelo hardware limitado que até pouco tempo atrás era um grande impeditivo para adaptação desta 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 possível 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.

Contudo, ao propor um sistema de verificação facial para *smartphones* é necessário levar em conta que a iluminação, pose e o cenário de captura não serão controlados e a forma de tratamento tem grande impacto no desempenho do sistema. Para contemplar estes fatores, os experimentos foram executados sobre a base de dados MOBIO (Marcel et al., 2010b), que é uma base de dados pública, capturada com câmeras de dispositivos móveis em cenários não controlados.

O texto está estruturado da seguinte forma: a Seção 2 apresenta o conceito de biometria; a Seção 3 apresenta os conceitos envolvidos no reconhecimento biométrico utilizando a face; a Seção 4 descreve as diferentes etapas do sistema proposto; a Seção 5 descreve a base de dados, o protocolo de testes e as configurações de sistema adotados; a Seção 6 apresenta os resultados obtidos e, por fim, a Seção 7 apresenta a conclusão e possíveis trabalhos futuros.

2. Biometria

Biometria é a ciência de reconhecer a identidade de uma pessoa baseada nos atributos físicos e comportamentais do indivíduo, tais como a face, as impressões digitais, a voz e a íris (Jain et al., 2008).

As características humanas para serem utilizadas em uma aplicação biométrica devem satisfazer alguns requisitos, dentre eles a universalidade (toda pessoa deve possuí-la), unicidade (ela deve permitir distinguir as pessoas), permanência (ela não deve se alterar demasiadamente ao longo do tempo) e coletabilidade (ela deve poder ser medida quantitativamente). Como nenhum dado biométrico consegue atender todos os requisitos com eficácia, a escolha de qual utilizar deve levar em conta a natureza e exigências da aplicação (Jain et al., 1998).

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 uma referência biométrica previamente gravada no banco de dados. Ele pode operar de duas formas: como identificação, onde as características extraídas são comparadas com todas as referências biométricas cadastradas 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 a referência biométrica associada com a identidade declarada pelo usuário, verificando se ele é quem diz ser (comparação 1:1).

3. Reconhecimento de Face

O reconhecimento de face é uma tarefa que as pessoas realizam rotineiramente e sem esforço em suas relações diárias. O uso da face tem várias vantagens sobre outras tecnologias biométricas: ela é natural, não intrusiva, pode ser capturada a distância, não exige equipamentos sofisticados para seu funcionamento, e é de fácil uso (Li & Jain, 2011).

Embora a pesquisa em reconhecimento automático de face tenha iniciado em 1960, ela ainda apresenta um grande número de desafios para pesquisadores das áreas de visão computacional e reconhecimento de padrões, especialmente em imagens capturadas em ambientes não controlados (Li & Jain, 2011).

No entanto, nos últimos anos observou-se avanços significativos neste assunto, novas abordagens propostas, e uma grande atenção ao assunto, sobretudo, devido as crescentes preocupações pela segurança pública, a necessidade de verificação de identidade para acesso físico e lógico, e a necessidade de análise de rostos e técnicas de modelagem no gerenciamento de dados de multimídia e de entretenimento digital (Zhao et al., 2003). Parte desta 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.

As abordagens de verificação de faces podem ser divididas em duas categorias: as abordagens discriminativas e generativas (Rodriguez & Marcel, 2006). 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 típico de reconhecimento de face envolve algumas etapas, tais como: a captura da imagem ou de quadros de um vídeo, a detecção da face na imagem capturada, um pré-processamento visando reduzir o ruído introduzido pela iluminação e outros fatores do ambiente, a extração de características da imagem da face, e por fim, a criação da referência biométrica do usuário para gravação no banco de dados (durante o cadastro), ou a comparação das características extraídas com as referências biométricas previamente gravadas no banco de dados (durante a verificação ou identificação). A Figura 1 apresenta um diagrama com as etapas de um sistema de reconhecimento de face.

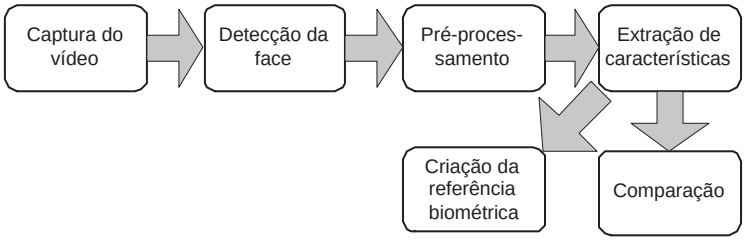


Figura 1. Diagrama de blocos de um sistema típico de reconhecimento de face.

4. Sistema Proposto

O método de verificação facial apresentado neste capítulo pertence a um projeto de pesquisa e desenvolvimento que envolve métodos de autenticação para uso em *smartphones*, aparelhos estes que apresentam desafios de usabilidade, de segurança, de memória, de capacidade de processamento e de tráfego de dados. Uma das frentes em andamento é a verificação através de características biométricas do usuário, e como resultado está sendo construída uma API (*Application Programming Interface*) para uso em *smartphones*.

O modelo de API que está sendo desenvolvido é inspirado na especificação da BioAPI 2.0 (BioApi Consortium, 2006). 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. Desta forma, as diferentes funções especificadas nela (como a captura, o processamento, a verificação, e a criação da referência biométrica) possuem diferentes graus de liberdade na implementação, ou seja, parte da implementação pode estar alocada no *smartphone* e parte em um servidor.

A abordagem de verificação de face proposta foi avaliada em uma base de dados pública com vídeos capturados por câmeras de *smartphones*, e foi desenvolvida uma aplicação para análise da viabilidade da técnica em *smartphones* da plataforma Android[1]. O sistema de verificação facial proposto é composto por quatro fases: detecção da face, pré-processamento, extração de características, e criação da referência biométrica/verificação.

[1] http://developer.android.com/sdk/index.html

4.1 Detecção da face

O processo de detecção da face consiste em localizar a região em que ela se encontra na imagem capturada e destacá-la do fundo. O método de detecção facial utilizado neste trabalho é baseado no algoritmo Viola & Jones (2004).

Este algoritmo procura em uma imagem características que codifiquem o padrão procurado, no caso da detecção da face são utilizadas as características de Haar, que exploram de modo multi-escala o contraste presente naturalmente na face e são representadas por formas geométricas, como se pode observar em algumas delas ilustradas na Figura 2. Posicionadas sob uma região da imagem, o valor de cada característica é calculado somando-se os pixels da parte branca e subtraindo pelo somatório dos pixels da parte preta da forma geométrica.

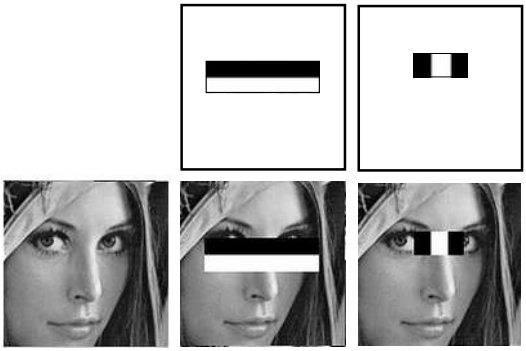


Figura 2. Representação geométrica das características de Haar (Viola & Jones, 2004).

Para tornar o detector mais simples e eficiente é selecionado apenas um subconjunto com as características mais discriminativas dentre as várias de Haar disponíveis, através do procedimento modificado do método *AdaBoost* (Tieu & Viola, 2000). Este método constrói um classificador "forte" com a combinação das características de Haar selecionadas (que são considerados classificadores "fracos" por dependerem de apenas uma característica). Outra etapa que aumenta significativamente a velocidade do detector é através do encadeamento destes classificadores fracos em uma estrutura de cascata, onde os primeiros desta estrutura são mais simples e de cálculo mais rápido, desta forma concentrando o esforço computacional apenas nas regiões promissoras da imagem. Um resultado positivo do classificador corrente dispara a avaliação pelo classificador seguinte, e um resultado negativo de qualquer classificador que compõe a estrutura levam a imediata rejeição da imagem. A Figura 3 mostra o funcionamento da estrutura em cascata de classificadores.

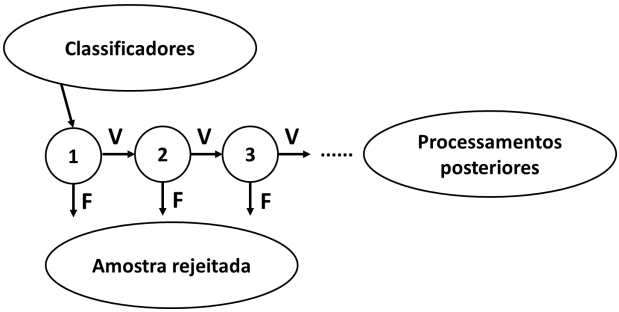


Figura 3. Árvore degenerada do AdaBoost (Viola & Jones, 2004).

A Figura 4 mostra o algoritmo Viola & Jones (2004) sendo aplicado sobre alguns quadros de vídeos da base de dados MOBIO, detectando a face em tempo real.

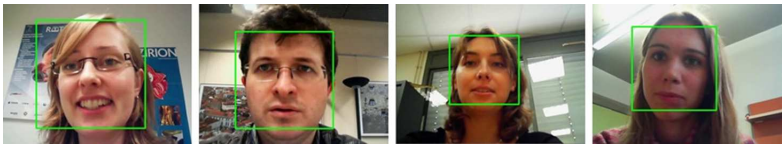


Figura 4. Algoritmo de detecção de face Viola & Jones (2004) aplicado sobre quadros dos vídeos da base MOBIO.

4.2 Pré-processamento

Após a detecção da face na imagem (ou nos quadros amostrados do vídeo) é necessário um pré-processamento da mesma, para amenizar os ruídos introduzidos pelo ambiente e tornar a imagem da face mais adequada para a extração de características.

Primeiramente, a imagem de face segmentada é convertida para escala de cinza, pois o descritor adotado não tira proveito da informação de cor. Depois, a imagem é redimensionada para um tamanho padrão (definido pela aplicação) através de interpolação bilinear, porque conforme ilustrado na Figura 4, a área segmentada pode apresentar dimensões diferentes.

Por fim, para amenizar a influência dos diferentes tipos de iluminação dos ambientes de captura, a imagem é normalizada fotometricamente, usando a seguinte sequência de passos: correção gama; filtro de Diferença de Gaussianas (DoG); normalização do intervalo de variações de saída; e compressão baseada em uma função sigmoidal para remoção de picos de

iluminação remanescentes. Uma descrição detalhada deste eficiente procedimento de normalização fotométrica pode ser encontrado em (Tan & Triggs, 2010).

Dois exemplos das etapas de pré-processamento são apresentados na Figura 5. Na Figura 5a são mostradas duas imagens de face segmentadas, sujeitas a diferentes tipos de iluminação. Na Figura 5b elas foram convertidas para escala de cinza, para então serem submetidas à normalização fotométrica, produzindo um resultado mais uniforme no que diz respeito a iluminação, conforme pode ser observado na Figura 5c.

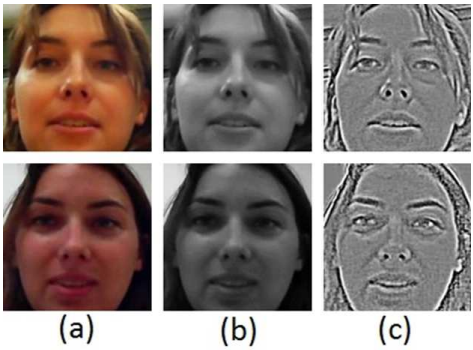


Figura 5. Pré-processamento das imagens de face. (a) Imagens segmentadas; (b) imagens convertidas para escala de cinza; (c) normalização fotométrica das imagens.

4.3 Extração de características

As características utilizadas neste trabalho para descrição da face são baseadas em operadores LBP (*Local Binary Pattern*). O operador LBP foi originalmente proposto por Ojala et al. (1996) para descrição de textura em imagens em escala de cinza. Este operador é calculado no nível dos pixels, em sua versão original usando um núcleo de tamanho 3×3, onde a intensidade dos pixels da vizinhança são comparados com a intensidade do pixel na posição central do núcleo, e os valores binários obtidos nas comparações são agrupados formando um único valor binário, que é o código LBP. O código LBP na forma decimal pode ser expresso da seguinte forma:

$$LBP(x_c, y_c) = \sum_{n=0}^{N-1} f(i_n - i_c)2^n, \tag{1}$$

onde i_c corresponde a intensidade de cinza do pixel central (x_c, y_c), N é o número de pontos amostrados na vizinhança do pixel, i_n é a intensidade

de cinza do n-ésimo pixel da vizinhança de i_c, e $f(x)$ é definido segundo a Equação 2.

$$f = \begin{cases} 0 & se \quad x < 0 \\ 1 & se \quad x \geq 0 \end{cases} . \tag{2}$$

Depois, Ojala et al. (2002) estendeu este operador para suportar vizinhanças de pixels com tamanhos, raios e formas variadas, permitindo lidar com texturas em diferentes escalas. A Figura 6 ilustra o cálculo do operador LBP e a distribuição dos pontos vizinhos considerando uma vizinhança circular de raio 2, onde os pontos amostrados que não coincidem com o centro de um pixel são bilinearmente interpolados.

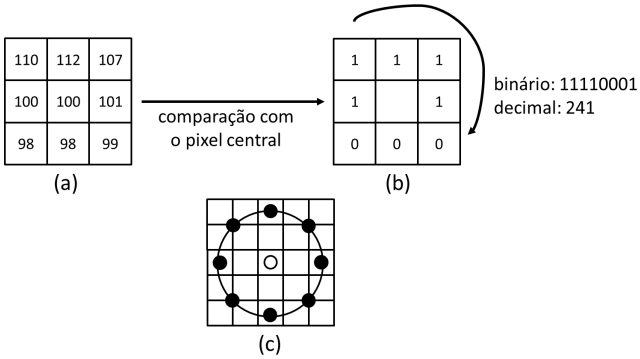


Figura 6. Operador LBP. (a) e (b) ilustram o operador LBP básico, onde cada pixel é comparado com seus vizinhos e apresenta como resultado um número binário. (c) Um exemplo de vizinhança circular (com 8 pontos de amostra e raio 2) (Ahonen et al., 2004).

Outra importante extensão apresentada por Ojala et al. (2002) foi o conceito de padrões uniformes. Um operador LBP é considerado um padrão uniforme se ele contém até duas transições de bits 0-1 ou 1-0 quando analisado como uma cadeia de bits circular, e segundo Ojala et al. (2002), aproximadamente 90% dos operadores LBP observados em seus experimentos são uniformes. Em termos espaciais, os padrões uniformes representam alguns padrões específicos de textura: pontos, planos, arestas, quinas e terminações de linhas. Em uma configuração de operadores LBP com vizinhança 3×3 (representação em 8 bits), existem 58 padrões com até duas transições de bits. A Figura 7, extraída de Chan (2008), mostra todos os padrões uniformes disponíveis em uma configuração com 8 pontos vizinhos.

Ahonen et al. (2004, 2006) adotam a seguinte notação para representar a configuração do operador LBP em uso: $LBP^{u2}_{P,R}$, onde o subscrito representa a configuração da vizinhança utilizada, com P pontos de amostragem

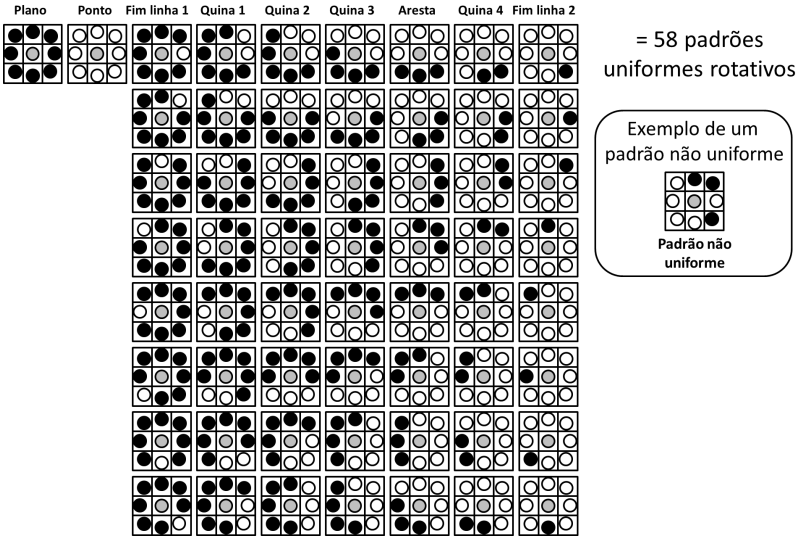


Figura 7. Todos os padrões uniformes para uma configuração do LBP com 8 vizinhos (Chan, 2008).

em um círculo de raio R, e o sobrescrito $u2$ denota que são considerados apenas os padrões uniformes, e os padrões não uniformes são agrupados em uma classe extra.

Recentemente, o LBP foi adaptado para representação de imagens de face e obteve resultados promissores (Ahonen et al., 2004). A descrição de face utilizando LBP consiste no cálculo de um histograma de operadores LBP, e bons resultados foram alcançados utilizando a configuração com 8 pixels em uma vizinhança circular de raio 2 (Ahonen et al., 2004, 2006; Rodriguez & Marcel, 2006). Esta abordagem de descrição de face baseada em histogramas LBP também tira proveito das extensões propostas por Ojala et al. (2002), ou seja, utilizando apenas padrões uniformes o número de classes do histograma é consideravelmente reduzido, visto que todos padrões não uniformes são agrupados em uma única classe. Esta é a principal vantagem do uso de padrões uniformes, a redução de dimensionalidade que ele promove no espaço de características, visto que ao invés das 256 classes possíveis em cada histograma LBP calculado, serão computadas apenas 59 classes.

Uma importante modificação no cálculo do LBP para representação de face é a ideia de particionar a imagem de face em pequenos blocos (os quais podem ser sobrepostos ou não) e calcular o histograma LBP para

cada bloco individualmente, depois os concatenando em um único, desta forma retendo a informação espacial. Portanto, a imagem de face é descrita em três diferentes níveis: nível de pixel, com o cálculo de cada operador LBP individualmente; nível regional, com o cálculo do histograma de cada bloco; e um nível global, com a concatenação de todos os histogramas de bloco (Rodriguez & Marcel, 2006). A Figura 8 mostra os três diferentes níveis de descrição da face disponíveis nesta abordagem.

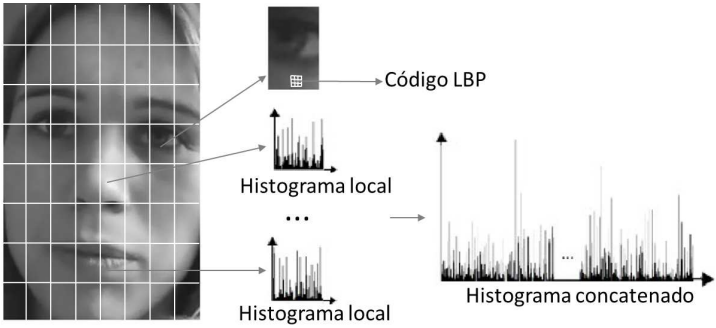


Figura 8. Descrição da face nos três diferentes níveis: nível de pixel (código LBP), nível regional (histograma local) e nível global (histograma concatenado) (Rodriguez & Marcel, 2006).

Embora a informação espacial seja preservada quando a imagem é particionada em blocos, parte da informação é perdida nas bordas dos blocos. Para evitar isto, os operadores LBP são primeiro calculados para a imagem toda e depois a matriz LBP resultante é dividida em blocos. Assim, tanto a informação espacial quanto as bordas dos blocos são preservadas.

4.4 Criação da referência biométrica e verificação de usuários

O método generativo implementado neste trabalho é baseado no Modelo de Misturas Gaussianas (GMM), que é um modelo estocástico que descreve uma função densidade probabilidade genérica como um produto de componentes gaussianas multivariadas.

A abordagem de GMM utilizada na verificação de face usando características LBP foi baseada nos trabalhos de Rodriguez & Marcel (2006) e Cardinaux et al. (2004) e opera da seguinte forma: dada uma sequência de histogramas LBP procedentes de um indivíduo, sendo que cada um representa um quadro do vídeo avaliado, podemos calcular a probabilidade da ocorrência desta sequência observada pela seguinte equação:

$$P(X|\lambda) = \prod_{t=1}^{Nv} p(\vec{x}_t|\lambda), \tag{3}$$

onde $X = \{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_{N_v}\}$ é a sequência de observações, $\vec{x}_t$ é uma única observação, N_v é o número de observações, λ é o modelo GMM associado ao indivíduo e $p(x)$ é a função probabilidade associado com o GMM. Esta função descreve a probabilidade de observações e é dada por:

$$p(\vec{x}_t|\lambda) = \sum_{j=1}^{N_G} m_j \eta\left(\vec{x}|\vec{\mu}_j, \Sigma_j\right), \tag{4}$$

onde $\eta\left(\vec{x}|\vec{\mu}_j, \Sigma_j\right)$ é uma função gaussiana N-dimensional com média μ_j e matriz de covariância diagonal Σ_j, N_G é o número de gassianas e m_j é o peso de uma gaussiana (Cardinaux et al., 2004).

Os parâmetros do modelo GMM (m_k, μ_k, Σ_k) são estimados a partir das amostras de treinamento. A ideia básica é iterativamente atualizar os parâmetros do modelo a fim de maximizar a probabilidade da observação das amostras de treinamento. O algoritmo K-médias (Linde et al., 1980) é usado para inicializar os parâmetros do modelo; então, o algoritmo de Maximização da Expectativa (EM) (Dempster et al., 1977) é executado para maximizar a probabilidade em relação aos dados de treinamento.

Um melhor desempenho da abordagem utilizando GMM pode ser obtido se os modelos dos indivíduos forem derivados de um modelo comum, chamado *Universal Backgroung Model* (UBM). O UBM é treinado com o procedimento descrito anteriormente, e os dados de treinamento, neste caso, são formados por amostras de vários indivíduos diferentes. Este modelo comum é a base para o procedimento de adaptação do modelo que é usado para gerar os modelos dos indivíduos.

O algoritmo usado para adaptar os modelos dos indivíduos é uma modificação do algoritmo EM, conhecido como adaptação *Maximum A Posteriori* (MAP).

Neste trabalho, apenas as médias das funções gaussianas N-dimensionais do UBM são adaptadas. A adaptação média é descrita pela Equação 5.

$$\mu_k = 1 - \alpha_k u_k^w + \alpha_k \frac{\sum_{t=1}^{T} P(k|\vec{x}_t)\vec{x}_t}{\sum_{t=1}^{T} P(k|\vec{x}_t)}, \tag{5}$$

onde μ_k é a média da k-ésima gaussiana, u_k^w é a média correspondente na gaussiana do UBM e α_k é um fator de adaptação (Cardinaux et al., 2004).

A dinâmica do fator de adaptação α_k segue uma estratégia diferente da descrita em Rodriguez & Marcel (2006) e Cardinaux et al. (2004). Seguindo

a ideia apresentada por Reynolds et al. (2000), o valor de α_k é adaptado para cada gaussiana, de acordo com a Equação 6:

$$\alpha_k = \frac{n_k}{n_k + R}, \tag{6}$$

onde n é a medida probabilística que representa o número de vezes que a gaussiana k é ativada pelas amostras de treinamento e R é um fator de relevância fixado pelo usuário. O uso desta contagem probabilística torna possível para o algoritmo controlar a quantidade de adaptações das misturas. Componentes da mistura com alta contagem probabilística são adaptadas para os novos dados, enquanto que as com contagem probabilística baixa levam a uma pequena adaptação. O fator de relevância R controla a importância dos novos dados para a adaptação.

Durante uma verificação a pontuação é calculada de acordo com a Equação 7.

$$\text{pontuação} = \log(P(X|\lambda_c)) - \log(P(X|\lambda_w)), \tag{7}$$

onde $P(X|\lambda_c)$ é a probabilidade de X no modelo do usuário λ_c e $P(X|\lambda_w)$ é a probabilidade de X no UBM. Visto que as probabilidades estão no domínio logarítmico, a operação realizada na Equação 7 desempenha o papel de uma normalização de pontuação, tornando pontuações que vieram a partir de modelos de diferentes usuários comparáveis.

5. Experimentos

Esta seção apresenta os experimentos que foram realizados para avaliar a abordagem proposta com dois enfoques: primeiro avaliando o desempenho do sistema sobre uma base grande e desafiadora, com vídeos capturados por câmeras de *smartphones* em ambientes não controlados (com notável variação de iluminação, cenário de fundo e pose); e segundo, avaliando a viabilidade de todas as etapas de reconhecimento de face serem executadas em um *smartphone*, que retrata o pior cenário da API, onde todo o processamento é realizado no celular.

5.1 Base de dados e protocolo de avaliação

Capturada em cenários do mundo real usando *smartphones*, a base de dados MOBIO é uma grande e desafiadora base de dados multimodal (McCool & Marcel, 2009). Apresentada durante a competição *"Face and Speaker Verification Evaluation"* no ICPR 2010 (Marcel et al., 2010a), ela é composta por 15444 amostras de áudio e vídeo de 162 usuários (sendo 105 homens e 57 mulheres). Para os experimentos do presente trabalho, apenas os vídeos foram utilizados. A Figura 9 apresenta algumas imagens extraídas da base de dados MOBIO, onde é possível notar a variabilidade de iluminação, fundo e pose.

Figura 9. Imagens da base de dados MOBIO.

O protocolo definido para a base de dados MOBIO separa a base em três partições de usuários não sobrepostas (treinamento, desenvolvimento e teste), onde para cada usuário há seis sessões de vídeos. Os vídeos dos usuários do conjunto de treinamento podem ser usados da forma que convier para a abordagem proposta (por exemplo, para compor o modelo de mundo (UBM), ou o espaço inicial de faces). Os usuários do conjunto de desenvolvimento são utilizados para cálculo do ponto de operação (limiar) do algoritmo de verificação, que será utilizado pelo conjunto de teste para avaliar o desempenho do algoritmo.

Como métrica de avaliação dos sistemas de verificação facial sobre a base MOBIO, o protocolo emprega a Taxa de Erro Total Médio (HTER), a qual é calculada seguindo a Equação 8.

$$HTER(\tau, D) = \frac{FAR(\tau, D) + FRR(\tau, D)}{2}, \tag{8}$$

onde τ é o limiar utilizado (definido pela base de desenvolvimento), D é o conjunto de dados, FAR é a taxa de falsa aceitação e FRR é a taxa de falsa rejeição.

Essa taxa deve ser obtida separadamente para os usuário do sexo masculino e do feminino, conforme define o protocolo da base de dados MOBIO (McCool & Marcel, 2009).

5.2 Configurações dos experimentos

Para cada vídeo da base de dados MOBIO foram selecionados quadros igualmente espaçados no tempo, com intervalo de 5 quadros (valor escolhido empiricamente). Para cada um dos quadros foi detectada a face usando a

implementação da biblioteca OpenCV[2] do algoritmo Viola & Jones (2004) descrito na Seção 4.1.

As imagens de face segmentadas foram convertidas para escala de cinza, redimensionadas para 108 × 108 pixels e normalizadas fotometricamente, conforme apresentado na Seção 4.2.

Depois do pré-processamento, os histogramas LBP foram extraídos das imagens, utilizando a configuração $LBP^{\mu 2}_{8,2}$ para os operadores LBP. Foram geradas matrizes de 104 × 104 operadores LBP para cada imagem, e cada uma delas foi particionada em 64 blocos (sub-matrizes) com 13 × 13 operadores LBP cada. O histograma LBP global de cada imagem foi obtido concatenando os histogramas de cada um dos blocos da mesma, gerando um descritor com dimensionalidade 3776 (64 × 59).

A implementação do GMM utilizada neste trabalho foi baseada em uma adaptação da *Speech Signal Processing Toolkit* (SPTK[3]). Foram gerados 2 UBMs utilizando o conjunto de treinamento da base de dados MOBIO e seguindo os passos descritos na Seção 4.4. O primeiro deles foi treinado usando apenas vídeos de mulheres (1890 vídeos), e o segundo foi gerado usando apenas vídeos de homens (4914 vídeos).

Durante a fase de treinamento a escolha do UBM para inicializar os parâmetros do GMM (médias, variâncias e pesos) foi baseada em gênero, ou seja, para treinar um usuário masculino, o modelo de UBM masculino foi selecionado, e para treinar um usuário feminino foi escolhido o UBM feminino, seguindo as regras estipuladas pelo protocolo MOBIO.

Os impostores usados para os testes de falsa aceitação foram selecionados de acordo com a divisão dos conjuntos da base de dados MOBIO. Desta forma, para usuários pertencentes ao conjunto de desenvolvimento foram usados apenas impostores do conjunto de desenvolvimento, e para usuários do conjunto de testes foram usados apenas impostores do conjunto de teste.

O número de misturas gaussianas adotado no GMM e o fator R, foram escolhidos baseados em simulações anteriores, e os valores usados foram 1 e 4 respectivamente.

Para verificar o desempenho computacional deste algoritmo embarcado em um *smartphone*, foram selecionadas algumas amostras de vídeos e elas foram processadas pelo aparelhos Galaxy S e Galaxy S II da Samsung, ambos com o sistema operacional Android 2.3. 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.

[2] http://opencv.willowgarage.com
[3] http://sp-tk.sourceforge.net

6. Resultados

A Tabela 1 exibe os resultados em termos de HTER seguindo o protocolo descrito na Seção 5.1, e a Figura 10 exibe os resultados através da curva DET (*Detection Error Tradeoff*) (Martin et al., 1997) no conjunto de teste da base de dados do MOBIO. A análise de resultados utilizando a curva DET é muito comum em biometria, pois com ela é possível observar os resultados ponderando a influência dos dois tipos de erro (falsos positivos e falsos negativos).

Tabela 1. Resultados dos experimentos utilizando o protocolo MOBIO.

Experimento	**HTER**
Base de dados masculina	23,54%
Base de dados feminina	25,54%

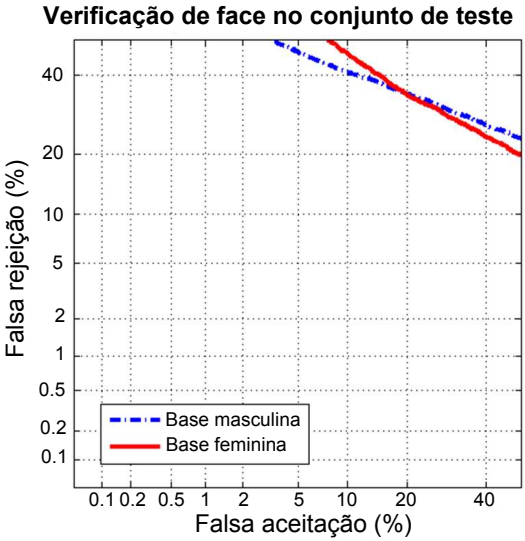


Figura 10. Curva DET dos diferentes experimentos de verificação facial no conjunto de teste.

Os resultados obtidos executando o protocolo da base de dados MOBIO apesar de, aparentemente, apresentarem valores altos em termos de HTER são resultados competitivos com os publicados na literatura, dada a complexidade da base de dados. Os resultados apresentados no *"Face*

and Speaker Verification Evaluation" do ICPR 2010 apresentaram taxas que variaram de 10% a 30% em termos de HTER (Marcel et al., 2010b).

Conforme descrito na Seção 5.2, algumas amostras de vídeo de diferentes tamanhos foram selecionadas para verificação de desempenho computacional embarcado em *smartphones*. Para isto foram feitos testes de criação da referência biométrica e verificação dos usuários. Como métrica de desempenho foi utilizada o tempo de execução em milissegundos. A Tabela 2 mostra os tempos médios de execução de cada operação embarcada em dois diferentes *smartphones* da plataforma Android.

Tabela 2. Tempo médio de execução das operações de criação da referência biométrica e verificação facial embarcadas em *smartphones* (em milissegundos).

Smartphone	**Criação da ref. biométrica**	**Verificação**
Samsung Galaxy S	490,6	135,3
Samsung Galaxy S II	325,0	93,3

Os tempos de execução das operações (criação da referência biométrica e verificação) assinalam um desempenho computacional promissor embarcado em *smartphones*.

É importante salientar que, neste trabalho, a escolha das técnicas utilizadas focaram no compromisso em equilibrar desempenho computacional e taxas de acerto.

7. Conclusão

Este trabalho explorou uma alternativa para autenticação biométrica através da face em dispositivos móveis, 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. Experimentos com a base de dados de face MOBIO mostraram taxas de erro competitivas com a literatura em um ambiente próximo ao cenário real de uso da biometria de face em um *smartphone*. A técnica utilizada apresentou resultados promissores em tal ambiente. Vale ressaltar que a principal contribuição deste trabalho foi estabelecer um compromisso entre taxas de acerto do algoritmo de verificação facial e um tempo de resposta adequado para uma aplicação real. Como atividade futura está previsto o aprimoramento da detecção da face juntamente com o seu registro, bem como novas formas de descrição da face.

Agradecimentos

A elaboração deste capítulo foi feita dentro do Projeto BIOMODAL, desenvolvido pela equipe da Fundação CPqD, realizado com apoio da FINEP e recursos do FUNTTEL.

Referências

Ahonen, T.; Hadid, A. & Pietikainen, M., Face recognition with local binary pattern. In: *Proceedings of the Eighth European Conference Computer Vision*. p. 469–481, 2004.

Ahonen, T.; Hadid, A. & Pietikainen, M., Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 28:2037–2041, 2006.

BioApi Consortium, , *Biometric Application Programming Interface, Part 1: BioAPI Specification*. ISO/IEC 19784-1, 2006.

Cardinaux, F.; Sanderson, C. & Bengio, S., Face verification using adapted generative models. In: *Proceedings of the International Conference on Automatic Face and Gesture Recognition*. p. 825–830, 2004.

Chan, C.H., *Multi-Scale Local Binary Pattern Histogram for Face Recognition*. Phd thesis, Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK, 2008.

Dempster, A.P.; Laird, N.M. & Rubin, D.B., Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society – Series B*, 39(1):1–38, 1977.

FOLHA.com, , Vendas de smartphones alcançarão 1 bilhão de unidades em 2014, diz banco. http://www1.folha.uol.com.br/tec/1075341-vendas-de-smartphones-alcancarao-1-bilhao-de-unidades-em-2014-diz-banco.shtml, 2012. Acesso em abril de 2012.

IDC Brasil Releases, , Estudo da IDC revela que foram vendidos aproximadamente 9 milhões de smartphones no Brasil em 2011. http://www.idclatin.com/news.asp?ctr=bra&id_release=2213, 2012. Acesso em abril de 2012.

Jain, A.K.; Bolle, R. & Pankanti, S. (Eds.), *Biometrics: Personal Identification in Networked Society*. Norwell, USA: Kluwer Academic Publishers, 1998.

Jain, A.K.; Flynn, P. & Ross, A.A. (Eds.), *Handbook of Biometrics*. New York, USA: Springer-Verlag, 2008.

Jain, A.K.; Ross, A. & Prabhakar, S., An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004.

Li, S.Z. & Jain, A.K. (Eds.), *Handbook of Face Recognition*. 2a edição. New York, USA: Springer-Verlag, 2011.

Linde, Y.; Buzo, A. & Gray, R., An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28(1):84–95, 1980.

Marcel, S.; McCool, C.; Matejka, P.; Ahonen, T. & Cernocky, J., Mobile Biometry (MOBIO) Face and Speaker Verification Evaluation. Research Report RR-09-2010, IDIAP Research Institute, Martigny, Switzerland, 2010a.

Marcel, S.; McCool, C.; Matejka, P.; Ahonen, T.; Cernocky, J.; Chakraborty, S.; Balasubramanian, V.; Panchanathan, S.; Chan, C.H.; Kittler, J.; Poh, N.; Fauve, B.; Glembek, O.; Plchot, O.; Jancik, Z.; Larcher, A.; Levy, C.; Matrouf, D.; Bonastre, J.F.; Lee, P.H.; Hung, J.Y.; Wu, S.W.; Hung, Y.P.; Machlica, L.; Mason, J.; Mau, S.; Sanderson, C.; Monzo, D.; Albiol, A.; Nguyen, H.V.; Bai, L.; Wang, Y.; Niskanen, M.; Turtinen, M.; Nolazco-Flores, J.A.; Garcia-Perera, L.P.; Aceves-Lopez, R.; Villegas, M. & Paredes, R., On the results of the first mobile biometry (MOBIO) face and speaker verification evaluation. In: *Proceedings of the Twentieth International Conference on Recognizing Patterns in Signals, Speech, Images, and Videos*. Heidelberg, Germany: Springer-Verlag, p. 210–225, 2010b.

Martin, A.; Doddington, G.; Kamm, T.; Ordowski, M. & Przybocki, M., The DET curve in assessment of detection task performance. In: *Proceedings of the 5th European Conference on Speech Communication and Technology*. v. 4, p. 1895–1898, 1997.

McCool, C. & Marcel, S., MOBIO Database for the ICPR 2010 Face and Speech Competition. Communication Report Com-02-2009, IDIAP Research Institute, Martigny, Switzerland, 2009.

Ojala, T.; Pietikainen, M. & Harwood, D., A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

Ojala, T.; Pietikainen, M. & Maenpaa, T., Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

Reynolds, D.A.; Quatieri, T.F. & Dunn, R.B., Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10(1-3):19–41, 2000.

Rodriguez, Y. & Marcel, S., Face authentication using adapted local binary pattern histograms. In: *Proceedings of European Conference on Computer Vision*. p. 321–332, 2006.

Tan, X. & Triggs, B., Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, 19(6):1635–1650, 2010.

Tieu, K. & Viola, P., Boosting image retrieval. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE Press, v. 1, p. 228–235, 2000.

Viola, P. & Jones, M.J., Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

Zhao, W.; Chellappa, R.; Phillips, P.J. & Rosenfeld, A., Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):433–440, 2003.

Notas Biográficas

Tiago de Freitas Pereira é graduado em Ciência da Computação (USP, 2009) e é pesquisador da Fundação CPqD. Atualmente é mestrando no programa de pós-graduação da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas (UNICAMP).

Marcus de Assis Angeloni é graduado em Ciência da Computação (UNESP, 2008) e é pesquisador da Fundação CPqD. Atualmente é mestrando no programa de pós-graduação em Ciência da Computação da Universidade Estadual Paulista (UNESP).

Capítulo 11

Detecção de Tipos de Tomadas em Vídeos de Futebol Utilizando a Divergência de Kullback-Leibler

Guilherme Alberto Wachs-Lopes*, Werner Fukuma e Paulo S. Rodrigues

Resumo: Atualmente, os sistemas de TV Digital apresentam vários desafios na área de análise de vídeo e imagem. Entre eles, há a quantificação do tempo de exposição de logotipos em eventos esportivos. Uma maneira tradicional de lidar com estes desafios é identificar o tipo de tomada como câmera principal e câmera secundária. Os passos seguintes, necessários para extração dos logotipos na cena, dependem diretamente da qualidade desta classificação. Trabalhos recentes mostram que a análise de histograma baseado no sistema HSV gera resultados com bom desempenho. Além disto, pesquisas têm mostrado que a análise de imagens e vídeos usando entropia não-extensiva como uma ferramenta de classificação é uma nova e promissora abordagem de investigação. Neste trabalho, propõe-se o uso de entropia não-extensiva para um classificador binário de tomadas de câmera principal. Os resultados confirmam os desempenhos encontrados na literatura.

Palavras-chave: Processamento de vídeos, Teoria da informação, Entropia, Classificador binário.

Abstract: *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. A traditional way to face these challenges is to classify the camera shots as main camera shot and secondary camera shot. The main camera shot detection is a step that all the following processes depend of. Recent works show that the histogram analysis based on the HSV system generates results with good performance. Furthermore, researches have shown that the video and image analysis using non-extensive entropy as a classification tool is a new and promising 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.*

Keywords: *Video processing, Information theory, Entropy, Binary classifier.*

*Autor para contato: guilhermewachs@gmail.com

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 dos próprios comerciais e a segunda está relacionada com a detecção de cenas. Outros trabalhos, como Hsu et al. (2003) e 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 caracteriza-se por imagens detalhadas, com foco em um determinado jogador ou objeto. Nestes tipos de tomadas, as propagandas são geralmente 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 ou placares ao redor do campo. O reconhecimento destas tomadas pode ser útil para eliminar partes do vídeo onde não há propagandas. Além disto, 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, Watve & Sural (2008) propuseram um método que, inicialmente, detecta o tipo de cada cena e utiliza segmentação para encontrar possíveis regiões de interesse. Tais regiões podem potencialmente conter *outdoors*. Finalmente, utilizando *template matching*, os *outdoors* são reconhecidos individualmente.

Trabalhos recentes na área de análise de imagens (Rodrigues & 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 os histogramas dos *frames* do vídeo e do histograma médio de uma classe melhora a detecção dos *frames* como câmera 1 e não-câmera 1, tratando-se esta a conclusão e contribuição principal deste trabalho.

Assim, é proposto neste trabalho a medida da informação mútua entre *frames* de um vídeo com a média de uma classe, calculadas sob a distribuição de probabilidade da mantissa em um sistema 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 capítulo está 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 os resultados obtidos são apresentados e discutidos.

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. Estes sistemas possuem uma das propriedades físicas mais conhecidas, chamada entropia, cujo tipo mais estudado é a entropia de *Shannon*, dada pela equação 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, sob a restrição que $0 \leq p_i \leq 1$ e $\sum p_i = 1,0$.

Tome-se como um exemplo um sistema que contém 2 estados: o lançar de uma moeda. Neste tipo de sistema, se a moeda não for "viciada", temos as probabilidades $p_1 = 0,5$ e $p_2 = 0,5$. Neste caso, o sistema se comporta de forma totalmente aleatória e não temos certeza em qual estado a moeda pode cair. Desta forma, o sistema é imprevisível e a quantidade de informação é máxima. Porém, caso a moeda seja viciada e caia mais com o mesmo lado no total de jogadas, temos um sistema previsível e a quantidade de informação é baixa. A Figura 1 ilustra um gráfico do resultado da entropia em função da probabilidade p_i. Note que a entropia máxima $S = log(w)$ é alcançada quando as probabilidades dos estados são iguais. Pode-se concluir então que a entropia está relacionada com a quantidade de desordem do sistema.

Dado o significado relevante de sua medida, a entropia chamou atenção de diversos cientistas, abrindo possibilidades de novas aplicações em diversas áreas. No final da década de 40, esta medida teve sua primeira aplicação na área da Teoria da Informação, proposta por Claude Shannon (Shannon, 1948). A ideia de Shannon era medir a quantidade de informação transmitida em uma mensagem (Equação 1). De forma mais específica, Shannon considerou um microestado (da termodinâmica) como sendo a probabilidade de um possível acontecimento. Se a probabilidade de uma mensagem

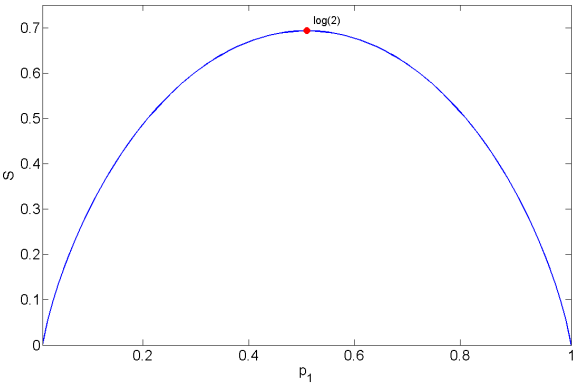


Figura 1. Entropia máxima para um sistema de dois estados.

ocorrer for pequena, então o sistema contém muita informação (problema da moeda não viciada). Porém, se uma mensagem ocorre muito frequentemente, o sistema terá pouca informação (problema da moeda viciada).

Uma propriedade importante da entropia de Shannon é conhecida por **aditividade**. Esta propriedade considera que, para dois sistemas totalmente independentes A e B, a entropia do sistema composto é dada por

$$S(A \oplus B) = S(A) + S(B) \tag{2}$$

onde $S(A)$ e $S(B)$ são as entropias dos sistemas A e B, consideradas independentes.

Contudo, a entropia de Shannon 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 fronteiras. Tais sistemas são chamados sistemas não-extensivos.

Partindo deste princípio, Tsallis (1988, 1999, 2001) propôs uma generalização da entropia tradicional, criando o conceito de entropia não-extensiva, definida por:

$$S_q = k\frac{1 - \sum_{i=1}^{n} p_i^q}{q - 1} \tag{3}$$

onde k é a constante de Boltzmann, n é o número de estados do sistema físico considerado, p_i, tal como na seção anterior, é a probabilidade do estado i ocorrer e q é o parâmetro entrópico ajustável ou parâmetro de não-extensividade. É importante notar que, quando q tende a 1, a equação 3 resume-se à entropia tradicional de Shannon, sendo portanto uma generalização da mesma.

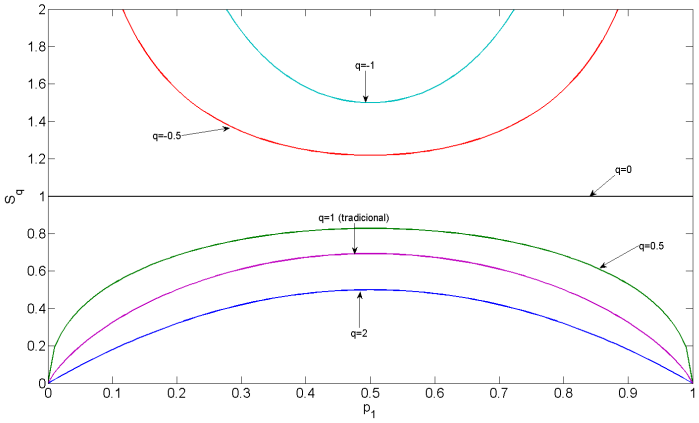


Figura 2. Distribuições de entropias para diferentes valores de q em um sistema de dois estados.

Da mesma forma como foi abordado anteriormente, a Figura 2 ilustra a entropia não-extensiva com diversos valores de q para o sistema de lançamento de moeda.

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** ou **ganho de informação de Kullback-Leibler**. A entropia relativa é definida como sendo:

$$D_{KL}(P, P') = \sum_{i=1}^{k} p_i \cdot \log \frac{p_i}{p'_i} \tag{4}$$

onde P e P' são as distribuições e k o número de estados do sistema físico considerado. É importante destacar que, para aplicar a equação 4, 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,

$$D_{KL}(p, p') \neq D_{KL}(p', p) \tag{5}$$

Desta forma, em Jeffreys (1939) foi proposta uma versão simétrica para entropia relativa:

$$D(p, p') = D_{KL}(p, p') + D_{KL}(p', p) \tag{6}$$

Borland et al. (1998) propuseram a generalização da entropia relativa para sistemas não-extensivos, adicionando o parâmetro entrópico q à comparação estatística entre duas distribuições, conforme a Equação 7:

$$D_{KL_q}(p, p') = \sum_{i=1}^{k} \frac{p_i^q}{1-q} \cdot (p_i^{1-q} - p_i'^{1-q}) \tag{7}$$

A vantagem da utilização da divergência de Kullback-Leibler estendida é a adição do parâmetro q como um ajuste fino na equação. Desta forma, podemos obter um q que maximiza os resultados.

4. Metodologia

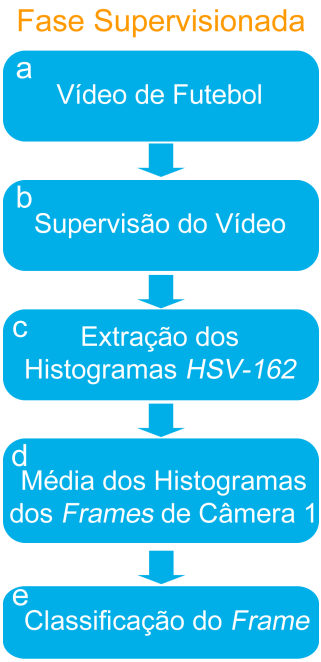
Neste trabalho, propomos a classificação dos *frames* como: tomadas de câmera principal (classe 1) e tomadas de câmera secundária (classe 2) (ver Figura 4). Então, pode-se considerar este sistema como um classificador binário.

De maneira genérica, 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. Este processo requer geralmente o uso de algoritmos computacionalmente pesados, muitas vezes inviáveis para processamento em 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.

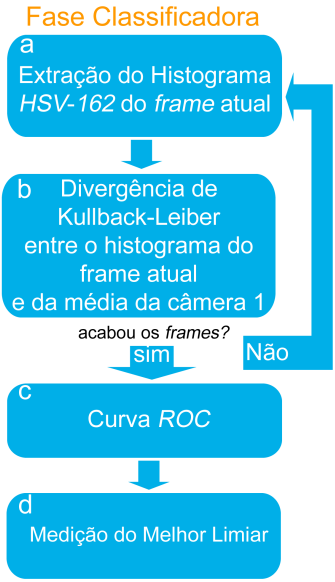
A maneira mais tradicional para se medir a quantidade de informação em uma distribuição é através da entropia clássica de Shannon. Porém, o surgimento da entropia não-extensiva de Tsallis permitiu a inclusão de um novo parâmetro que possibilitou o uso da entropia para sistemas onde a teoria clássica não era válida. Um exemplo é o trabalho de Rodrigues & Giraldi (2009) onde foi proposto um método de segmentação de imagens com cálculo automático do parâmetro q. Os resultados mostraram que esta é uma técnica promissora para o tratamento de imagens naturais.

Com isto em mente, propõe-se o fluxograma da Figura 3 para a classificação dos *frames* utilizando a teoria da informação.

A parte esquerda da Figura 3 refere-se à base de dados utilizada neste trabalho. Esta base é 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 Figura 4 mostra alguns quadros característicos da base.



3.1: Fluxograma da Fase Supervisionada da Metodologia. a) Entrada de um vídeo de futebol. b) Supervisão manual de cada *frame* do vídeo. c) Extração do histograma *hsv-162* de todos os *frames* do vídeos. d) Média dos histogramas dos *frames* classificados como câmera 1. e) Fase Classificadora.



3.2: Fluxograma da Fase Classificadora de *Frames* da Metodologia. a) Extração do histograma *HSV-162* do *frame* atual. b) Cálculo da divergência de Kullback-Leiber entre o histograma do *frame* atual e da média dos histogramas classificados como câmera 1. c) Traça a curva *ROC*, se todas as divergências foram calculadas. d) Calcula o melhor limiar da curva *ROC*

Figura 3. Fluxogramas da parte supervisionada e da parte classificadora do sistema.

Com o objetivo de estudar as características mais discriminantes entre as classes, os vídeos foram supervisionados e cada *frame* foi manualmente classificado como câmera 1 ou câmera 2. Este processo corresponde à parte direita da Figura 3.

Figura 4. Exemplo de alguns quadros da base. Os quadros 2, 3, 4, 6 e 8 são considerados câmera 1, e os demais não câmera 1.

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 isto em vista, propôs-se 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. Esta discretização gera um histograma de 162 posições possíveis. Para efeitos

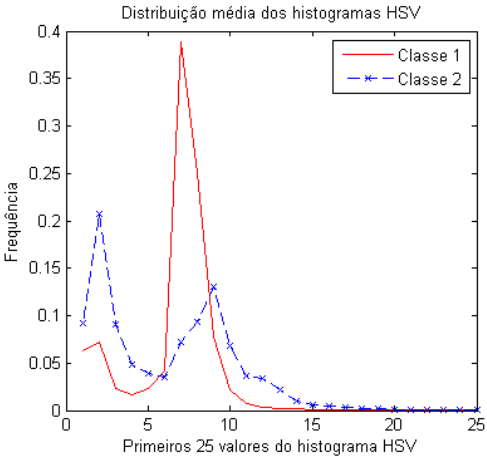


Figura 5. Histograma *HSV* médio para ambas as classes, c_1 e c_2

de comparação, foram calculados os histogramas médios das classes 1 e 2. Estes histogramas são calculados a partir da Equação 8.

$$Hist_m = \frac{1}{|C_1|} \sum_{i \in C1} Hist(i) \tag{8}$$

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 Figura 5. A extração dos histogramas *HSV-162* corresponde ao processo c da Figura 3.1.

Em seguida, o processo d esta relacionado ao cálculo da média dos histogramas *HSV-162* dos *frames* supervisionados como câmera 1. A Figura 5 mostra o histograma médio dos *frames* supervisionados como câmera 1. Por motivos de comparação, o histograma médio dos *frames* classificados como câmera 2 são exibidos também na Figura 5. Note que as médias das classes são diferentes, indicando que a cor pode ser uma característica discriminante. Desta forma, considerou-se a média da classe 1 como a distribuição padrão.

O processo e da Figura 3.1 equivale a Fase Classificadora representada pela Figura 3.2. Nesta fase de classificação dos *frames*, utilizou-se a divergência de Kullback-Leibler estendida para medir o quanto o histograma de cada *frame* diferencia do histograma padrão (processos a e b da Figura 3.2).

Um experimento foi realizado comparando, através da divergência de Kullback-Leibler estendida, todos os *frames* da classe 1 e da classe 2 com

a média da classe 1. Esta medida de distância servirá como base para classificar os *frames*. Se a distância for abaixo de um valor t, o *frame* será classificado como câmera 1, caso contrário como câmera 2. Uma das contribuições deste trabalho é um método automático para o cálculo deste limiar t, representados pelos processos c e d da Figura 3.2. Este método é explicado com maiores detalhes na Seção 5.

5. Resultados Experimentais

Para o modelo proposto neste 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 estendida) e é conhecido como parâmetro entrópico q.

É proposto o uso da curva ROC para medir a qualidade da classificação. Desta forma, variando os parâmetros t e q, é possível observar o desempenho do classificador. De acordo com Fawcett (2006), a curva ROC é uma técnica que relaciona a quantidade de falsos positivos e verdadeiros positivos. Esta 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, pode-se maximizar a área sob a curva ROC (Az).

Para este experimento, definiu-se 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; 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 da ordem de $O(h \times w \times n)$, onde h é a altura do *frame*, w é a largura e n é quantidade de *frames* analisados. A divergência de Kullback-Leibler é utilizada para fazer 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 $O(h \times w \times n)$.

Os resultados da classificação são apresentados Figura 6. Nesta distribuição, cada ponto representa a distância de Kullback-Leibler do histograma HSV do *frame* para o histograma HSV médio da classe 1. A linha tracejada separa a distribuição de duas formas: o lado esquerdo representa os *frames* que foram supervisionados como pertencentes à classe 1 e, do lado direito, os que pertencem à classe 2. Desta forma, nota-se que a mai-

oria 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 (t_{opt}). Os pontos que estão do lado esquerdo devem estar abaixo da linha horizontal contínua. E os pontos que estão do lado direito devem estar acima da linha horizontal. Quando isto não ocorre, há uma classificação como Falso Negativo e Falso Positivo, respectivamente. A Figura 7 representa a melhor curva ROC gerada nos experimentos. A Figura 8 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 Figura 7, o 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 desta curva, variando de 0 ao máximo da distribuição. Com isto, o valor de t_{opt} (que melhor classifica os dois grupos) é $t_{opt} = 0,44$. Na Figura 6, 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).

6. Discussão

Os resultados obtidos na Seção 5 mostram que o classificador teve um desempenho de 97% dos *frames* supervisionados com $q = 0,5$. Este resultado sugere que o sistema em estudo se comporta de maneira não-extensiva.

Porém, mesmo utilizando a teoria da informação não-extensiva, os resultados não foram totalmente corretos. Por este motivo, decidiu-se analisar quais *frames* tiveram as piores classificações. A Figura 9 mostra um *frame* supervisionado como câmera 1 e classificado como não câmera 1. Uma justificativa para este resultado pode ser dada pela concentração da cor verde no campo. Na Figura 9, nota-se que a câmera 1 está ampliada em uma determinada parte do campo. Isto pode ter feito com que a distribuição de verde diminuísse, uma vez que as faixas de grama estão ampliadas, causando uma diferença maior entre o histograma HSV do *frame* e o histograma HSV padrão.

Com relação ao pior falso positivo, a Figura 10 ilustra o *frame* em questão. Esta figura trata-se de uma transição entre cenas. As transições são criadas considerando informações tanto da cena anterior quanto da próxima cena. Isto significa que o histograma HSV de um *frame* de transição próximo a um *frame* de câmera 1 é semelhante ao histograma HSV médio da classe 1, justificando o resultado obtido.

Os erros de classificação obtidos são justificados pela própria composição histogrâmica dos *frames*. Os resultados mostraram que esta característica não é suficiente para classificar corretamente os *frames* como câmera 1. Isto significa que, para se alcançar 100% de acerto, deve-se eleger uma nova característica discriminante.

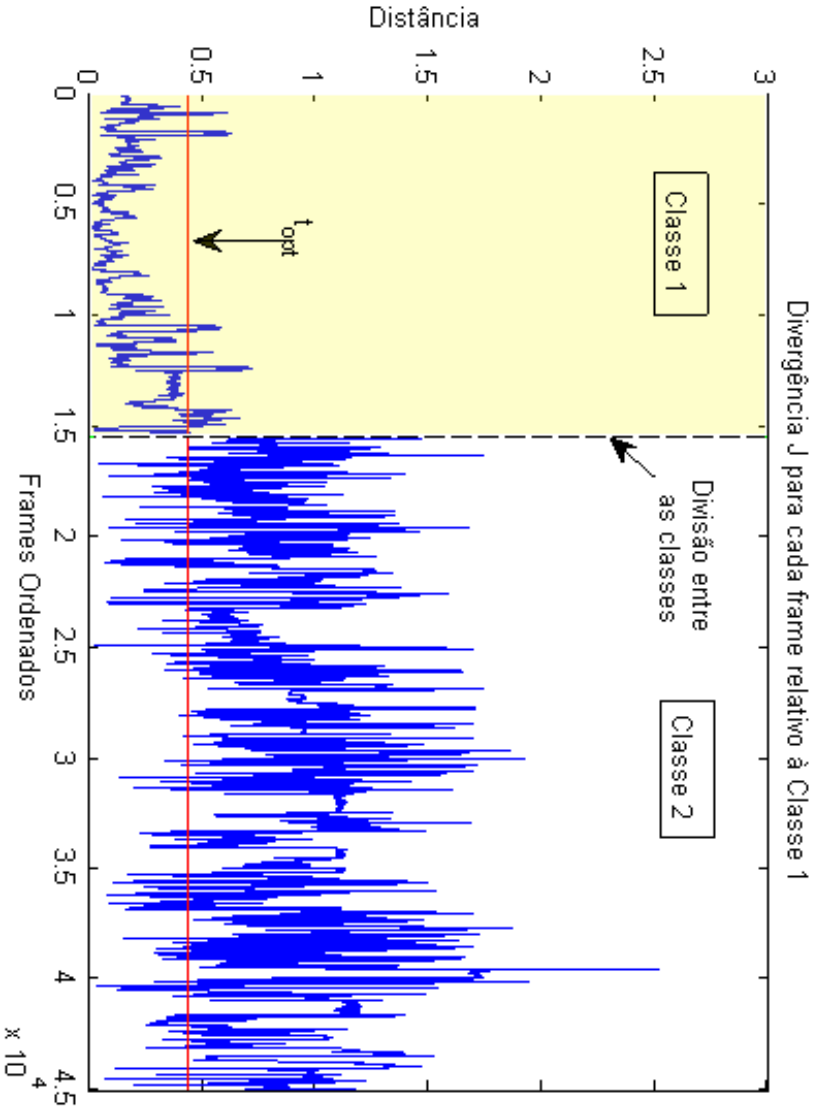


Figura 6. Distâncias entre os histogramas dos *frames* e o histograma médio da classe 1.

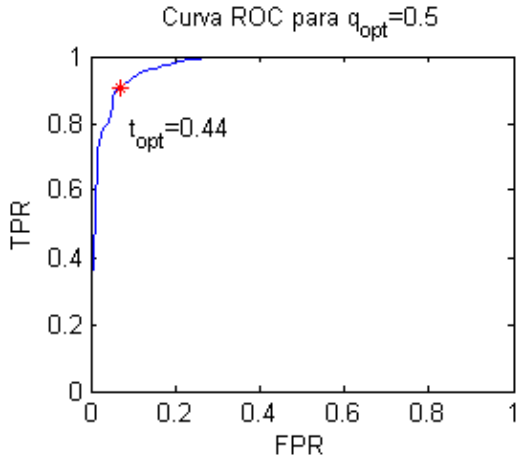


Figura 7. Curva ROC para q_{opt}.

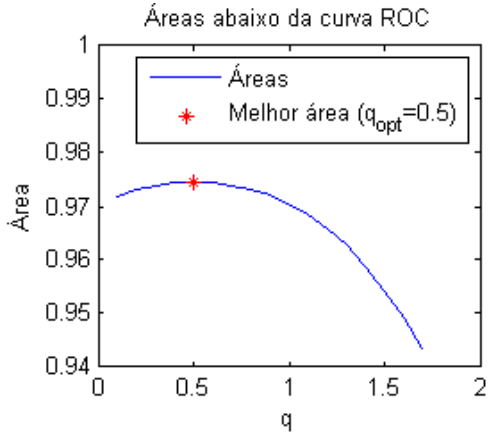


Figura 8. Áreas abaixo da curva ROC para diferentes valores de q.

Figura 9. Pior Falso Negativo.

Além disto, pode-se notar que a variação do valor do parâmetro entrópico q não alterou os resultados de forma significativa (menor que 4%),

Figura 10. Pior Falso Positivo.

como pode ser observado na Figura 8. Os mesmos testes foram efetuados para diferentes discretizações do histograma *HSV*. Os resultados ficaram próximos aos apresentados neste artigo.

7. Conclusão

Neste artigo propôs-se um método para classificação da tomada de câmeras em jogos de futebol através da divergência de Kullback-Leibler estendida. Foram analisados uma hora e meia de vídeo em cores num total de 45.000 *frames* de três vídeos diferentes. Os tipos de tomadas classificadas 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. Este resultado também reforça os dados apresentados em trabalhos prévios da literatura que não usam entropia não-extensiva (Halin et al., 2009), mas indicam também que a mantissa verde é a mais discriminante.

No uso da divergência de Kulback-Leibler estendida, 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 deste 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. Assim, pela literatura da entropia não-extensiva sugere-se que o sistema estudado aqui pode ser não-extensivo. Uma consequência imediata desta 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 sub-classes, 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 vôlei, 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 ao Centro Universitário da FEI (Fundação Educacional Inaciana), pelo suporte a este trabalho.

Referências

Bimbo, A.D., *Visual Information Retrival*. San Francisco, USA: Morgan Kaufmann, 1999.

Borland, L.; Plastino, A.R. & Tsallis, C., Information gain within nonextensive thermostatistics. *Journal Of Mathematical Physics*, 39(12):6490–6501, 1998.

Esqueff, I.A., *Técnicas de Entropia em Processamento Digital de Imagens*. Dissertação de mestrado em instrumentação científica, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, RJ, 2002.

Fawcett, T., An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874, 2006.

Halin, A.A.; Rajeswari, M. & Ramachandram, D., Shot view classification for playfield-based sports video. In: *Proceedings of IEEE International Conference on Signal and Image Processing Applications*. p. 410–414, 2009.

Hsu, W.; Chang, S.F.; Huang, C.W.; Kennedy, L.; Lin, C.Y. & Iyengar, G., Discovery and fusion of salient multi-modal features towards news story segmentation. In: Yeung, M.M.; Lienhart, R.W. & Li, C.S. (Eds.), *Proceedings of Storage and Retrieval Methods and Applications for Multimedia*. v. 5307, p. 244–258, 2003.

Jeffreys, H., *Theory Of Probability*. Oxford, UK: Oxford University Press, 1939.

Kuhmunch, C., On the detection and recognition of television commercials. In: *Proceedings of the 1997 International Conference on Multimedia Computing and Systems*. Piscataway, USA: IEEE Computer Society, p. 509, 1997.

Rodrigues, P.S. & Giraldi, G.A., Computing the q-index for Tsallis nonextensive image segmentation. In: *Proceedings of XXII Brazilian Conference on Computer Graphics and Image Processing*. Los Alamitos, USA: IEEE Computer Society, p. 232–237, 2009.

Shannon, C.E., A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423; 623–656, 1948.

Tsallis, C., Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52(1/2), 1988.

Tsallis, C., Nonextensive statistics: Theoretical, experimental and computational evidences and connections. *Brazilian Journal Of Physics*, 29(1):1–35, 1999.

Tsallis, C., Nonextensive statistical mechanincs and thermodynamics: historical background and present status. In: Abe, S. & Okamoto, Y. (Eds.), *Nonextensive Statistical Mechanics And Its Applications*. Berlin, Germany: Springer, v. 560 de *Lecture Notes In Physics*, p. 3–98, 2001.

Watve, A. & Sural, S., Soccer video processing for the detection of advertisement billboards. *Pattern Recognition Letters*, 29:994–1006, 2008.

Yeh, J.H.; Chen, J.C.; Kuo, J.H. & Wu, J.L., TV commercial detection in news program videos. In: *Proceedings of IEEE International Symposium on Circuts and Systems*. Piscataway, USA: IEEE Press, v. 5, p. 4594–4597, 2005.

Notas Biográficas

Guilherme Alberto Wachs Lopes é bacharel e mestre em Ciência da Computação (Centro Universitário da FEI, 2009 e 2011), atuando principalmente nos seguintes temas: redes complexas, reconhecimento de padrões, visão computacional, computação gráfica e simulação de fluidos. Atualmente, é doutorando no Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP).

Werner Fukuma é graduado em Ciência da Computação (Centro Universitário da FEI, 2009) e atualmente é mestrando no grupo de Inteligência Artificial Aplicada à Automação Industrial (Depto. Engenharia Elétrica no Centro Universitário da FEI). Tem como foco de seu trabalho a área de redes complexas e visão computacional.

Paulo Sergio Rodrigues é bacharel, mestre e doutor em Ciência a Computação (1996, 1999 e 2003, Universidade Federal de Minas Gerais), com estágio na Univertità Degli Studi di Ancona, Itália (1999). Durante os anos de 2003 a 2006 fez pós-doutorado no Laboratório Nacional de Computação Científica (LNCC). Há cerca de 15 anos, suas principais áreas de interesse têm sido visão computacional, processamento de imagens, realidade aumentada e reconhecimento de padrões. Tem a área médica como um dos principais alvos dos resultados de seus trabalhos. Em 2005-2006 publicou vários trabalhos na área de análise de imagens de câncer de mama e atualmente vem desenvolvendo técnicas para reconstrução de próteses craniofacial. Desde 2007 é professor do Departamento de Ciência da Computação do Centro Universitário da FEI e membro do Grupo de Inteligência Artificial do Departamento de Elétrica da mesma Instituição. É professor do mestrado em Engenharia Elétrica ministrando as disciplinas de visão computacional e geometria computacional.

Capítulo 12

EVEREVIS: Sistema de Navegação em Vídeos

Bruno do Nascimento Teixeira*, Júlia Epischina Engrácia de Oliveira, Tiago de Oliveira Cunha, Fillipe Dias Moreira de Souza, Lucas Gonçalves, Christiane Okamoto Mendoça, Vinícius de Oliveira Silva, Arnaldo de Albuquerque Araújo

Resumo: Este capítulo relata um sistema desenvolvido para a navegação de vídeo baseado na análise multimodal. A abordagem multimodal proposta realiza a transcrição de áudio para categorização de cenas (esportes, clima, política e economia) combinando informações de áudio e de vídeo. Suas principais características incluem resumos estáticos e dinâmicos, segmentação usando detecção de face, classificação em cenas internas e externas e transcrição de áudio para a busca de palavras-chave do tema. Palavras-chave são selecionadas para representar os vídeos. Uma série de experimentos foram conduzidos para avaliar a eficácia da categorização usando as informações de transcrição de áudio.

Palavras-chave: Processamento de imagens, Analise multimodal, Classificação e Sumarização.

Abstract: *This chapter reports a system developed for video browsing based on multimodal analysis. The proposed multimodal approach performs audio transcription for shot categorization (sports, weather, politics and economy) combining audio and visual information for theme categorization. Its main features include static and dynamic summaries, segmentation using face detection, classification into indoor and outdoor scenes based on Support Vector Machine (SVM) and audio transcription for theme keyword search. Keywords are selected to represent the subjects, followed by a simple text search. A set of experiments was conducted for evaluating the effectiveness of the shot subject categorization using audio transcription information.*

Keywords: *Image processing, Multimodal analysis, Classification and summarization.*

*Autor para contato: bruno.teixeira@dcc.ufmg.br

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 usuário limitem-se aos seus interesses e descartem o que não for relevante.

Este trabalho apresenta 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, estes algoritmos de processamento de vídeo visam auxiliar a navegação entre cenas internas e externas no vídeo, fornecer uma sumarização estática e dinâmica do vídeo, navegar no vídeo através da detecção de faces, e reconhecer eventos específicos nos vídeos. Além disto, esta plataforma executa uma abordagem multimodal, através do fornecimento da transcrição do áudio para a categorização de tomadas (esportes, tempo, política e economia), combinando informações visuais e auditivas para esta categorização.

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 (2008) 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. Esta 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. (2010) 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. (2010). 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, Rui et al. (1998) 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 Kim et al. (2006) 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* (Geisler et al., 2002), facilitando o acesso a trechos do vídeo.

Mais recentemente, Bouaziz et al. (2010) investigam a localização de texto nos vídeos para contribuir nesta área de indexação dos conteúdos dos vídeos. E mudando um pouco o contexto em que os sistemas de recuperação de vídeos vêm sendo desenvolvidos, Kim et al. (2011) propõem um sistema de navegação em vídeos musicais, de forma que os usuários busquem vídeos através de temas ou emoções, como por exemplo, a busca por um vídeo musical calmo de natureza.

As próximas seções apresentam, respectivamente, a metodologia utilizada para o desenvolvimento dos algoritmos integrados ao sistema interativo de análise do vídeo baseado em conteúdo, o resultado do sistema e considerações finais.

2. Sumarização de Vídeos

Um resumo de vídeo compreende a síntese do conteúdo do vídeo, preservando suas sequências mais importantes ou mais representativas, fornecendo uma versão concisa da mensagem do vídeo original. Resumos de vídeo podem ser classificados em duas categorias: estáticos (uma sequência de quadros-chave) ou dinâmicos (uma sequência de segmentos de vídeo). Em geral, este último é considerado mais atrativo por incorporar elementos de movimentação e áudio. Com um resumo em mãos, o usuário pode tomar decisões sobre a relevância do vídeo para tarefa desejada sem ter que vê-lo inteiramente, economizando tempo e esforço.

Neste trabalho, para a sumarização estática do vídeo foi utilizado o método proposto em de Avila et al. (2008). Este método foi concebido para ser simples e eficiente. Histogramas de cor e perfis de linha são utilizados para representar as imagens do vídeo. Após a extração das características visuais, as imagens são agrupadas pelo algoritmo *k-means*. Então,

um quadro-chave de cada grupo é identificado, e por fim as imagens são dispostas em ordem cronológica, produzindo o resumo estático do vídeo.

2.1 Sumarização dinâmica

Um método que tem como objetivo de tirar proveito de características de baixo nível e de alto nível foi desenvolvido para a sumarização dinâmica, o qual é baseado em características espaciais e espaço-temporais representadas por uma abordagem de histogramas de palavras visuais (*Bags of Visual Features*) (BoVF). BoVF é um tipo de representação inspirada no saco das palavras, uma abordagem comumente usada na recuperação de informação para representar coleções de texto (de Campos et al., 2011). No entanto, em vez de palavras, BoVF usa características visuais, onde cada característica visual é representa por uma região de interesse na imagem, gerada a partir de descritores extraídos dos segmentos. BoVF tenta reduzir a diferença semântica entre características de baixo nível e o conteúdo visual do vídeo, e tem sido utilizado na literatura em vários cenários de detecção de padrões e classificação, alcançando bons resultados devido a sua robustez a uma série de transformações na imagem e oclusão. No entanto, de acordo com a literatura pesquisada, esta técnica não foi empregada na área de sumarização automática de vídeo.

Além de representar os segmentos utilizando característica de alto nível geradas a partir de características de baixo nível, o método proposto também emprega uma estratégia inspirada em aprendizado multivisão (*multi-view learning*). O aprendizado multivisão utiliza diferentes representações (chamadas de visões) extraídas de um mesmo objeto (segmento) e aprende a partir deles independentemente (Muslea et al., 2002). Neste trabalho, são utilizadas três visões, representadas pelos descritores SIFT, Hue-SIFT e STIP. Cada descritor é utilizado para gerar um BoVF e o processo de aprendizado corresponde a encontrar os segmentos mais similares (representados por BoVFs gerados a partir dos descritores) utilizando um algoritmo de agrupamento para cada visão separadamente.

Após agrupar os segmentos, o mais representativo de cada grupo é extraído, e o resumo é gerado modelando a tarefa de sumarização com um problema de otimização. Isto é feito com o objetivo de que os resumos tenham uma duração máxima pré-definida. No entanto, esta restrição de tempo tem que ser obedecida enquanto preservam-se os segmentos de vídeo mais importantes. A importância de um segmento é medida de acordo com as suposições que dizem que quanto maior o movimento no segmento, maior é a informação contida (Pan et al., 2007; Laganière et al., 2008). Assim, a geração do resumo final é modelada como o problema da mochila, com a duração do resumo sendo equivalente ao peso da mochila e a quantidade de movimento ao benefício de um item. Finalmente, depois dos resumos de cada visão (descritor) terem sido criados, eles são unidos para formar um sumário único e final, usando a mesma abordagem descrita acima. A

Figura 1 apresenta o esquema geral do método proposto para sumarização dinâmica de vídeos.

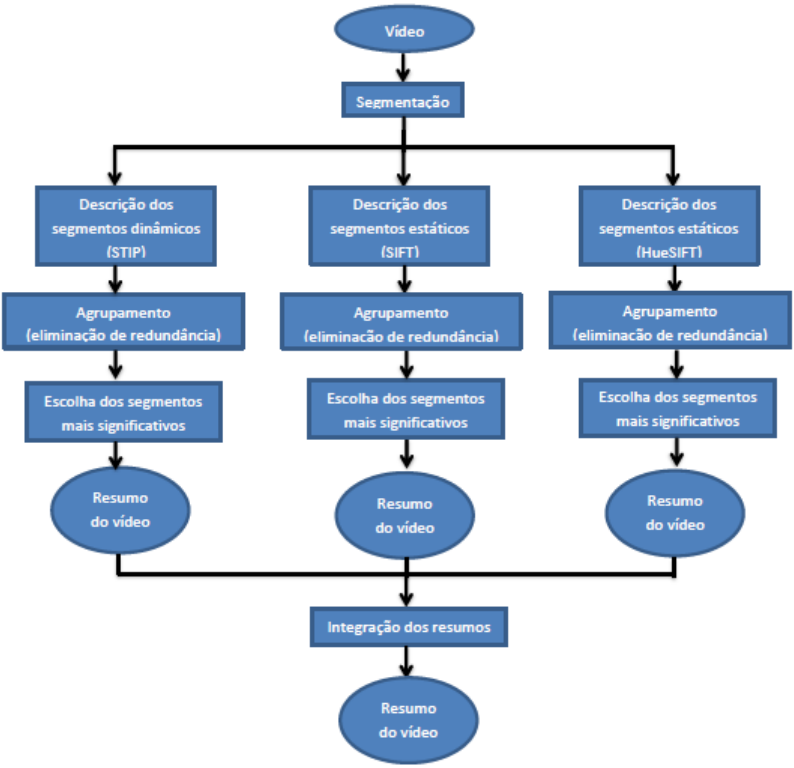


Figura 1. Arquitetura do método proposto.

Para a segmentação dos vídeos é utilizado o método proposto em Pan et al. (2007), por possuir baixo custo computacional e bons resultados em termos de detecção de tomadas, já que o método beneficia-se do uso de histogramas locais de cor e vetores de movimento.

2.1.1 Descrição e caracterização dos segmentos

No método proposto, as unidades básicas do vídeo são descritas por dois descritores espaciais e um descritor espaço-temporal. Estes descritores detectam e descrevem pontos de interesse, que são pontos específicos (ou regiões) em uma imagem ou vídeo que apresentam significante variação de intensidade em mais de uma direção. Além disto, são amplamente utilizados em visão computacional para tarefas de rastreamento, comparação e reconhecimento (Yan & Pollefeys, 2004; Laganière et al., 2008).

Para representar os segmentos previamente definidos, foi utilizada a abordagem de BoVF. BoVF (Csurka et al., 2004) é uma representação robusta para imagens, onde cada imagem é vista como um conjunto de regiões ou pontos e apenas as informações visuais são levadas em consideração sem ser necessária a informação sobre a localização do ponto na imagem. Estes pontos são chamados de palavras visuais.

O método BoVF é executado em algumas etapas. Primeiro, um método para detectar e descrever os pontos da imagem é aplicado. Os descritores extraídos da imagem precisam ser invariantes a mudanças que são irrelevantes para a tarefa de categorização (transformações de imagem, variações de iluminação e oclusões), mas ricos o suficiente para discriminar categorias. Os segmentos de vídeo são descritos usando os descritores STIP (Laptev, 2005), SIFT (Lowe, 2004) e Hue-SIFT (van de Sande et al., 2010a), que têm as características mencionadas anteriormente.

Em uma segunda etapa, um vocabulário é definido. Esta definição é baseada na escolha de um conjunto de pontos interessantes (palavras visuais), que é feita aleatoriamente a partir de todos os pontos disponíveis.

O terceiro passo é associar cada descritor para um visual da palavra no vocabulário. Esta associação é feita através do cálculo da distância Euclidiana entre os pontos da imagem e do vocabulário. O mais próximo visual da palavra no vocabulário a um ponto de imagem é armazenado de forma a gerar um histograma de ocorrência de palavras visual.

BoVF fornece uma representação mais informativa em termos de características de baixo nível, e é esperado que as características que compõem os histogramas sejam realmente padrões representativos capazes de descrever o conteúdo da imagem.

2.2 Eliminação de redundância

Para a remoção de redundância entre os segmentos do vídeo é aplicado um algoritmo de agrupamento. A idéia é que após o processo de agrupamento, os histogramas de palavras visuais que representam segmentos semelhantes façam parte dos mesmos grupos, e que de cada grupo seja escolhido somente um único segmento como representante.

Para a escolha do segmento mais representativo por grupo, é calculada a proximidade dos histogramas pertencentes a um grupo em relação ao seu centróide, e como resultado o histograma de maior proximidade é escolhido como o representante do grupo. Para o cálculo da proximidade foi utilizada a distância Euclidiana. Na Figura 2 é visto o processo de agrupamento e escolha do segmento mais representativo por grupo.

3. Seleção dos Segmentos mais Informativos

De maneira a gerar o sumário dinâmico, é necessário fazer uma seleção dos segmentos mais significativos. Parte-se da suposição de que quanto

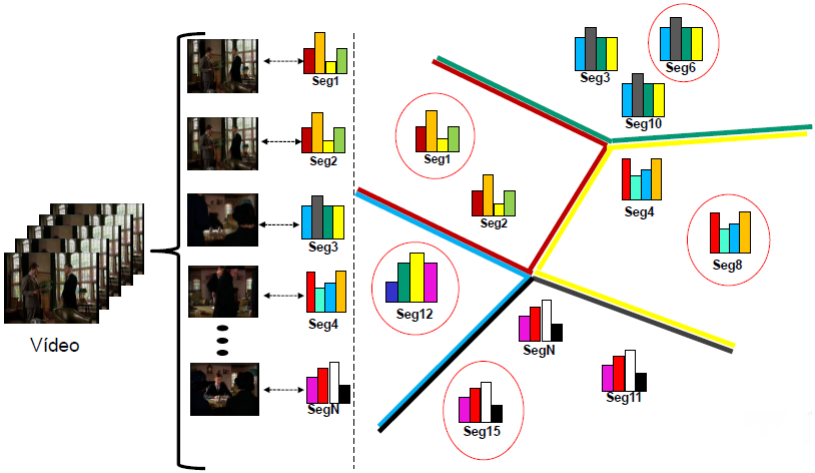


Figura 2. Processo de agrupamento.

maior o nível de atividade no segmento, maior é a quantidade de informação fornecida. Neste trabalho, assim como nos trabalhos de Pan et al. (2007), Laganière et al. (2008) e Putpuek et al. (2008), utiliza-se medidas de movimentação para o cálculo do nível de atividade, em que a média das magnitudes dos vetores de movimentação entre os quadros de um segmento é utilizada como medida de informação.

Para garantir que todos os resumos estejam de acordo com um tamanho máximo predefinido, o problema foi modelado como o amplamente conhecido problema da mochila binária (Cormen et al., 2001). Dados um conjunto de n objetos e uma mochila com:

- c_j = benefício do objeto j.
- w_j = peso do objeto j.
- b = capacidade da mochila.

Deseja-se determinar quais objetos devem ser colocados na mochila para maximizar o benefício total de tal forma que o peso da mochila não ultrapasse sua capacidade. Formalmente, o objetivo é:

$$\begin{cases} maximizar\ z = \sum_{j=1}^{n} c_j s_j \\ Sujeita\ a\ \sum_{j=1}^{n} w_j s_j \leq b \\ s_j \in \{0,1\} \end{cases}$$

Os resumos serão formados pela união dos segmentos que possuem o maior valor de movimentação de maneira que esta união ainda esteja de

acordo com o tamanho máximo preestabelecido para o resumo. Assim, foi definido que o número de quadros do resumo seria correspondente ao peso da mochila e o valor de movimentação correspondente ao benefício. Com a resolução do problema da mochila obtêm-se os segmentos que unidos obedecem a um limite de tempo predefinido, ao mesmo tempo em que possuem o valor máximo de movimentação. Para a resolução do problema da mochila foi utilizado o algoritmo baseado em programação dinâmica.

4. Classificação em Cenas Internas e Externas

Uma característica da aplicação é o reconhecimento automático de cenas interiores e exteriores em vídeos de notícias. O objetivo desta tarefa é a de facilitar a busca pelo conteúdo de interesse nas bases de dados de vídeo volumosos. Separar cenas internas de cenas de cenas externas pode ser visto como uma etapa de pré-processamento de um sistema de busca baseado em hierarquia. Para ilustrar, podemos assumir que é mais fácil encontrar clips de destaques dos jogos da Copa do Mundo de Futebol, se considerar apenas a pesquisa no conjunto de cenas externas do que quando utilizando o conjunto de dados inteiro. Ou seja, basta trabalhar com apenas com um subconjunto dos dados disponíveis.

Como cenas externas são separadas, um conjunto de imagens estáticas que representam estas cenas do vídeo de notícias é exibida em uma guia independente do sistema. Este guia se destina a fornecer aos usuários um meio alternativo para facilmente procurar o conteúdo de interesse. O usuário seria capaz de verificar visualmente se algum tema de interesse está presente neste vídeo particular.

4.1 Descritores

Cenas internas podem ser definidas como cenas em que o âncora lê a notícia dentro do estúdio, enquanto as cenas externas são aqueles onde as entrevistas e as apresentações são realizadas em um ambiente externo. Naturalmente, a cor parece ser uma característica bem a divergir interior de ambientes externos, embora a descrição forma ainda desempenhe o seu papel. Por esta razão, recomenda-se utilizar uma representação baseada em cor, forma de caracterizar as cenas de vídeo. van de Sande et al. (2010b) propuseram descritores baseados em cores, que têmum desempenho muito bom para categorização de objetos e cenas, sendo um deles o HueSIFT. HueSIFT é uma combinação do detector de características locais SIFT (para imagens estáticas) com histogramas matiz. Neste caso, a vizinhança de cada característica local é descrito em termos de matiz (a partir do espaço de cores HSI) e seu histograma é concatenado com histogramas de gradientes orientados (descrição da forma).

4.2 Representação *bag-of-visual-words*

Dado um clip de vídeo, uma representação de alto nível para o conjunto de características locais podem ser fornecidos pela abordagem *bag-of-feature*. Nesta abordagem, *visual codebook* é construído usando um algoritmo de agrupamento como o *k-means* (MacQueen, 1967), sobre uma amostra do conjunto de dados de treinamento. Características locais de uma imagem são atribuídos à palavra mais próxima visual (pode-se usar como métrica de distância função de distância Euclidiana). Como resultado, um histograma de palavras visual é construído para representar o conjunto de características locais descrevendo o conteúdo da imagem.

Foi usado aprendizado supervisionado com Support Vector Machines (SVM) Vapnik (1995) para classificar os vídeos. Uma vez que é um problema binário, é recomendado usar *kernels* lineares, uma vez que demonstraram bons resultados em problemas envolvendo apenas 2 classes como por exemplo cenas internas e externas.

Na validação do método de classificação, pode-se testar o desempenho através de uma abordagem de validação cruzada. Neste esquema, uma amostra de quadros é extraída de cada cena do vídeo de ambos os tipos. O conjunto é uniformemente dividido em cinco grupos de tal forma que cada grupo é ponderado em termos da classe considerada. Desta forma, cada conjunto é testado com a união dos grupos restantes, isto é, se *fold0* é o conjunto de teste, então o conjunto de treinamento é a concatenação de *fold1*, *fold2*, *fold3* e *fold4*. Todos os quadros são rotulados e correspondem a entrada para treinar os classificadores. Para cada quadro dos conjuntos é atribuído um rótulo (interna ou externa) e um esquema de votação por maioria é aplicado, determinando a classe da cena. A Tabela 1 mostra a avaliação de desempenho em termos de *shots*.

Tabela 1. Performance da classificação. As colunas *#inshot* e *#outshot* correspondem ao número de cenas *indoor* e *outdoor*, respectivamente. As colunas *Inclass* e *outclass* correspondem a classificação dos conjuntos de *shots indoor* e *outdoor*, respectivamente.

fold	***#inshot***	***#outshot***	***inclass***	***outclass***
fold0	46	40	86,96%	92,50%
fold1	46	40	97,83%	92,50%
fold2	46	40	89,13%	87,50%
fold3	46	40	91,30%	90,00%
fold4	45	40	73,33%	90,00%

A perfomace geral é mostrada na tabela de matriz de confusão (Tabela 2), o que demonstra que o método empregado é poderoso para o contexto de aplicação pretendida.

Tabela 2. Matriz de confusão.

	outdoor	***indoor***
outdoor	**90,00%**	10,00%
indoor	11,11%	**88,89%**

5. 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 frequentemente na posição frontal.

O algoritmo deste trabalho utiliza o método de Viola & Jones (2001), que é implementado na biblioteca *OpenCV* (*Open Source Computer Vision*[1]) (Bradsky & Kaehler, 2008). Este 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* (Freund & Schapire, 1995) 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. O método 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.

6. Transcrição de Áudio

Para categorizar as cenas, a transcrição de áudio gera uma descrição de cada cena segmentada que representa os assuntos apresentados em vídeo (fala). Em vídeos de notícias, é muito comum que informações de áudio descreve a informação visual. Esta observação sugere para converter áudio em informações de texto usando Julius (*Open-Source Large Vocabulary CSR Engine Julius*) (Kawahara et al., 2000), motor de reconhecimento de voz para transcrição de áudio que usa modelos de gramática e acústico. Um conjunto de cinco palavras-chave é definido para representar quatro temas: esportes, política, economia e tempo. A taxa de segmentos classificados é de aproximadamente 5%. Com um vídeo com 300 segmentos usando características visuais, apenas 15 serão classificados, e 9 segmentos foram classificados corretamente. O tempo de processamento é de aproximadamente 5 horas para todos os algoritmos. A Figura 3 mostra o princípio de funcionamento do reconhecimento de fala baseado na interface entre modelos acústicos e de linguagem. Modelos ocultos de Markov (*Hidden*

[1] Disponível em: *http://opencv.org/*

Markov Model – HMM), base dos modelos acústicos, podem ser usados em aplicações de reconhecimento de padrões.

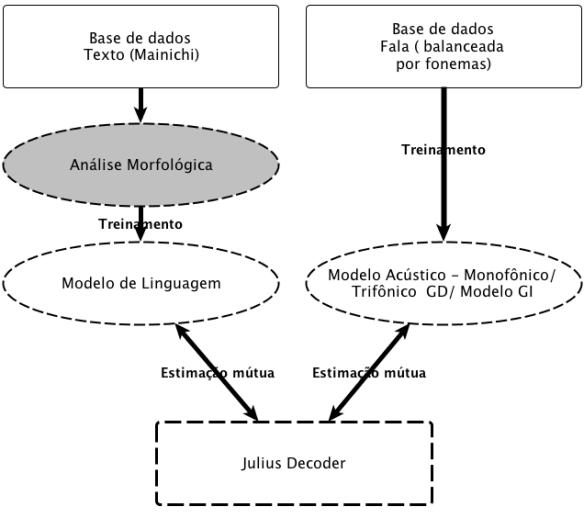


Figura 3. Plataforma LVCSR.

O reconhecimento de fala realiza uma pesquisa de dois passos (para frente e para trás) usando modelos 2-gram e 3-gram. Modelos n-gram são modelos probabilísticos para a previsão do próximo termo em uma sequência. Alguns modelos de linguagem construído a partir de n-grams são modelos de Markov de ordem $(n-1)$. Um n-gram é a sequência contínua de n termos para uma da sequência de texto ou fala. A primeira passagem gera uma "índice de palavras" que consiste em um conjunto nós de palavras-finais por quadro, com suas pontuações. Ele será usado eficientemente para procurar palavras candidatas no segundo passo. A segunda passagem realiza outra pontuação Viterbi permitindo superar a perda de precisão por aproximações. No segundo passo, modelo de linguagem e dependência de contexto é aplicado para re-pontuação. A busca é realizada no sentido inverso, e precisa ser dependente da sentença.

Na fala espontânea, como palestras e reuniões, o segmento de entrada é incerto e longo. Decodificações sucessivas são realizadas usando pausa de curta-segmentação, que consiste em automaticamente dividir a entrada e usar o reconhecimento por pequenas pausas. Quando uma pequena pausa é detectada, ela finaliza a pesquisa atual neste ponto e, em seguida, reinicia o reconhecimento.

7. Interface

A interface do sistema *web* de vídeos consiste em três áreas principais: navegação, exibição e montagem (Figuras 6, 7 e 8). 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. Estas imagens são clicáveis para exibição no *player* e arrastáveis até a área de montagem.

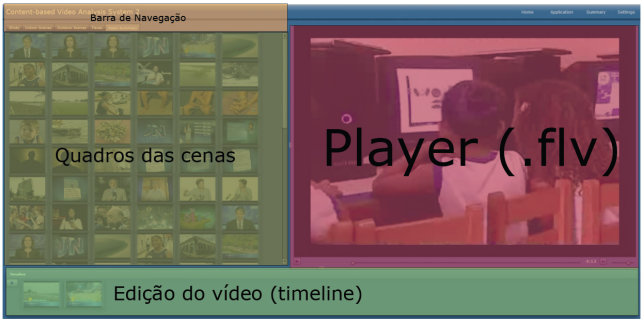


Figura 4. Interface *web* com as ferramentas para navegação e visualização de vídeos de noticiários. A barra de navegação apresenta os algoritmos usados na segmentação e caracterização das cenas do vídeo.

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 sequencialmente e sem intervalos. Também, os segmentos podem ser reordenados facilmente na linha temporal, formando uma nova edição para o vídeo. A tela *home* do sistema apresenta o resumo dinâmico dos vídeos, onde o usuário escolhe o vídeo desejado (Figura 5). Além destas 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.

8. Considerações Finais

Foi apresentada neste capítulo uma interface visual para melhorar a efetividade do acesso à informação em vídeos. O sistema de navegação na interface é composto por sumarização estática e dinâmica, detecção de fa-

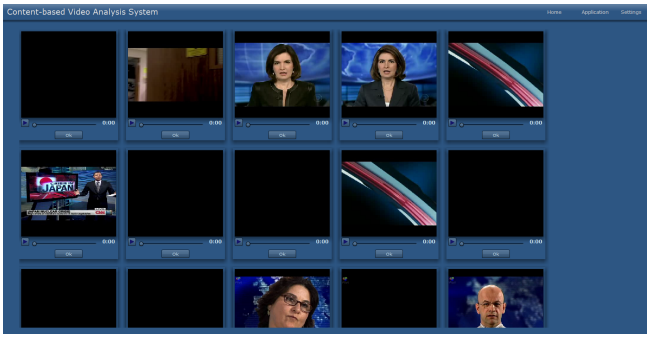
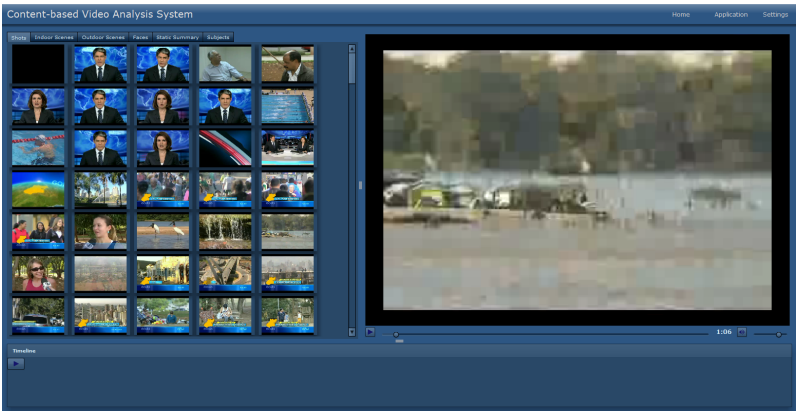


Figura 5. Tela *home* do sistema de navegação em vídeos com os vídeos e os resumos dinâmicos.

ces, classificação em cenas internas e externas, e transcrição de áudio para a classificação das cenas em temas.

A sumarização estática utilizou um método simples que forneceu as imagens mais representativas do vídeo, ao passo que a sumarização dinâmica apresentou um vídeo curto como resumo. Ambos sumários forneceram rapidamente a informação concisa do vídeo, permitindo ao usuário ter uma idéia do conteúdo dos vídeos disponíveis na plataforma sem a necessidade de assisti-los inteiramente. Outra forma disponibilizada para a navegação nos vídeos de notícias foi através das faces presentes. A detecção de faces é um dos passos do modelo computacional para o reconhecimento da face, cujo objetivo é identificar ou verificar pessoas no vídeo e é um dos trabalhos futuros a ser realizado. A classificação em cenas internas e externas baseada em aprendizagem supervisionada com SVM demonstrou ser uma ferramenta poderosa para o contexto desta aplicação, ou seja, vídeos de notícias. O algoritmo utilizou HueSIFT como detector das características locais, principalmente devido ao padrão azul observado nas cenas internas. Um trabalho futuro inclui expandir a caracterização das cenas através de outros atributos, testando também outros *kernels* do SVM. Finalmente, a transcrição de áudio forneceu uma taxa baixa de classificação devido ao uso de poucas palavras-chave para a descrição do tema e à segmentação baseada em características visuais. Esta taxa pode ser aumentada usando a segmentação semântica de texto e áudio.

Além do aperfeiçoamento dos algoritmos já existentes, algumas direções de pesquisa são a inclusão de outros tipos de vídeos e a combinação de áudio e recursos visuais utilizando grafos para a busca vídeo-vídeo.



(a)

(b)

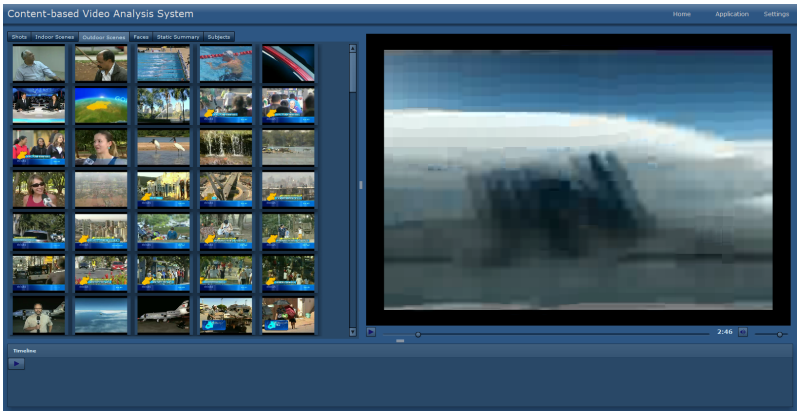
Figura 6. (a) Navegação por cenas e classificação das cenas em internas (b) usando o método proposto com *kernel* linear.

Agradecimentos

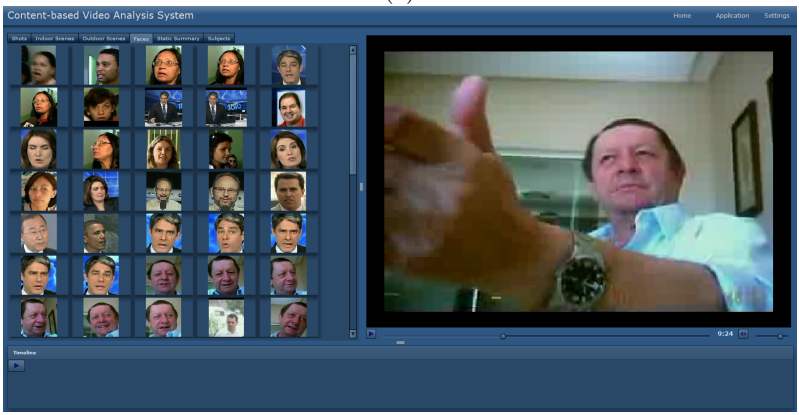
Os autores agradecem ao CNPq, CAPES, FAPEMIG pelo suporte financeiro do projeto.

Referências

de Avila, S.E.F.; da Luz, A.; de Albuquerque Araújo, A. & Cord, M., VSUMM: An Approach for Automatic Video Summarization and



(a)



(b)

Figura 7. (a) Classificação das cenas em internas e indexação de cenas usando detecção de face (b) usando o método de Viola-Jones.

Quantitative Evaluation. In: *Proceedings of XXI Brazilian Symposium on Computer Graphics and Image Processing*. p. 103–110, 2008.

Bouaziz, B.; Mahdi, W.; Zlitni, T. & ben Hamadou, A., Content-based video browsing by text region localization and classification. *International Journal of Video & Image Processing and Network Security*, 10(1):40–50, 2010.

Bradsky, G.R. & Kaehler, A., *Learning OpenCV: Computer Vision with the OpenCV Library*. 1a edição. Sebastopol, USA: O'Reilly Media, 2008.

(a)

(b)

Figura 8. (a) Sumário estático e (b) classificação das cenas temas usando transcrição e busca texto-texto simples.

Cormen, T.H.; Leiserson, C.E.; Rivest, R.L. & Stein, C., *Introduction to Algorithms*. Cambridge, USA: The MIT Press, 2001.

Csurka, G.; Dance, C.R.; Fan, L.; Willamowski, J. & Bray, C., Visual categorization with bags of keypoints. In: *Proceedings of Workshop on Statistical Learning in Computer Vision*. p. 1–22, 2004.

de Campos, T.; Barnard, M.; Mikolajczyk, K.; Kittler, J.; Yan, F.; Christmas, W. & Windridge, D., An evaluation of bags-of-words and spatio-temporal shapes for action recognition. In: *Proceedings of IEEE*

Workshop on Applications of Computer Vision. Washington, USA: IEEE Computer Society, p. 344–351, 2011.

Forlines, C., Content aware video presentation on high-resolution displays. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. New York, USA: ACM Press, p. 57–64, 2008.

Freund, Y. & Schapire, R.E., A decision-theoretic generalization of online learning and an application to boosting. In: *Proceedings of the Second European Conference on Computational Learning Theory*. London, UK: Springer-Verlag, p. 23–37, 1995.

Geisler, G.; Marchionini, G.; Wildemuth, B.M.; Hughes, A.; Yang, M.; Wilkens, T. & Spinks, R., Video browsing interfaces for the Open Video Project. In: *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*. p. 514–515, 2002.

Kawahara, T.; Lee, A.; Kobayashi, T.; Takeda, K.; Minematsu, N.; Sagayama, S.; Itou, K.; Ito, A.; Yamamoto, M.; Yamada, A.; Utsuro, T. & Shikano, K., Free software toolkit for japanese large vocabulary continuous speech recognition. In: *Proceedings of Sixth Annual Conference Spoken Language Processing*. ISCA, v. 4, p. 476–479, 2000.

Kim, H.G.; Kim, J.Y. & Baek, J.G., An integrated music video browsing system for personalized television. *Expert Systems with Applications*, 38:776–784, 2011.

Kim, J.; Kim, H. & Park, K., Towards optimal navigation through video content on interactive tv. *Interactive Computing*, 18:723–746, 2006.

Laganière, R.; Bacco, R.; Hocevar, A.; Lambert, P.; Païs, G. & Ionescu, B.E., Video summarization from spatio-temporal features. In: *Proceedings of the 2nd ACM TRECVid Video Summarization Workshop*. New York, USA: ACM Press, p. 144–148, 2008.

Laptev, I., On space-time interest points. *International Journal of Computer Vision*, 64:107–123, 2005.

Lowe, D.G., Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

MacQueen, J.B., Some methods for classification and analysis of multivariate observations. In: Le Cam and J. Neyman, L.M. (Ed.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, v. 1, p. 281–297, 1967.

Morand, C.; Benois-Pineau, J.; Domenger, J.P.; Zepeda, J.; Kijak, E. & Guillemot, C., Scalable object-based video retrieval in hd video databases. *Signal Processing: Image Communication*, 25(6):450–465, 2010.

Muslea, I.; Minton, S. & Knoblock, C.A., Active + semi-supervised learning = robust multi-view learning. In: *Proceedings of the Nineteenth International Conference on Machine Learning*. San Francisco, USA: Morgan Kaufmann, p. 435–442, 2002.

Pan, C.M.; Chuang, Y.Y. & Hsu, W.H., NTU TRECVID-2007 fast rushes summarization system. In: *Proceedings of the International Workshop on TRECVID Video Summarization*. New York, USA: ACM Press, p. 74–78, 2007.

Putpuek, N.; Le, D.D.; Cooharojananone, N.; Satoh, S. & Lursinsap, C., Rushes summarization using different redundancy elimination approaches. In: *Proceedings of the 2nd ACM TRECVid Video Summarization Workshop*. New York, USA: ACM, p. 100–104, 2008.

Rui, Y.; Huang, T.S. & Mehrotra, S., Constructing table-of-content for videos. *Multimedia Systems*, 7(5):359–368, 1998.

van de Sande, K.E.A.; Gevers, T. & Snoek, C.G.M., Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010a.

van de Sande, K.E.A.; Gevers, T. & Snoek, C.G.M., Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582 –1596, 2010b.

Su, J.H.; Huang, Y.T.; Yeh, H.H. & Tseng, V.S., Effective content-based video retrieval using pattern-indexing and matching techniques. *Expert Systems with Applications*, 37(7):5068–5085, 2010.

Vapnik, V.N., *The nature of statistical learning theory*. New York, USA: Springer-Verlag, 1995.

Viola, P. & Jones, M., Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE Press, p. 511–518, 2001.

Yan, J. & Pollefeys, M., Video synchronization via space-time interest point distribution. In: *Proceedings of Advanced Concepts for Intelligent Vision Systems*. p. 1–5, 2004.

Notas Biográficas

Bruno do Nascimento Teixeira é graduado em Engenharia de Controle e Automação (Universidade Federal de Minas Gerais – UFMG, 2006) com estágio em Technion - Israel Institute of Technology, mestre em Engenharia Elétrica (UFMG, 2009) com estágio em University of British Columbia (UBC). Atualmente é doutorando em Ciências da Computação (UFMG, 2010).

Júlia Epischina Engrácia de Oliveira é graduada emEngenharia Elétrica (Universidade Presbiteriana Mackenzie, 2001), mestre em Física aplicada à Medicina e Biologia (Universidade de São Paulo, 2005) e doutor em Bioinformática pela (UFMG, 2009). Fez um doutorado-sanduíche em Aachen, Alemanha na Universidade RWTH (2007). Concluiu um pós-doutorado em Ciência da Computação (UFMG, 2011). Atualmente, trabalha em um projeto de pesquisa em imagens mamográficas no Centro de Desenvolvimento da Tecnologia Nuclear (CDTN).

Fillipe Dias Moreira de Souza é bacharel e mestre em Ciência da Computação (Universidade Estadual de Santa Cruz, 2009 e UFMG, 2011, respectivamente). Participou do programa de formação de Analistas de Teste de Software promovido pela Universidade Federal de Pernambuco em parceria com a Motorola Ltda (2008). Participou do programa de treinamento em projeto de circuitos integrados digitais, CI-Brasil, promovido pela Universidade Federal do Rio Grande do Sul em parceria com o governo brasileiro e Cadence Ltda, obtendo o título de projetista de circuitos integrados digitais (2009). Atualmente é doutorando em Ciência da Computação (University of South Florida, EUA).

Tiago de Oliveira Cunha é graduado e mestre em Ciência da Computação (pela Universidade Estadual de Santa Cruz, 2008 e UFMG, 2011, respectivamente). Possui experiência na área de Ciência da Computação, com ênfase em processamento digital de imagens e algoritmos evolucionários, atuando principalmente nos seguintes temas: compressão de imagens digitais e sumarização automática de vídeos.

Lucas Figueiredo Gonçalves é bacharel em Ciência da Computação (Pontifícia Universidade Católica de Minas Gerais, 2011).

Christiane Okamoto Mendoça é graduanda em Ciência da Computação (UFMG).

Vinícius de Oliveira Silva é graduando em Engenharia Elétrica (UFMG).

Arnaldo de Albuquerque Araújo é graduado, mestre e doutor em Engenharia Elétrica (Universidade Federal da Paraíba – UFPB – em Campina Grande, 1978, 1981, 1987, respectivamente). Tem pós-graduação em Processamento de Imagens (Rheinisch-Westfaelische Technische Hochschule Aachen, 1981-1985) e pós-doutorado em Processamento de Imagens (ESIEE Paris, 1994-1995; ENSEA Cergy-Pontoise, 2005; e UPMC Paris 6, 2008-2009). Foi professor adjunto da UFPB, departamento de Engenharia Elétrica (1978-1989) e atualmente é professor associado da UFMG, Departamento de Ciência da Computação (desde 1990).

Capítulo 13

SLPTEO e SCORC: Abordagens para Segmentação de Linhas, Palavras e Caracteres em Textos Impressos

Josimeire do Amaral Tavares,* Igor Santos Peretta,
Gerson Flávio Mendes de Lima, Keiji Yamanaka e Mônica Sakuray Pais

Resumo: Sistemas de reconhecimento óptico de caracteres possibilitam várias aplicações, dentre elas, o reconhecimento automático de caracteres em textos impressos. Para o sucesso de tais sistemas, é essencial uma etapa de segmentação confiável. Este capítulo apresenta dois métodos de segmentação: o SLPTEO para segmentação de linhas de texto e palavras, e o SCORC para segmentação de caracteres. O primeiro é aplicado a textos impressos, mas pode ser aplicado a textos manuscritos. O segundo resolve problemas de sobreposição de caracteres impressos e conexões entre os mesmos, trabalhando diretamente em imagens em níveis de cinza. Os resultados experimentais indicam grande robustez dos métodos apresentados.

Palavras-chave: Segmentação de linhas de texto, Segmentação de palavras, Segmentação de caracteres, Textos impressos.

Abstract: *Optical Character Recognition systems enable several applications, e.g. automatic character recognition in printed texts. For the success of such systems, reliable segmentation is an essential stage. This chapter presents two approaches to segmentation: the SLPTEO for segmentation of text lines and words, and SCORC for character segmentation. The first is applied to printed texts, but can be also applied to handwritten texts. The second handles printed overlapping and touching characters, working directly on grayscale images. Experimental results show great robustness of the methods presented.*

Keywords: *Text line segmentation, Word segmentation, Character segmentation, Printed texts.*

*Autor para contato: josimeiretavares@gmail.com

1. Introdução

O processamento digital de imagem em conjunto com a evolução da computação tem beneficiado várias áreas da ciência. Uma das aplicações de processamento digital de imagens, chamada automação de tarefas, visa atribuir ao computador a capacidade necessária para que ele desempenhe papéis e tarefas que são executadas com facilidades pelos seres humanos. Esta demanda na automatização de tarefas tem favorecido o desenvolvimento de sistemas de reconhecimento de padrões. Dentre eles, existem os chamados sistemas de Reconhecimento Óptico de Caracteres (*Optical Character Recognition* – OCR). Um sistema de OCR possibilita vários tipos de aplicações: identificação de assinaturas (Pavlidis et al., 1998); reconhecimento de textos manuscritos e impressos (Marti & Bunke, 2002); reconhecimento de placas de veículos (Conci et al., 2009); reconhecimento de textos em Braille (Bezerra, 2003); dentre outros.

Para o sucesso de sistemas OCR é necessária uma etapa primordial: a etapa de segmentação. Segundo Gonzalez & Woods (2010) a etapa de segmentação subdivide uma imagem em regiões ou objetos que a compõem. O nível de detalhes em que a subdivisão é realizada depende do problema a ser resolvido. Por exemplo, para a segmentação de documentos impressos, a subdivisão da imagem normalmente é feita até que se extraia os caracteres da imagem. Para textos manuscritos, Silva (2009) considera que o nível de segmentação de palavras já é suficiente.

Neste capítulo são apresentadas abordagens de segmentação de textos impressos através de dois métodos: "Segmentação de Linhas e Palavras baseado no Operador de Energia de Teager" (SLPTEO) e "Segmentação de Caracteres Orientados a Regiões em níveis de Cinza" (SCORC).

O método SLPTEO é aplicado a imagens de texto para segmentação de linhas e palavras. Este método possui como diferencial o fato de que o mesmo método pode ser aplicado a textos impressos e manuscritos sem nenhum ajuste prévio para adequação ao tipo de texto. Além disto, o mesmo algoritmo é utilizado para a segmentação de linhas e para a segmentação de palavras, com a distinção de um único parâmetro.

O método SCORC é aplicado a imagens de palavras de texto impresso, visando a subdivisão da mesma em caracteres. Este método é baseado em segmentação de regiões, aplicando-se uma abordagem semelhante à rotulação de componentes conectados. Este método requer que a imagem da palavra seja capturada em níveis de cinza, ou que seja convertida da original colorida.

Muitos trabalhos têm abordado a segmentação de caracteres conectados e sobrepostos em dígitos, mas poucos em alfabetos (Saba et al., 2010). Alguns trabalhos tem sido desenvolvidos nos últimos anos com o objetivo de leitura automática por máquinas. Para isto, várias estratégias tem sido investigadas objetivando-se a solução dos caracteres conectados, levando

em conta que existe uma grande variação entre os vários tipos de caracteres em diversos idiomas (Saba et al., 2010).

No trabalho de Lue e colaboradores (Lue et al., 2010), os autores apresentam a segmentação de caracteres capturados por câmeras digitais de aparelhos móveis. Neste trabalho, a extração da linha de texto é realizada primeiramente, identificando os componentes conectados da imagem capturada. Os autores utilizam como base deste trabalho as técnicas de projeção horizontal e vertical, criando um vetor de características extraídas destas projeções. Em seguida, uma rede neural *support vector machine* (SVM) realiza a classificação dos caracteres.

No trabalho de Nikolaou et al. (2010), os autores efetuam todas as etapas de segmentação de linhas, palavras e caracteres em textos históricos impressos. Para a etapa de segmentação de caracteres é utilizado um algoritmo de esqueletização e, a partir desta etapa, é realizada a segmentação isolando-se os possíveis caracteres conectados.

O trabalho apresentado por Jung (2010), o autor utiliza a projeção horizontal, vertical e lateral dos caracteres para extrair as características dos mesmos. O autor considera que a conexão entre caracteres não interfere na visão lateral. Assim, ele agrupa os caracteres em 13 classes distintas, de acordo, com as características apresentadas. Em seguida, realiza os cálculos para verificar o melhor ponto de corte para a devida segmentação.

O restante deste capítulo está estruturado como se segue: a Seção 2 descreve os conceitos teóricos necessários para o entendimento dos métodos propostos; a Seção 3 descreve ambos os métodos propostos (SLPTEO e SCORC), com a apresentação de exemplos práticos; a Seção 4 descreve os resultados alcançados para a correta segmentação de linhas, palavras e caracteres quando os métodos são aplicados a uma base de dados com textos impressos; finalmente, a Seção 5 traz conclusões sobre o trabalho, assim como uma breve discussão sobre outros trabalhos a serem desenvolvidos.

2. Fundamentação Teórica

Nesta seção estão descritos alguns conceitos importantes para o entendimento dos métodos a serem apresentados.

2.1 Projeção linear

A projeção horizontal de uma imagem é definida como a contagem dos pixels de interesse existentes em cada linha do objeto, como mostrada na equação (1) (Pedrini & Schwartz, 2008).

$$P_h(y) = \sum_{x=0}^{N-1} f(x, y) \tag{1}$$

A técnica de projeção linear (horizontal ou vertical) é muito utilizada em segmentação de linhas e palavras, uma vez que através do histograma gerado é possível identificar "vales" e "picos" de intensidade. Estes picos são concentrações de pixels e são facilmente identificadores de objetos dentro de uma imagem.

Em se tratando de imagens de textos, quando se utiliza a projeção horizontal, pode-se identificar que os picos apresentados pelo histograma são as linhas de textos e os vales são os espaços entre linhas. A projeção é um método simples e eficiente quando se trata de textos bem comportados (textos impressos, por exemplo).

Analogamente, a Equação (2) mostra a projeção vertical P_v que é definida como a soma dos pixels de interesse em cada coluna da imagem.

$$P_v(x) = \sum_{y=0}^{M-1} f(x, y) \tag{2}$$

No entanto, a segmentação de palavras é um pouco mais complexa. Os espaçamentos existentes são diferenciados: entre palavra ou entre caracteres de uma mesma palavra. Normalmente se usa o histograma vertical em conjunto com alguma estatística que leva em conta as medidas de distância de espaçamento para segmentar as palavras de um texto.

2.2 Operador de energia de Teager

Teager & Teager (1990), em um trabalho sobre modelamento não-linear da voz, apresentaram um operador de energia que mais tarde foi denominado como operador de energia de Teager (*Teager Energy Operator* – TEO), também conhecido como operador Teager-Kaiser. Este operador, de acordo com Kaiser (1990), 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 este sinal.

Kaiser (1993) também definiu os operadores TEO em ambos os espaços contínuo e discreto como "ferramentas muito úteis para analisar sinais com um único componente de um ponto de vista da energia" [tradução dos autores].

O operador de energia de Teager é definido no domínio discreto (Kaiser, 1993) como:

$$\Psi[x(n)] = x_n^2 - x_{n+1} \cdot x_{n-1}, \tag{3}$$

onde Ψ é o operador de Teager e x_n é o valor da n-ésima amostra do sinal.

Note que, no domínio discreto, este algoritmo utiliza apenas três operações aritméticas aplicadas a três amostras adjacentes do sinal para cada deslocamento no tempo.

Uma importante característica do TEO analisada por Kaiser (1990) é que, quando aplicado a sinais compostos por dois ou mais componentes de

frequência, "seria como se o algoritmo [TEO] fosse capaz de extrair a função envelope do sinal" [tradução dos autores]. Como exemplo, a dissertação de Peretta (2010) utiliza esta característica do TEO para detectar as fronteiras de um comando de voz inserido em um sinal de áudio. Outro trabalho dos autores utiliza esta mesma premissa (Peretta et al., 2010).

Considerando os valores de projeção (histograma) de uma imagem de texto binarizada como sendo um sinal amostrado, pela teoria de séries de Fourier (Dirichlet, 1829), este sinal (função arbitrária dentro de dado intervalo) pode ser expresso por uma série de senos e cossenos. Ou seja, o resultado desta abstração é um sinal composto por diversos componentes de frequência. Ao aplicar o TEO a este sinal abstrato, a "função envelope" seria extraída.

A Figura 1 mostra a imagem do documento "f04-020" onde se pode observar a projeção horizontal e o envelope do TEO. O método SLPTEO é aplicado aos valores de projeção horizontal desta imagem 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.

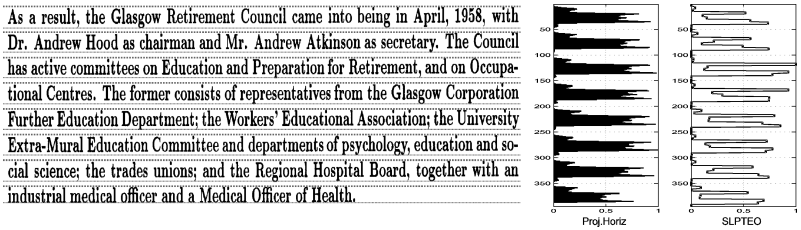


Figura 1. Projeção horizontal e envelope do TEO extraídos do texto impresso "f04-020".

2.3 Rotulação de componentes conectados

A rotulação de componentes conectados é utilizada como uma etapa intermediária em visão computacional. As imagens binárias se constituem de duas regiões: o *foreground*, onde estão os objetos de interesse; e o *background*, o fundo da imagem. Rotular uma imagem binária visa distinguir os objetos significativos da mesma, chamados objetos de interesse, através da conectividade de seus pixels.

A conectividade está diretamente condicionada pela relação de vizinhança entre pixels, uma vez que dois pixels só serão conexos se houver uma sequência de pixels vizinhos que os liga (Peccini & Ornellas, 2005).

Em imagens binárias, onde os possíveis valores assumidos por um pixel são 0 ou 1, dois pixels vizinhos devem possuir o mesmo valor para serem considerados conectados. No entanto, para imagens em níveis de cinza, é

necessário definir regras de similaridade entre os valores dos pixels para possibilitar a determinação de conectividade entre pixels vizinhos.

Existem alguns algoritmos de rotulação de componentes conectados utilizados para percorrer a imagem em uma direção definida rastreando pixel a pixel. Neste caso, os componentes de interesse – que podem ser os componentes 4-conectados ou 8-conectados – são verificados. Ao serem visitados, os pixels que são conexos recebem um rótulo que os distingue dos demais pixels que não apresentam conectividade. Ao final do processo, estes rótulos são ordenados e separados em classes de equivalência (Gonzalez & Woods, 2010; Monteiro, 2002).

2.4 Inferência estatística

De acordo com Barros Neto et al. (2010), "usando planejamentos experimentais baseados em princípios estatísticos, os pesquisadores podem extrair do sistema em estudo o máximo de informação útil, fazendo um número mínimo de experimentos".

Também segundo Barros Neto et al. (2010), chama-se de **população** qualquer coleção de indivíduos ou valores, finita ou infinita. A **amostra** se refere a uma parte da população, normalmente selecionada com o objetivo de se fazer inferências sobre a população. Levando em conta que a amostra precisa ser uma representação realista (não tendenciosa) da população completa, "é necessário que seus elementos sejam escolhidos de forma rigorosamente aleatória". Uma amostra representativa apresenta características relevantes na mesma proporção que ocorrem na população de origem.

Para análise dos resultados deste trabalho, optou-se pelo uso do Intervalo de Confiança (IC) para a média de uma população. Montgomery & Runger (2009) apresentam a formulação para o cálculo do IC em diversas configurações. Dentre elas, adotou-se o cálculo de IC para amostras grandes com "independência da função de distribuição de probabilidade da população". Montgomery & Runger (2009) também especificam que uma amostra com 40 ou mais indivíduos é suficiente para garantir sua classificação como uma "amostra grande".

A equação (4) (Montgomery & Runger, 2009) é o IC da média μ para uma amostra grande com nível de significância α.

$$IC = \left\{ \mu \in \mathbb{R} \;\middle|\; \bar{x} - z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \right\} \tag{4}$$

onde IC é o Intervalo de Confiança; $\bar{x}$ é a média amostral; S é o desvio-padrão amostral; $z_{\alpha/2}$ é o valor fornecido pela tabela da distribuição normal padrão (Z) para o nível de significância α; e n o número de elementos da amostra.

O IC calculado determina os limites bilaterais da média μ, com confiança de $100 \cdot (1-\alpha)\%$. Todos os valores dentro de um IC são equiprováveis.

3. Metodologia

A metodologia empregada neste trabalho trata da segmentação de textos impressos e a apresentação de um estudo de caso envolvendo a aplicação do método SLPTEO em textos manuscritos. Este método é baseado no Operador de Energia de Teager (TEO), aplicado ao sinal gerado pela projeção horizontal (linhas) ou vertical (palavras).

Para o método SCORC, a metodologia empregada é baseada em segmentação de regiões. Os pixels semelhantes são agregados a regiões de interesse durante o processo, utilizando como base a imagem em níveis de cinza. Para isto, foi utilizado os conceitos de componentes rotulados e análise de similaridade entre pixels.

3.1 Segmentação de linhas e palavras: SLPTEO

Para a segmentação de linhas e palavras é importante que a imagem de texto já tenha sido binarizada, pois o algoritmo de segmentação de linhas recebe como entrada a matriz binária da imagem do texto impresso ou manuscrito. Para segmentar linhas, é realizada a projeção horizontal desta imagem. Para segmentar palavras, é utilizada a projeção vertical da imagem binarizada de cada linha de texto.

3.1.1 Algoritmo

Para **segmentação de linhas**, o operador TEO é aplicado aos valores encontrados no histograma (a projeção horizontal) como se fossem parte de um sinal amostrado. Os valores calculados através do operador TEO são armazenados em um vetor, como é demonstrado em na Equação (5).

$$\Psi\left(P_h(y)\right) = P_h(y)^2 - P_h(y+1) \cdot P_h(y-1) \tag{5}$$

onde $\Psi\left(P_h\left(y\right)\right)$ é um vetor contendo os valores de TEO e $P_h\left(y\right)$ é a projeção horizontal da imagem de texto.

A seguir, o método SLPTEO converte os valores deste vetor em valores absolutos. Este vetor é designado como Ω, definido como:

$$\Omega\left(P_h\left(y\right)\right) = \left|\Psi(P_h(y))\right| \tag{6}$$

Em seguida, o vetor Ω é normalizado dentro do intervalo $[0, 1]$, passando a ser identificado por Ω^*, como pode ser observado na Equação (7).

$$\Omega^*\left(P_h\left(y\right)\right) = \frac{\Omega(P_h\left(y\right))}{max(\Omega)} \tag{7}$$

Depois de encontrar os valores de Ω^* para cada ponto da projeção horizontal, é necessário encontrar as fronteiras das linhas de texto. Estas fronteiras são constituídas pelos limites superior e inferior de cada linha

de texto contida na imagem binarizada. Uma janela de 6 pixels[1] (H_{min}) percorre o sinal Ω^* assumindo para cada valor o valor máximo encontrado dentro da janela.

O método SLPTEO varre o vetor Ω^* à procura de fronteiras, partindo de um limiar (pré-definido como 5% da média de Ω^*), como presente na Figura 2.

```
linha = FALSE
FOR EACH i IN Omega*
    IF (Omega*(i) >= limiar & NOT(linha))
        fronteira_superior.adiciona(i)
        linha = TRUE
    ELSEIF (Omega*(i) < limiar & linha)
        fronteira_inferior.adiciona(i)
        linha = FALSE
    ENDIF
ENDFOR
```

Figura 2. Algoritmo para detecção de fronteiras do método SLPTEO.

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 o 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 H_{min}. Se sim, as fronteiras superior e inferior são validadas. Se não, as fronteiras encontradas são removidas.

Após encontradas, as fronteiras sofrem ajustes com a finalidade de estendê-las ou contraí-las, para melhor definição das linhas de texto. Os seguintes ajustes são aplicados (ver Figura 3 para o diagrama de blocos):

1. Localizar linhas com espaço entre linhas

   Percorre-se cada linha identificada por suas fronteiras buscando alguma que contenha $P_h(y) = 0$ em algum y ou seja, projeção horizontal igual a 0 pixels. Se houver, quebra-se a linha em duas.

2. Contrair as fronteiras de linhas

   A cada linha detectada, verifica-se a existência de $P_h(y) = 0$ abaixo de sua fronteira inferior e acima de sua fronteira superior. Caso afirmativo, contrai as fronteiras até algum y tal que $P_h(y) \neq 0$.

[1] Este tamanho foi definido ao se analisar qual seria o menor número de pixels necessário para conter uma linha de texto legível.

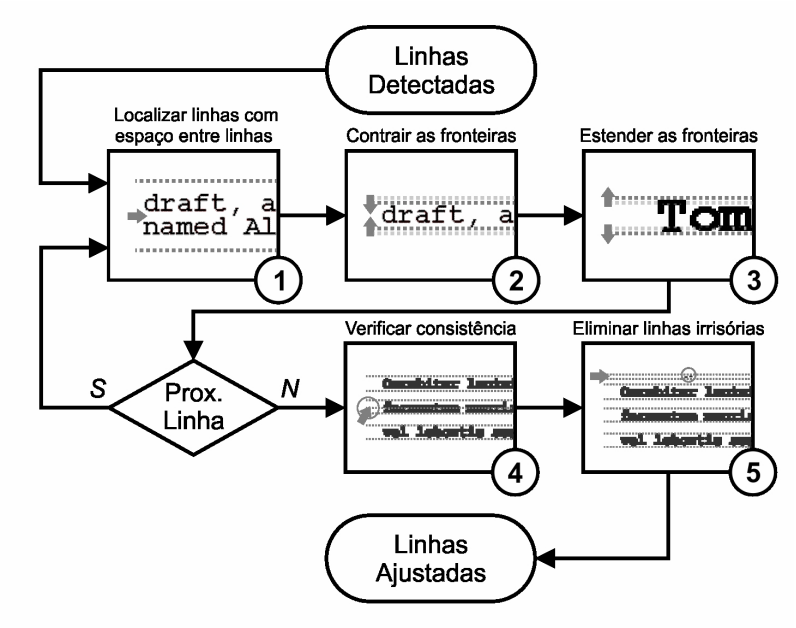


Figura 3. Diagrama de blocos para ajuste de fronteiras.

3. Estender fronteiras

   A cada linha detectada, verifica-se a existência de $P_h(y) \neq 0$ acima de sua fronteira superior e abaixo de sua fronteira inferior. Caso afirmativo, reajusta as fronteiras até algum y tal que $P_h(y) = 0$.

4. Verificar consistência da linhas

   Após o cálculo da altura média das linhas detectadas, é feita a verificação se duas ou mais linhas adjacentes (fronteira inferior de uma está a no máximo 1 pixel da fronteira superior da outra). Caso afirmativo, verifica se pelo menos uma das linhas possui uma altura de no máximo 90% da altura média das linhas detectadas. Se for o caso, unifica as duas linhas definindo uma nova linha com a fronteira superior da 1^a linha e a fronteira inferior da 2^a linha.

5. Eliminação de linhas irrisórias

   Nesta etapa é verificado a existência de linhas. Se for encontrada 1 linha com a altura menor do que o valor-controle de menor altura (H_{min}), esta linha é eliminada.

De maneira análoga, o método funciona para **segmentação de palavras**, só que trabalhando com imagens de linhas de texto já segmentadas

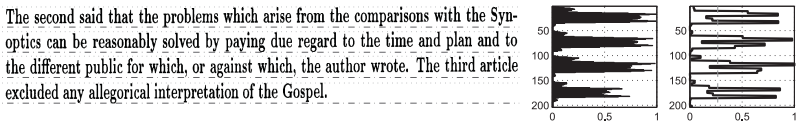
e a projeção vertical $P_v(x)$ da mesma. O algoritmo é **o mesmo** utilizado para a segmentação de linhas. A única diferença é que, ao invés de utilizar H_{min}, tem-se a largura mínima L_{min}. Esta é também o tamanho da janela que percorre Ω^*, e é obtida pela seguinte equação:

$$L_{min} = min(10; 0,15 * H_L) \tag{8}$$

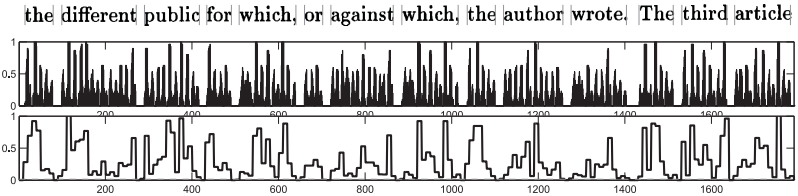
onde H_L é a altura em pixels da imagem da linha de texto segmentada. A ideia é que esta largura será de 10 pixels, no máximo, ou de 15% da altura da linha de texto (estimador sobre tamanho mínimo de letra mais espaço).

3.1.2 Exemplo prático

A Figura 4 mostra o resultado da segmentação de linhas e palavras de um texto impresso através do método SLPTEO.



(a) Linhas (imagem de texto / projeção horizontal / Ω^*).



(b) Palavras (imagem da linha / projeção vertical / Ω^*).

The second said that the problems which arise from the comparisons with the Syn-
optics can be reasonably solved by paying due regard to the time and plan and to
the different public for which, or against which, the author wrote. The third article
excluded any allegorical interpretation of the Gospel.

(c) Texto com fronteiras de segmentação.

Figura 4. Resultado do método SLPTEO de segmentação, aplicado a um texto impresso.

Como se pode observar na Figura 4(a), a partir do vetor Ω^* – obtido através da projeção horizontal da imagem do texto – pode-se definir, com o auxílio do algoritmo presente na Figura 2 e os ajustes posteriores, as fronteiras superiores e inferiores de cada linha de texto. Na Figura 4(b), o mesmo algoritmo é aplicado para segmentação de palavras (apenas trocando H_{min} por L_{min}), baseado no vetor Ω^* – obtido através da projeção vertical da imagem da linha de texto – para definir as fronteiras esquerda

e direita de cada palavra. Na Figura 4(c), tem-se a imagem do texto junto às fronteiras detectadas pelo SLPTEO.

3.2 Segmentação de caracteres: SCORC

Comparando-se uma imagem de palavra de texto em níveis de cinza com uma imagem do mesmo texto binarizada, pode-se verificar que os métodos de limiarização da imagem em seu processo decisório falham na definição da classe de alguns pixels. Ao se analisar a imagem em tons de cinza, pode-se perceber o engano no processo de limiarização, uma vez que o mesmo nível de cinza pode encontrar-se em ambas as classes (objeto ou fundo) em diferentes localizações da imagem, especialmente quando a imagem sofre um processo de suavização em sua digitalização. Este tipo de estratégia (binarização global) frequentemente gera caracteres conectados em imagens de um texto. Com isto, desenvolver o método de segmentação de caracteres que atue em imagens em níveis de cinza, e não em imagens binárias, tornou-se importante para melhoria de desempenho.

A etapa de segmentação de palavras através do método SLPTEO nos fornece como saída as fronteiras da imagem de uma palavra, obtidas através de sua imagem binarizada. O método SCORC, entretanto, parte da imagem da palavra em tons de cinza[2]. Apoiado no limiar de Otsu (1979), o método SCORC se propõe a não incorrer no erro de decisão local dos métodos de binarização globais, possibilitando a correta segmentação dos caracteres que seriam conectados. Note que, para este trabalho, os níveis de cinza da imagem foram distribuídos dentro do intervalo $[0; 1] \in \mathbb{R}$, onde 0 é o nível de cinza equivalente ao preto puro (mínima intensidade luminosa) e 1 ao branco puro (máxima intensidade).

O método de Otsu define um limiar de decisão para considerar um pixel com dado nível de cinza como parte do objeto constituinte ou do fundo da imagem. Os pixels com níveis de cinza próximos ao limiar de Otsu podem pertencer à quaisquer das classes (objeto ou fundo), dependendo do local da imagem em que se encontram (ver Figura 5). Como efeito colateral, eventualmente tem-se a conexão de caracteres originalmente desconexos. Para segmentar estes caracteres, o método SCORC precisa encontrar um pixel inicial que possua um nível de cinza que se enquadre certamente à classe constituinte do objeto[3].

3.2.1 Caracteres especiais

Na Figura 6, são apresentados três tipos de caracteres com estratégias de segmentação não convencional, ou seja, técnicas de segmentação de carac-

[2] Para aplicação do método SCORC, é necessário utilizar as coordenadas geradas pelo SLPTEO para extrair a mesma da imagem original do texto, em tons de cinza.

[3] O método SCORC se aproveita do fato de que o interior do caracter a ser segmentado é mais escuro do que suas bordas.

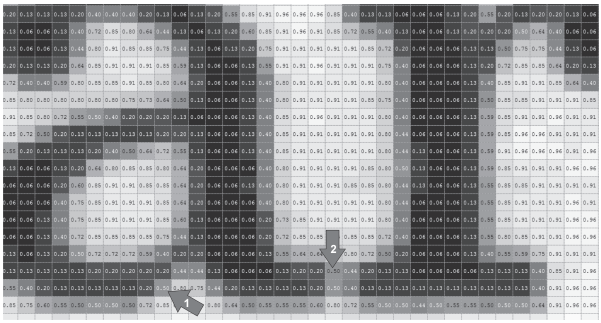


Figura 5. Pixels com mesmo tom de cinza: (1) pode ser considerado parte do objeto; (2) deve ser considerado parte do fundo.

teres "simples" são insuficientes para sua correta segmentação. O método SCORC, além de caracteres simples, consegue lidar com dois destes tipos de problema.

1. **Caracteres com problemas gerados pelo desenho da fonte** – Exemplos na Figura 6(a). A segmentação destes tipos de caracteres é um problema complicado não relacionado ao processo de digitalização da imagem do texto. São resultados intrínsecos às escolhas feitas pelo *designer* das fontes em questão. Este tipo de problema não é resolvido pelo método proposto.

2. **Caracteres sobrepostos** – Exemplos na Figura 6(b). Um método baseado em segmentação orientada a regiões com rotulação de componentes pode resolver este tipo de problema, inclusive quando aplicado a imagens binárias.

3. **Caracteres conectados** – Exemplos na Figura 6(c). Conexão entre caracteres é um típico problema de segmentação de caracteres. Pode ser gerado por um processo de digitalização ruidoso, ou simplesmente pela limiarização de uma imagem em tons de cinza com estratégia global. Segundo Saba et al. (2010), as pesquisas que tratam deste tipo de problema focam na solução de conexão entre dígitos (ou sequências de dígitos), mas poucas tratam de alfabetos. Outro ponto é que soluções para este problema são fortemente vinculadas à linguagem escrita e à língua de apresentação. O SCORC se propõe a resolver conexões de caracteres, indiferente da fonte ou da língua utilizadas, restringindo-se às conexões geradas pelo processo de limiarização aplicado em imagens de texto impresso em tons de cinza.

(a) Desenho de fontes (b) Sobrepostos (c) Conectados

Figura 6. Três tipos de caracteres com estratégias de segmentação não convencional.

3.2.2 Algoritmo

Na primeira etapa, todos os valores únicos de níveis de cinza da imagem da palavra são relacionados e ordenados em um vetor. Os níveis de cinza acima do limiar de Otsu (fundo da imagem) são descartados do vetor. A seguir, é escolhido o maior nível de cinza menor ou igual à *mediana* deste vetor. Este valor é o máximo nível de cinza aceitável (C_{max}) para um pixel a ser procurado na imagem da palavra segmentada. O algoritmo então percorre os pixels da imagem de cima para baixo e da esquerda para a direita, procurando um pixel com o nível menor ou igual ao C_{max}. A ideia é que este pixel sinaliza a existência de um caractere (objeto) a ser segmentado. A partir deste pixel, o método SCORC começa o processo de segmentação.

O método SCORC teve inspiração nos métodos de segmentação por regiões, onde os pixels visitados são agregados a determinadas regiões segundo a similaridade que existe entre eles (Gonzalez & Woods, 2010). Logo, após encontrarmos o pixel com C_{max}, o método começa a visitar recursivamente seus 8 vizinhos, identificado – dentro das regras pré-estabelecidas – se são pertencentes à mesma "região". Em outras palavras, o método procura recursivamente os pixels conectados que possuam um tom de cinza semelhante ao do pixel original.

A cada pixel visitado é calculada a máxima variação de tons de cinza para que seja permitida ou não a transição para cada um dos pixels vizinhos. A Equação (9) fornece este nível máximo de "salto" que o pixel observado pode dar, ou seja, qual a diferença máxima entre tons para que um vizinho possa ser considerado parte da região do pixel observado ou não. Quanto mais próximo de zero o nível deste pixel for (mais escuro), maior vai ser o valor encontrado para uma possível transição. Por outro lado, quanto mais o nível de cinza se aproximar do valor do limiar de Otsu (mais claro), mais restrito será o valor de permissão para uma transição.

$$\Delta g = A \cdot T \cdot \left(\frac{g - T}{g_0 - T} \right) \tag{9}$$

onde Δg é a variação máxima (superior ou inferior) entre tons vizinhos para que se aceite que pertençam à mesma região; A é o fator de escala para a inclinação da reta que é fronteira de decisão entre a similaridade ou não de um tom[4]; T é o valor do limiar de Otsu para a imagem da palavra analisada; g é o nível de cinza do pixel observado; e g_0 é o menor valor de nível de cinza encontrado na imagem da palavra analisada (i.e., o pixel mais escuro).

Se a intensidade de um pixel vizinho estiver dentro do intervalo[5] definido por $[g - \Delta g; T]$, este será incorporado à região a ser segmentada e se torna o pixel observado, para a análise de seus vizinhos, seguindo assim de forma recursiva. Caso contrário, não será rotulado como constituinte da região de interesse, nem terá seus vizinhos analisados.

3.2.3 Exemplo prático

Observe a Figura 7 que mostra o mapa de pixels da imagem da letra "a", extraída da base de imagens utilizada.

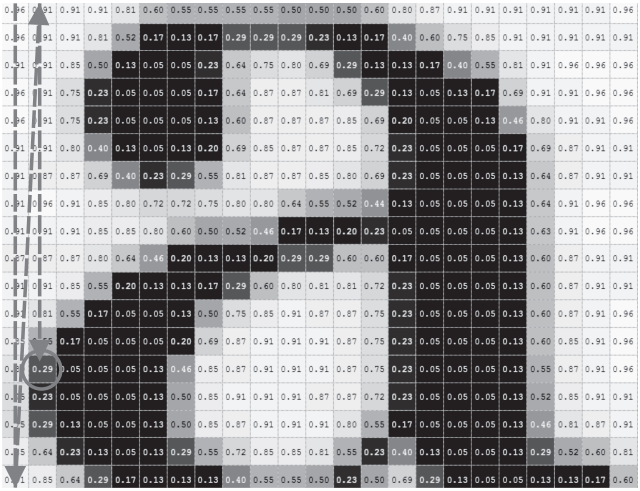


Figura 7. Imagem da letra "a" para exemplo.

O método SCORC seleciona os níveis de cinza únicos da imagem da palavra e em seguida os ordena em um vetor. A Tabela 1 mostra os níveis de cinza da Figura 7.

A partir do valor do limiar de Otsu – neste exemplo, $0,5725$ – é realizada uma seleção dos níveis de cinza menores ou iguais a este limiar (classe objeto), e são descartados os pixels que possuem valores maiores do que este

[4] Para este trabalho, o fator de escala foi definido empiricamente como $A = 1,2$.
[5] A grandeza $g + \Delta g$ sempre é maior do que T (limiar de Otsu).

Tabela 1. Níveis de cinza únicos da imagem "a" (em negrito, os níveis permitidos para o início do SCORC; em itálico, os níveis abaixo do limiar de Otsu).

0,0549	*0,5176*	0,8078
0,1333	*0,5490*	0,8509
0,1686	0,5999	0,8666
0,2000	0,6313	0,9058
0,2274	0,6392	0,9568
0,2862	0,6901	0,9685
0,4000	0,7176	0,9725
0,4353	0,7489	0,9764
0,4588	0,7529	0,9803
0,5019	0,7999	0,9842

limiar (classe fundo). De posse dos níveis de cinza relevantes é calculada a mediana destes valores. Neste caso, o valor encontrado é $C_{max} = 0,2862$ (na Figura 7, tem-se $0,29$).

Na primeira etapa, o método SCORC começa a visitar os pixels da imagem, começando pela coordenada superior esquerda na direção cima-baixo e esquerda-direita, como representado na Figura 7. O método busca um pixel cujo valor seja menor ou igual a C_{max}, garantindo com relativa certeza que este pixel é parte constituinte do caractere. Assim, este primeiro pixel visitado torna-se a porta de entrada para o método começar o processo de rotulação.

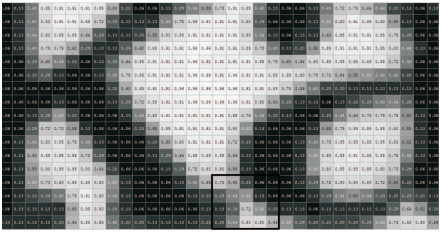
Para todo pixel observado é realizada uma análise de possibilidades de transição para este pixel baseado na Equação (9). Para este exemplo prático, tem-se:

$$\Delta g = 1,2 \cdot 0,5725 \cdot \left(\frac{g - 0,5725}{0,0549 - 0,5725} \right) = -1,33 \cdot g + 0,76 \qquad (10)$$

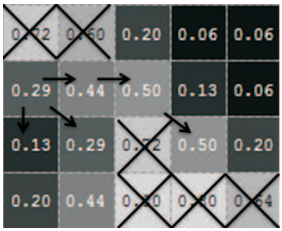
onde g é o valor do nível de cinza de cada pixel observado; e Δg é a variação máxima de intensidade para aceitação de um pixel vizinho como sendo da mesma região do pixel observado. Por exemplo, para $g = 0,2862$, a transição só é possível se o nível de cinza do pixel vizinho analisado estiver entre $-0,0938$ (considera-se 0, pois é menor do que o preto puro) e o limiar de Otsu. Já para $g = 0,4353$, a transição só é possível se o vizinho estiver entre $0,2532$ e o limiar de Otsu.

A partir do cálculo de Δg para cada pixel observado, é determinado um intervalo de permissão de transição para seus pixels vizinhos. Portanto, para todo pixel visitado estas transições são avaliadas, isto é feito de forma recursiva obedecendo a vizinhança-de-8 dos pixels. Ou seja, para cada pixel observado é realizada uma análise de transição de todos os oito vizinhos

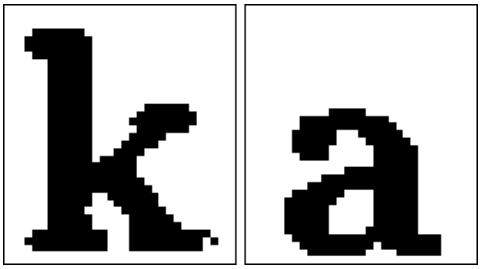
deste pixel. Quando estes vizinhos se enquadram no intervalo definido eles automaticamente são rotulados e se constituem parte do objeto de interesse, caso contrário, não são rotulados. O método SCORC encerra sua busca quando ele não mais encontra vizinhos semelhantes. Neste momento, o método extrai da imagem original toda a região rotulada, salva este caractere (região extraída), e inicia um novo processo procurando C_{max} na imagem resultante da extração. Na Figura 8 é mostrado o resultado de segmentação de dois caracteres conectados ("k" e "a").



(a) Mapa de pixels.



(b) Decisão do SCORC.



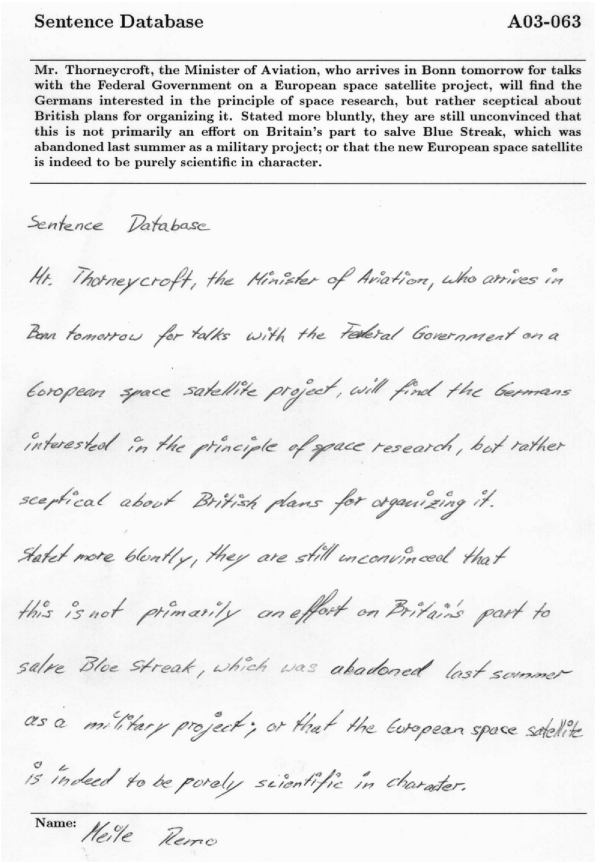
(c) Caracteres segmentados.

Figura 8. Exemplo de segmentação de "k" e "a".

3.3 Base de dados

A base de imagens *IAM-Database* (Marti & Bunke, 2002) é uma conceituada base de imagens para reconhecimento de documentos de textos impressos e manuscritos. A base contém formulários de textos digitalizados com resolução de 300 dpi e armazenados no formato de imagem PNG com 256 níveis de cinza. Cada formulário da *IAM-Database* contém uma parte de texto impressa e a transcrição em escrita manual do mesmo texto. Contribuíram para a criação da base de dados 657 autores. Dos 1.539 formulários, foram disponibilizados via internet 1.507 páginas de texto digitalizadas[6]. A Figura 9 apresenta um dos formulários extraídos desta base de dados utilizado neste trabalho.

[6] http://www.iam.unibe.ch/fki/databases/iam-handwriting-database



Sentence Database A03-063

Mr. Thorneycroft, the Minister of Aviation, who arrives in Bonn tomorrow for talks with the Federal Government on a European space satellite project, will find the Germans interested in the principle of space research, but rather sceptical about British plans for organizing it. Stated more bluntly, they are still unconvinced that this is not primarily an effort on Britain's part to salve Blue Streak, which was abandoned last summer as a military project; or that the new European space satellite is indeed to be purely scientific in character.

Sentence Database

Mr. Thorneycroft, the Minister of Aviation, who arrives in Bonn tomorrow for talks with the Federal Government on a European space satellite project, will find the Germans interested in the principle of space research, but rather sceptical about British plans for organizing it. Stated more bluntly, they are still unconvinced that this is not primarily an effort on Britain's part to salve Blue Streak, which was abadoned last summer as a military project; or that the European space satellite is indeed to be purely scientific in character.

Name: Heike Renn

Figura 9. Formulário "a03-063" da *IAM-Database*.

Para este trabalho, foram escolhidos aleatoriamente 40 textos da base e extraídos apenas a parte impressa de cada formulário. A parte manuscrita dos mesmos foram utilizadas em um estudo de caso não supervisionado[7] (Seção 4.1.1). Em uma primeira etapa, foram calculados apenas os 'percentuais de acerto' dos resultados. Posteriormente, foram realizadas inferências sobre os resultados dos métodos abordados para toda a população de 1.507 textos.

[7] Entenda-se "estudo de caso não supervisionado" como um experimento *per se*, sem pretensões de utilização dos resultados para a melhoria do método.

4. Resultados

O método SLPTEO se mostrou bastante robusto para textos impressos. Todas as linhas e palavras dos 40 textos impressos constituintes da nossa base de dados foram segmentadas corretamente[8]. Este indicativo é importante para inferirmos sobre nossa base de dados, já que a mesma mostrou-se extremamente bem comportada. Para avaliarmos a efetividade do método, aplicou-se o SLPTEO a textos manuscritos – o mesmo algoritmo – apresentando os resultados na Seção 4.1.1, como um estudo de caso.

O método SCORC conseguiu solucionar o problema dos caracteres sobrepostos, além de resolver grande parte do problema dos caracteres conectados. De complexidade linear ao tamanho da imagem, o método mostrou-se flexível em solucionar dois tipos de problema existentes sem a necessidade de prévia identificação dos mesmos. Assim, o mesmo método pôde ser usado para segmentar caracteres simples, conectados ou sobrepostos, à medida que percorre a imagem da palavra em níveis de cinza.

4.1 Segmentação de linhas e palavras

Os resultados encontrados para o método SLPTEO aplicados aos 40 textos impressos estão presentes na Tabela 2.

Tabela 2. Resultados do SLPTEO em textos impressos.

	Total	**Segmentadas**	**Acerto**
Linhas	202	202	100%
Palavras	2655	2655	100%

Como se pode observar, o método foi suficiente para segmentar linhas e palavras de textos impressos. Nota-se que, ao alcançar 100% de correta segmentação, pode-se dizer que o método é extremamente eficaz para textos bem comportados.

4.1.1 Estudo de caso: textos manuscritos

Apresentando como um estudo de caso não supervisionado, tem-se os resultados da aplicação do SLPTEO a textos manuscritos. Foi utilizado o mesmo algoritmo, sem nenhuma modificação. Os textos manuscritos utilizados foram extraídos dos 40 formulários da *IAM-Database* utilizados para os experimentos com textos impressos. Os resultados podem ser conferidos na Tabela 3.

Pode-se observar que o método SLPTEO é bastante robusto com relação à segmentação de linhas, com uma pequena queda de desempenho com relação à segmentação de palavras. Foi verificado que esta queda de desempenho se deve aos textos que possuem escrita inclinada. Uma maneira de

[8] O texto "d01-052" foi previamente rotacionado em 0,1°.

Tabela 3. Resultados do SLPTEO em textos manuscritos.

	Total	Segmentadas	Acerto
Linhas	341	337	98,8%
Palavras	2606	2197	84,3%

minimizar este problema seria efetuar a correção da inclinação da escrita em cada linha de texto (normalização), como no trabalho de Vinciarelli & Luettin (2001). Em tempo de pré-processamento, este tipo de correção poderia melhorar o desempenho do SLPTEO na segmentação de palavras.

4.2 Segmentação de caracteres

A base de dados utilizada apresentou alguns tipos de problemas de caracteres: caracteres conectados, caracteres sobrepostos e desenhos de fonte. Destes, somente o desenho de fonte não foi abordado na pesquisa, como mencionado na Seção 3.2.1.

A Tabela 4 mostra os percentuais de caracteres segmentados corretamente na base de dados. Para cada problema encontrado, neste caso caracteres conectados e caracteres sobrepostos, tem-se a quantidade de ocorrências e o número de caracteres segmentados pelo método. Note que 71,90% dos casos de conectados foram solucionados. Já os casos de sobreposição deixaram de existir após a segmentação pelo método SCORC.

Tabela 4. Resultados do SCORC em Textos Impressos.

	Ocorrências	Segmentados	Acerto
Conectados	64	46	71,9%
Sobrepostos	445	445	100%

4.3 Inferências para o conjunto de dados completos

Baseados nos resultados obtidos para os 40 textos aleatórios, pode-se fazer inferências sobre os resultados dos métodos quando aplicados à toda população de 1.507 textos da base *IAM-Database*.

Para os experimentos deste trabalho, foi definido um nível de confiança igual a 99%. Isto significa que com 99% de certeza pode-se afirmar que a média populacional está dentro do IC, ou seja, os resultados estarão pautados neste intervalo encontrado.

Com base na Equação (4), foram substituidos os valores utilizados neste trabalho. Para o valor da distribuição normal padrão (Z) tem-se $z_{\alpha/2} = 2,58$ (Montgomery & Runger, 2009), o que corresponde à área de 99% na distribuição normal com amostra contendo $n = 40$ elementos. O IC para a média da população aqui tratada será:

$$IC = \left\{ \mu \in \mathbb{R} \;\middle|\; \bar{x} - 2,58 \cdot \frac{S}{\sqrt{40}} \leq \mu \leq \bar{x} + 2,58 \cdot \frac{S}{\sqrt{40}} \right\} \quad (11)$$

Para o caso do SLPTEO aplicado a textos impressos, onde o percentual de acerto foi de 100% para segmentação de linhas e palavras, a inferência para a população não é necessária. Estes resultados indicam que a base de dados de textos impressos é bastante comportada, indicando um perfeito funcionamento do método. Já para o estudo de caso realizado (Seção 4.1.1), podem-se inferir para a população. Os resultados são apresentados na Tabela 5, com 99% de confiança.

Tabela 5. Intervalo de confiança para ocorrências de linhas e palavras antes e depois da aplicação do método SLPTEO em textos manuscritos

Ocorrências	**Segmentação**	
	Visual	**Computacional**
Linhas	7,75 a 9,30	7,64 a 9,21
Palavras	60,20 a 70,10	49,44 a 60,41

A coluna "Segmentação Visual" da Tabela 5 mostra o intervalo de confiança gerado a partir dos dados obtidos pela contagem visual do número de linhas e palavras dos 40 textos amostrados. A coluna "Segmentação Computacional" apresenta o intervalo de confiança gerado a partir dos dados obtidos após o uso do método SLPTEO nos textos amostrados, extrapolando, assim, os resultados referentes à detecção automática de linhas e palavras.

Pode-se notar, para a segmentação automática (computacional) de linhas em textos manuscritos, o intervalo inferido para toda a base de imagens de textos é bastante próximo ao intervalo obtido quando se utilizou a segmentação visual (humana). Este resultado vem de encontro ao objetivo primário de todo sistema de automação de tarefas – capacitar o computador para que ele desempenhe um papel ou uma tarefa que seria fácil para um ser humano.

Com relação ao intervalo para a segmentação automática (computacional) das palavras, o método apresenta uma queda de desempenho com relação ao intervalo gerado pela contagem visual (humana), o que é coerente com as evocações descritas na Tabela 3. Embora fosse inviável para uma aplicação real genérica, uma etapa de normalização da escrita poderia melhorar o desempenho apresentado pelo SLPTEO com relação à segmentação de palavras manuscritas.

O método SCORC se mostrou bastante eficiente quando aplicado às palavras de textos impressos. A Tabela 6 mostra os intervalos de confianças para as médias, com 99% de confiança após a utilização do método SCORC:

Tabela 6. Intervalo de Confiança da segmentação de caracteres pelo método SCORC

Ocorrências	Segmentação	
	Visual	Computacional
Caracteres Conectados	0,75 a 2,45	0,02 a 0,88
Caracteres Sobrepostos	8,37 a 13,88	0
Caracteres Fragmentados	0	0 a 0,18

A coluna "Segmentação Visual" da Tabela 6 apresenta as inferências sobre a média por texto dos casos de caracteres conectados, sobrepostos e fragmentados, geradas a partir dos dados obtidos pela contagem visual (humana) dos mesmos nos 40 textos amostrados. A coluna "Segmentação Computacional" apresenta o intervalo de confiança gerado a partir dos dados obtidos após o uso do método SCORC nos textos amostrados, extrapolando assim os resultados referentes à detecção automática de caracteres.

Comparando-se os intervalos visuais e computacionais, pode-se notar um desempenho interessante para a correta segmentação de caracteres conectados. Por exemplo, se fosse esperada uma média de quase 3 caracteres conectados por texto na base de dados utilizada, contados visualmente, o método SCORC nos garantiria uma nova média de menos de um caractere conectado por texto. Quando se comparam os limites dos intervalos visual e computacional para a segmentação de caracteres conectados, tem-se uma redução na média de ocorrências mínima de 65%, quando comparados aos máximos dos intervalos, e uma máxima de $97,3\%$ quando comparados os mínimos dos intervalos.

No caso dos caracteres sobrepostos, o método SCORC solucionou todos os casos. Ou seja, pode-se dizer com 99% de confiança que todos os casos de caracteres sobrepostos da base de textos utilizadas (1.507 textos da base *IAM-DataBase*) são resolvidos por este método.

Surgiu um novo problema após a aplicação do método SCORC: apareceram 3 caracteres fragmentados nos resultados da aplicação do método nos 40 textos amostrados. Considerando os 12.816 caracteres, este problema representou $0,023\%$ dos caracteres presentes nos textos amostrados. Além disto, o método SCORC resultou em $99,10\%$ de caracteres segmentados corretamente, considerando todos os casos aqui abordados. Ou seja, este problema, embora indesejável, não chega a comprometer a eficiência do método. Extrapolando para a base de textos inteira, tem-se, no máximo, 0,18 caracteres fragmentados em média por texto.

5. Conclusões

Neste trabalho foram desenvolvidos dois métodos: SLPTEO e SCORC. Ambos contribuem para a pesquisa em reconhecimento de textos impressos,

na área de processamento digital de imagens, quando aplicados às etapas de segmentação de linhas, palavras e caracteres.

Baseado no operador de energia de Teager, o método SLPTEO foi desenvolvido e utilizado para a segmentação de linhas e palavras de textos impressos e também aplicado à textos manuscritos. O método SLPTEO apresentou resultados bastante interessantes quanto a segmentação de linhas tanto para textos impressos como para textos manuscritos. Embora para uma base comportada um simples método baseado em projeção horizontal seria suficiente para a separação de linhas, o mesmo não pode ser dito de um método simples baseado em projeção vertical para separação de palavras. Um dos diferenciais do SLPTEO está no fato de que um mesmo algoritmo é utilizado para segmentação de linhas para ambos os textos impressos e manuscritos. Além disto, o o SLPTEO demonstrou sua eficiência na segmentação de palavras para textos impressos ou manuscritos, independente de qualquer mensuração de distâncias entre palavras e/ou caracteres.

Para segmentação de caracteres, o método SCORC se apresentou suficiente para solução na segmentação de caracteres simples ou para os problemas de caracteres conectados e sobrepostos. O grande diferencial do método é que este utiliza imagens em tons de cinza e as inferências utilizadas são baseadas nas características únicas do nível de cinza de cada pixel analisado. Associado ao método de limiarização de Otsu, o método SCORC delimita as possibilidades de transição entre pixels, estabelecendo um degrau disponível de acordo com o valor de cada pixel na imagem da palavra.

O método SCORC, além de ter se mostrado eficiente, é simples e – por não depender de outras abordagens como a projeção linear ou o cálculo de distância entre caracteres – pode ser considerado rápido, com complexidade linear com referência ao número de pixels da imagem. Além disto, não há a necessidade de prévia identificação dos caracteres especiais dentro de um texto. O método SCORC trata todos os caracteres da mesma maneira, independente da natureza do caractere (simples, sobreposto ou conectado).

Este trabalho se propõe a dar suporte para a evolução de novos estudos e métodos. Assim, propomos algumas estratégias que poderão ser exploradas a partir dos métodos aqui abordados:

- Um sistema de reconhecimento desenvolvido a partir dos caracteres segmentados neste trabalho.
- O método SLPTEO pode ser aplicado a outras bases de dados, em conjunto com estratégias de pré-processamento que ajustem a qualidade da imagem dos textos.
- Estudos em identificação de linhas de textos e palavras em textos manuscritos podem fazer uso do método SLPTEO.

- Para o método SCORC, estudos adicionais são necessários para automatizar o valor do fator de escala A que afeta a inclinação da reta para o cálculo de Δg, presente na Equação (9).
- Suporte ao desenvolvimento de novos algoritmos utilizando as características de imagens em tons de cinza para segmentação de outros tipos de imagens.

Referências

Barros Neto, B.; Scarminio, I.S. & Bruns, R.E., *Como Fazer Experimentos*. 4a edição. Porto Alegre, RS: Bookman, 2010.

Bezerra, C.M.C., *BR BRAILLE: Programa Tradutor de Textos Braille Digitalizados para Caracteres Alfanuméricos em Português*. Dissertação de mestrado em engenharia elétrica, Universidade Estadual de Campinas, 2003.

Conci, A.; de Carvalho, J.E.R. & Rauber, T.W., A complete system for vehicle plate localization, segmentation and recognition in real life scene. *IEEE Latin America Transactions*, 7(5):497–506, 2009.

Dirichlet, P.G., Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données. *Journal für die Reine und Angewandte Mathematik*, 4(7):157–169, 1829.

Gonzalez, R.C. & Woods, R.E., *Processamento Digital de Imagens*. 3a edição. São Paulo: Pearson Prentice-Hall, 2010.

Jung, M., Character segmentation using side view feature in machine-printed optical characters recognition. *Journal of Korean Institute of Information Technology*, 8(12):271–280, 2010.

Kaiser, J., On a simple algorithm to calculate the 'energy' of a signal. In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*. Piscataway, USA: IEEE Press, v. 1, p. 381–384, 1990.

Kaiser, J.F., Some useful properties of Teager's energy operators. In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*. Piscataway, USA: IEEE Press, v. 3, p. 149–152, 1993.

Lue, H.T.; Wen, M.G.; Cheng, H.Y.; Fan, K.C.; Lin, C.W. & Yu, C.C., A novel character segmentation method for text images captured by cameras. *ETRI Journal*, 32(5):729–739, 2010.

Marti, U.V. & Bunke, H., The IAM-database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1):39–46, 2002.

Monteiro, L.H., *Utilização de Técnicas de Processamento de Imagens para o Reconhecimento de Placas de Veículos*. Dissertação de mestrado em computação, Universidade Federal Fluminense, 2002.

Montgomery, D.C. & Runger, G.C., *Estatística Aplicada e Probabilidade para Engenheiros*. 4a edição. Rio de Janeiro: LTC, 2009.

Nikolaou, N.; Makridis, M.; Gatos, B.; Stamatopoulos, N. & Papamarkos, N., Segmentation of historical machine-printed documents using adaptive run length smoothing and skeleton segmentation paths. *Image and Vision Computing*, 28(4):590–604, 2010.

Otsu, N., A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.

Pavlidis, I.; Papanikolopoulos, N.P. & Mavuduru, R., Signature identification through the use of deformable structures. *Signal Processing*, 71(2):187–201, 1998.

Peccini, G. & Ornellas, M.C., Segmentação de imagens por *Watersheds*: Uma implementação utilizando a linguagem java. *Revista Eletrônica de Iniciação Científica*, V(IV), 2005.

Pedrini, H. & Schwartz, W.R., *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. São Paulo: Thomson Learning, 2008.

Peretta, I.S., *A Novel Word Boundary Detector Based on the Teager Energy Operator for Automatic Speech Recognition*. Dissertação de mestrado em engenharia elétrica, Universidade Federal de Uberlândia, 2010.

Peretta, I.S.; Lima, G.F.M.; Tavares, J.A. & Yamanaka, K., A spoken word boundaries detection strategy for voice command recognition. *Learning and Nonlinear Models*, 8(3):148–156, 2010.

Saba, T.; Sulong, G. & Rehman, A., A survey on methods and strategies on touched characters segmentation. *International Journal of Research and Reviews in Computer Science*, 1(2):103–114, 2010.

Silva, L.F., *Distinção Automática de Texto Impresso e Manuscrito em uma Imagem de Documento*. Dissertação de mestrado em computação, Universidade Federal Fluminense, 2009.

Teager, H.M. & Teager, S.M., Evidence for nonlinear production mechanisms in the vocal tract. In: *Proceedings of the NATO Advanced Study Institute on Speech Production and Speech Modelling*. p. 241–261, 1990.

Vinciarelli, A. & Luettin, J., A new normalization technique for cursive handwritten words. *Pattern Recognition Letters*, 22(9):1043–1050, 2001.

Notas Biográficas

Josimeire do Amaral Tavares é graduada em Sistemas de Informação (PUC/MG, 2008) e mestre em Engenharia Elétrica (Universidade Federal de Uberlândia – UFU, 2011). Atualmente é professora de ensino superior atuando na área de projetos de sistemas, além de gestão da informação e Tecnologia da Informação.

Igor Santos Peretta é graduado em Engenharia Elétrica (UNICAMP, 2002) e mestre em Engenharia Elétrica (UFU, 2010). Possui experiência na área de engenharia elétrica, com ênfase em inteligência computacional. Atualmente é doutorando no programa de pós-graduação em Engenharia Elétrica (linha de pesquisa em inteligência artificial) da UFU. Tem interesse na área de meta-heurísticas, estatística, aprendizado de máquina e computação evolucionária.

Gerson Flávio Mendes de Lima é graduado em Engenharia Eletrotécnica (UFU, 1994), mestre em Engenharia Elétrica (UFU, 2008). Atualmente é doutorando em Computação Gráfica na UFU e Pesquisador - Sistemas Ópticos - FIBERWORK - Comunicações Ópticas. Possui experiência na área de projetos de instalações elétricas, projetos eletrônicos de redes de HFC, atuando principalmente nos seguintes temas: computação gráfica, cartografia digital, sistemas de mapas digitais na enternet, aplicações para telecomunicações e algoritmos de inteligência computacional, contemplando redes neurais artificiais e algoritmos genéticos.

Keiji Yamanaka é doutor em Engenharia Elétrica e de Computação (Nagoya Institute of Technology, Japão, 1999). Atualmente é Professor Associado-2 da UFU. Tem experiência na área de engenharia elétrica, com ênfase em inteligência computacional, atuando principalmente nos seguintes temas: redes neurais artificiais, algoritmos genéticos, e em reconhecimento de padrões.

Mônica Sakuray Pais é graduada e mestre em Ciência da Computação (UNICAMP, 1986 e UFU, 2004, respectivamente). Atualmente é Professora do Ensino Básico, Técnico e Tecnológico do Instituto Federal Goiano – Campus Urutaí e doutoranda em Engenharia Elétrica na UFU. Tem experiência na área de ciência da computação, com ênfase em metodologia e técnicas da Computação, atuando principalmente nos seguintes temas: bancos de dados, integração de dados, lógica paraconsistente, bancos de dados dedutivos, linguagem de consultas a bancos de dados e banco de dados inconsistentes.

Capítulo 14

Reconhecimento de Caracteres Baseado em Regras de Transições entre *Pixels* Vizinhos

Francisco Assis da Silva*, Almir Olivette Artero,
Maria Stela Veludo de Paiva e Ricardo Luís Barbosa

Resumo: Este capítulo trata do reconhecimento de caracteres impressos e manuscritos, apresentando um algoritmo totalmente baseado na análise do comportamento das transições entre os *pixels* vizinhos nas imagens dos caracteres. A partir desta análise, são definidas regras que determinam em qual classe cada caractere deve ser colocado, caracterizando uma classificação supervisionada. A baixa complexidade deste algoritmo tem tornado possível o seu uso em aplicações onde o tempo de reconhecimento é bastante crítico, como é o caso de sistemas de reconhecimento em tempo real, usados em sistemas de visão computacional, como robôs e veículos não tripulados.

Palavras-chave: Reconhecimento de caracteres, Classificação supervisionada, Análise de transições entre *pixels*, Processamento de vídeo.

Abstract: *This chapter deals with the recognition of printed and handwritten characters, and presents an algorithm based exclusively on the analysis of the behavior of transitions between neighboring pixels in the images of the characters. From this analysis, rules are defined to determine in which class each character should be placed, corresponding to a supervised classification scheme. The low complexity of this algorithm has made its use possible in applications where the recognition time is quite critical, such as real-time recognition systems used in computer vision systems for robots and autonomous vehicles.*

Keywords: *Character recognition, Supervised classification, Pixel transition analysis, Video processing.*

*Autor para contato: chico@unoeste.br

1. Introdução

O problema do reconhecimento de caracteres têm chamado atenção há bastante tempo, com os trabalhos pioneiros de Tauschek, que obteve a patente do OCR (*Optical Character Recognition*) na Alemanha em 1929 e, posteriormente, nos Estados Unidos (Tauschek, 1935), e de Handel (1933), que também registrou uma patente nos Estados Unidos. A partir da década de 50, com o impulso gerado pelos computadores (Dimond, 1957; Neisser & Weene, 1960; Eden, 1961; Eden & Halle, 1961; Frishkopf & Harmon, 1961), a área se tornou ainda mais atrativa, surgindo uma grande variedade de propostas de algoritmos para resolver este problema. Atualmente, o reconhecimento de caracteres continua sendo uma área de intensa pesquisa, que ainda apresenta vários problemas, por causa da grande diversidade de formas que os caracteres podem assumir, principalmente, no caso de caracteres manuscritos (ICR, *Intelligent Character Recognition*) (Gonzalez & Woods, 2001; Montaña, 2007; Jain & Ko, 2008; Pereira et al., 2010; Shrivastava & Gharde, 2010; Trentini et al., 2010; Lin et al., 2011). Mesmo no caso do reconhecimento de caracteres impressos (OCR), que é uma tarefa mais simples, ainda persistem diversas dificuldades, por causa da grande diversidade de fontes que podem ser usadas nos documentos.

Apesar do grande número de propostas apresentadas para o reconhecimento dos caracteres, existem duas tarefas básicas nesta área, que são a abordagem estrutural (Pavlidis, 1980; Schalkoff, 1992) e a abordagem estatística (Schalkoff, 1992; Duda et al., 2001; Jain et al., 2000), que inclui a extração de atributos e a classificação dos objetos a partir de informações obtidas a partir de objetos conhecidos (Zhu et al., 2000). Entretanto, para realizar uma classificação satisfatória dos caracteres, é preciso usar uma grande quantidade de atributos que, muitas vezes, compromete a execução da tarefa. Na segunda abordagem, o reconhecimento dos caracteres pode ser feito através de uma classificação supervisionada, quando se busca obter informações que permitam predizer, com precisão, a classe de cada amostra a partir de medições realizadas em caracteres cujas classes são conhecidas. Uma alternativa a este processo é a classificação não supervisionada, também chamada de agrupamento e, neste caso, procura-se, simplesmente colocar os caracteres em classes, onde eles apresentam grande similaridade entre seus integrantes e baixa similaridade entre os elementos de classes distintas.

Além de realizar a classificação correta dos caracteres, um dos principais desafios da área é o tempo de processamento, que precisa ser muito baixo, para que a tarefa possa ser feita em tempo aceitável. Assim, 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. Por causa de sua baixa complexidade, o algoritmo consegue excelente tempo de processamento, o

que permite a sua aplicação em tarefas consideradas de tempo real, como é o caso do reconhecimento de placas em rodovias, durante o deslocamento do veículo. As demais seções deste capítulo estão organizadas da seguinte maneira: na Seção 2 são apresentados alguns trabalhos relacionados ao 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 dos caracteres; a Seção 4 apresenta alguns experimentos realizados com esta proposta em um conjunto real de dados, bem como os resultados obtidos e a análise de desempenho; por fim, na Seção 5 são apresentados os comentários finais e trabalhos futuros.

2. Trabalhos Relacionados

Embora o problema do reconhecimento de caracteres tenha atraído atenção desde os primórdios da computação, com a atual tendência de uso dos *tablets*, o reconhecimento de caracteres manuscritos continua sendo uma área de grande interesse pelos fabricantes destes dispositivos, pois a inserção eficaz de textos em dispositivos sem o tradicional teclado QWERTY é um importante diferencial entre os aparelhos. Alguns trabalhos recentes nesta área são descritos a seguir.

O trabalho de Montaña (2007) realiza o reconhecimento de padrões de dígitos empregando redes neurais. Em seu trabalho são usadas duas redes neurais diferentes para alcançar os resultados, uma rede Perceptron Multicamadas e uma rede baseada no Mapa de Kohonen.

O trabalho de Jain & Ko (2008) apresenta um algoritmo de classificação para reconhecer dígitos numéricos manuscritos (0-9), sendo utilizada a implementação da Análise de Componentes Principais (*Principal Component Analysis* – PCA), combinada com o algoritmo do primeiro vizinho mais próximo (1-*nearest neighbor*) para reconhecer os dígitos.

A contribuição do trabalho de Pereira et al. (2010) é melhorar a precisão do reconhecimento de caracteres manuscritos usando um novo método de extração de características, também baseado em Análise de Componentes Principais. 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. (2010) 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 & Gharde (2010) é utilizado para reconhecimento de números *Devanagari* manuscritos. *Devanagari* é um alfabeto manuscrito 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 $m \times n$ gera um conjunto com $m.n$ atributos ($m.n - 1$ transições). Os atributos são definidos percorrendo os *pixels* na sequência indicada na Figura 1 (embora outras sequências possam ser experimentadas). Em seguida, busca-se determinar o comportamento das sequências dos *pixels* em cada classe e, então, este conhecimento pode ser usado para classificar as poligonais (coordenadas paralelas (Inselberg, 1985) – Figura 2) de outros registros, para os quais não se conhece a classe.

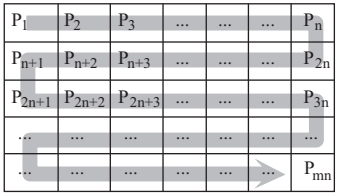
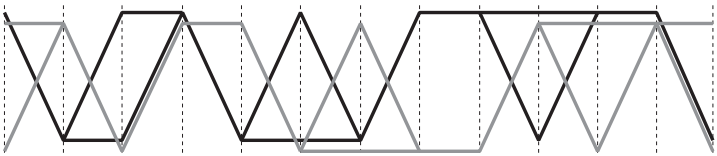


Figura 1. Sequência de *pixels* na imagem usada.

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 os comportamentos das poligonais de quatro caracteres em duas classes (dois em cada classe) são comparados. Em (a) tem-se a visualização em coordenadas paralelas do conjunto de dados apresentado em (b), destacando as transições. Em (c) observa-se que, entre os atributos a_1 e a_2, os caracteres da classe 1 possuem apenas a transição 10, não ocorrendo as transições 01, 11 e 00. Quanto à classe 2, em (d), nota-se que não ocorrem as transições 10 e 00, para esses atributos. Assim, para cada classe são determinadas todas as transições que

(a)



(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

(c)

(d)

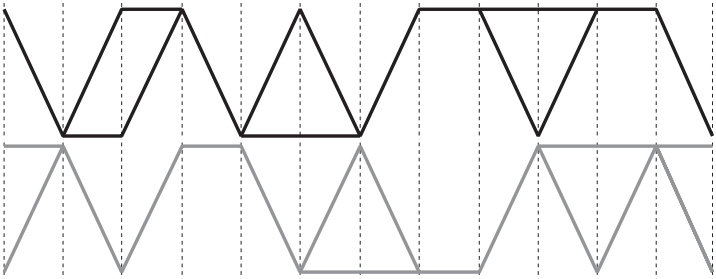


Figura 2. a) Exibição dos registros da classe 1 em preto e da classe 2 em cinza; b) conjunto de dados contendo 4 registros com treze atributos em duas classes; c) Exibição dos registros da classe 1; d) classe 2.

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.

Em seguida, a classificação dos caracteres pode então ser conduzida usando as três regras de classificação propostas:

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.

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.

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

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	10	00	00	10	01	00	00	00	00	00	00
	01	11	10	01	11	11	10	01	01	10	01	01
	11			11			11	10			10	11
Classe 2	00	00	00	00	00	10	01	01	00	00	00	00
	10	01	10	01	01	11	11	10	10	01	10	01
		11	11	10	11			11	11			

quantidades iguais de restrições, ou seja, não ocorrem as trinta transições na classe 1 e as trinta transições na classe 2, indicadas na Tabela 1. 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.

$$T = \sum_{i=1}^{n-1} c_i \cdot c_{i+1} \tag{1}$$

3.1 Alternativas para 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 ou mais *pixels* no lugar das transições entre dois *pixels*, propostas inicialmente. Assim, para três vizinhanças, as transições a serem verificadas seriam: 000, 001, 010, 011, 100, 101, 110 e 111. Com estas duas estratégias, aumenta-se a quantidade de informações/restrições usadas e, consequentemente, aumenta-se as chances de se classificar os caracteres corretamente. De fato, usando uma quantidade j de vizinhanças, se resolve ambiguidades que ocorrem quando se usa uma quantidade $j-1$ de vizinhanças. Isto é ilustrado na Figura 3, que mostra as ambiguidades que ocorrem usando transições entre dois *pixels* sendo resolvidas, usando três *pixels*. Nesta figura é possível observar que usando apenas as transições entre dois *pixels*,

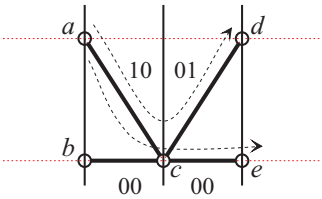


Figura 3. Ambiguidades que ocorrem usando transições entre dois *pixels*, resolvidas usando três *pixels*.

não se sabe se a poligonal que começa em a, vai para c e depois para d, ou se começa em a, vai para c e depois para e. O mesmo vale para a poligonal que começa em b (b,c,e ou b,c,d ?).

Nesta figura, a ambiguidade surge porque existem quatro possibilidades para as duas poligonais que iniciam em a e b, e terminam em d e e. Conforme se aumenta o número de vizinhanças nas transições, o classificador elimina tais ambiguidades, porém, se torna menos flexível. Assim, quando se considera as transições entre apenas dois *pixels*, o classificador pode inserir em uma classe, registros apenas parecidos com os usados no seu treinamento. Porém, se forem usadas as transições entre todas as vizinhanças possíveis, o classificador somente será capaz de classificar os objetos idênticos aos usados no seu treinamento.

4. Experimentos

Esta seção apresenta dois experimentos aplicando a técnica proposta neste trabalho. No primeiro é usado um conjunto de caracteres manuscritos, que tem sido amplamente utilizado para o teste de classificação, por causa da sua complexidade. O segundo experimento utiliza um conjunto contendo caracteres impressos, usando diferentes fontes.

4.1 Experimento 1

Nesse experimento é apresentada uma análise do conjunto de dados *binaryalphadigs* (Frank & Asuncion, 2010), usando a estratégia proposta. Este conjunto de dados é formado por 390 registros, obtidos a partir das imagens de 39 exemplos dos algarismos 0, 1, ..., 9, escritos à mão, conforme ilustra a Figura 4.

No experimento realizado com este conjunto, cada caractere foi reamostrado usando uma grade com uma resolução de 16×20 *pixels*, usando apenas as cores preto e branco (imagens binárias). Assim, cada caractere é representado neste conjunto através de 320 atributos (*pixels*) que podem assumir os valores zero (preto) ou um (branco).

A Figura 5 mostra as transições entre os 320 atributos usando coordenadas paralelas. A visualização dos registros em algumas de suas classes

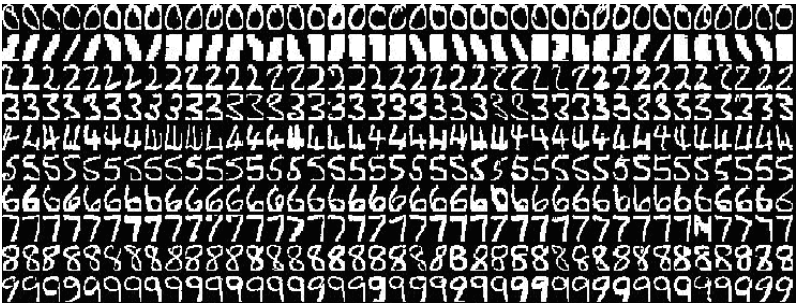


Figura 4. Conjunto de dados *binaryalphadigs* (Frank & Asuncion, 2010), contendo trinta e nove caracteres (números: 0, 1, ..., 9) escritos à mão.

revela um comportamento particular para cada classe, evidenciando diferentes conjuntos de restrições para cada classe.

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.

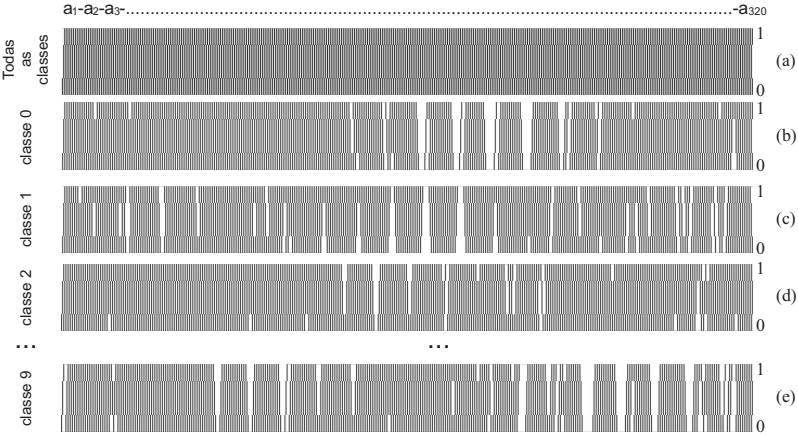


Figura 5. Visualização obtida usando: a) todas as transições de todos os caracteres nas dez classes; b) apenas os registros da classe do caractere "0"; c) apenas os registros da classe do caractere "1"; d) apenas os registros da classe do caractere "2"; e) apenas os registros da classe do caractere "9".

A matriz de confusão apresentada na Figura 6 ilustra a eficácia da estratégia para classificar os caracteres de acordo com as regras de classificação propostas, ou seja, o caractere é inserido na classe cujas transições não são violadas pelas transições impostas pela classe. Nos casos em que se encontra mais de uma classe satisfazendo esta condição, o caractere é inserido na classe menos violada 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 23^o caractere “5” na Figura 4, que atende todas as restrições das classes 3 e 5 e, pela regra R2 (regras de classificação propostas), foi reconhecido como da classe 3, porque ela possui um número maior de restrições que a classe 5.

Conforme apontado anteriormente, uma solução imediata para este problema é aumentar o tamanho da malha em que a imagem é amostrada, 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). Nos experimentos realizados, os caracteres classificados erroneamente são delimitados por retângulos.

Na Figura 8, tem-se o resultado já analisado na matriz de confusão da Figura 6, que usa a malha 20×16, e classifica um caractere “5”, na classe dos caracteres “3”.

Na Figura 9, amostrando os caracteres sobre uma malha com resolução 40×32, tem-se um resultado ótimo, 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.

Em seguida, utilizando transições entre três *pixels*, tem-se o seguinte resultado para a malha 10×8, ilustrado na Figura 10, onde se verifica

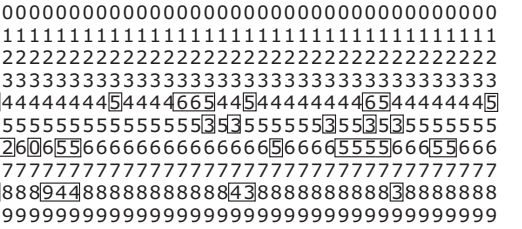


Figura 7. Resultado ruim da classificação usando uma malha 10×8 com transições entre 2 *pixels*.

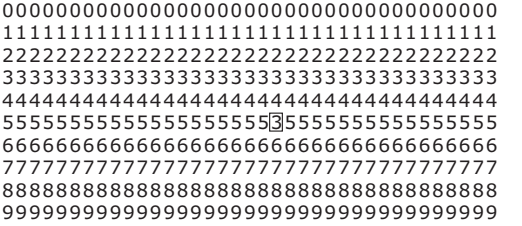


Figura 8. Resultado razoável da classificação usando uma malha 20×16 com transições entre 2 *pixels*.

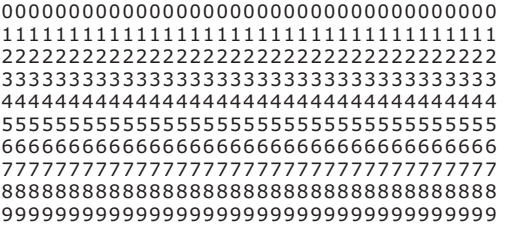


Figura 9. Resultado ótimo da classificação usando uma malha 40×32 com transições entre 2 *pixels*.

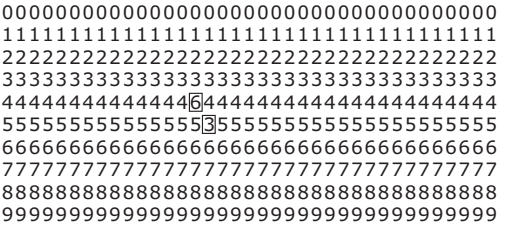


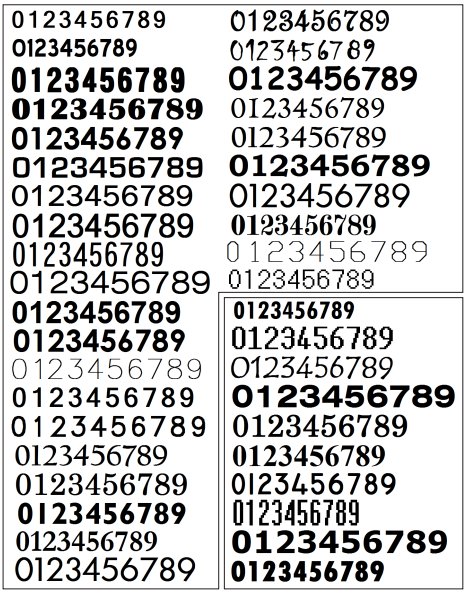
Figura 10. Resultado melhor para a classificação usando uma malha 10×8 com transições entre 3 *pixels*.

uma quantidade de erros bem menor, se comparada com o resultado da Figura 7, que também usa a malha 10×8, porém, usando transições entre dois *pixels*.

Para as malhas 20×16 (Figura 8), o uso das transições entre três *pixels* elimina completamente os erros de classificação. Do mesmo modo, o uso de uma vizinhança de quatro *pixels* para definir as transições, também elimina todos os erros, mesmo com a resolução baixa de 10×8.

4.2 Experimento 2

Nesse experimento é apresentada uma análise da classificação usando o conjunto de imagens de caracteres exibido na Figura 11. Neste experimento, o treinamento do classificador foi realizado usando apenas os caracteres em (a), enquanto que o teste foi feito usando apenas os caracteres em (b), observando que se trata de um conjunto 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, em (b), usando as transições entre três *pixels*, foi obtida a classificação correta para todos os caracteres, enquanto que em (c), usando as transições entre quatro *pixels*, também foi obtida a classificação sem erros para todos os caracteres.

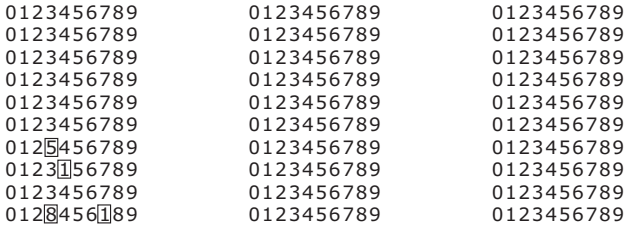


Figura 12. Resultado da classificação usando uma malha 40×32 com transições a) dois *pixels*; b) três pixels; c) quatro *pixels*.

O experimento a seguir ilustra o uso da proposta apresentada neste trabalho, com caracteres alfabéticos (a..z e A..Z), além dos dígitos (0..9). A Figura 13(a) mostra o conjunto de caracteres usados no treinamento deste classificador, que no caso corresponde aos caracteres alfanuméricos usando a fonte *Times New Roman*.

Neste caso, como existe uma diferença entre os tamanhos dos caracteres maiúsculos e minúsculos, deve-se realizar o recorte dos caracteres com um tamanho adequado, obedecendo a sua altura, tanto na etapa de treinamento, quanto na etapa de classificação. Isto é ilustrado na Figura 14, que mostra como os recortes dos caracteres "a" e "A" devem ser feitos. De fato, não existe dificuldade alguma neste processo, pois todos os caracteres são enquadrados automaticamente dentro de retângulos, todos com a mesma altura, independentemente de se tratar de um caractere maiúsculo ou minúsculo.

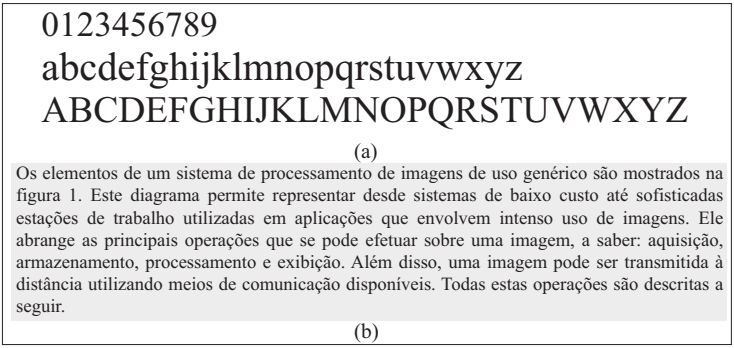


Figura 13. a) Caracteres alfanuméricos usando a fonte *Times New Roman*, usados no treinamento do classificador; b) Imagem de um parágrafo digitalizado do livro de (Marques Filho & Vieira Neto, 1999, página 2).

Figura 14. Enquadramento diferenciado para caracteres maiúsculos e minúsculos.

A Figura 15 apresenta o texto extraído da imagem na Figura 13(b), com esta classificação. Neste caso, pode-se observar apenas alguns erros: o caractere "1" (um) foi reconhecido como a letra "l" (L), por causa da grande semelhança entre estes caracteres usando a fonte *Times New Roman*; caracteres acentuados, devido a sua não inclusão no conjunto de treinamento do classificador (Figura 13(a)); pontuações (vígulas, pontos finais) também não foram reconhecidos pelo mesmo motivo.

Os elementos de um sistema de processamento de imagens de uso gen□rico s□o mostrados na figura 1 Este diagrama permite representar desde sistemas de baixo custo at□ sofisticadas esta□□es de trabalho utilizadas em aplica□□es que envolvem intenso uso de imagens Ele abrange as principais opera□□es que se pode efetuar sobre uma imagem a saber aquisi□□o armazenamento processamento e exibi□□o Al□m disso uma imagem pode ser transmitida a dist□ncia utilizando meios de comunica□□o dispon□veis Todas estas opera□□es s□o descritas a seguir

Figura 15. Texto extraído da imagem apresentada na Figura 14(b).

Em seguida é apresentada uma comparação entre alguns programas de OCR, que podem ser encontrados na Internet. Neste estudo são comparados a qualidade da classificação, o que é feito através de uma contagem dos erros de classificação dos caracteres. Também é feita uma comparação entre os tempos de execução. Os programas usados são:

- FreeOCR.net - Trata-se de um programa OCR que inclui o núcleo *Tesseract free OCR*, que pode ser usado com os *drivers* Twain e WIA. Este núcleo foi desenvolvido pela Hewlett Packard entre 1985 e 1995. Atualmente está disponível sob a forma *open-source*, mantido pela Google Inc. Este programa pode ser encontrado em: http://www.freeocr.net;
- SimpleOCR 3.1 - Este programa foi desenvolvido pela Simple-Software e é distribuído na modalidade *freeware*, podendo ser usado

por usuários domésticos, instituições educacionais e também usuários corporativos. Este programa pode ser encontrado no *website* http://www.simpleocr.com;

- A-PDF-OCR - Trata-se de um programa comercial, desenvolvido pela APDF, que pode ser encontrado em http://www.a-pdf.com;
- Cuneiform Pro OCR 6.0 - Este programa foi desenvolvido pela Give Me Freeware e usa alfabetos para vinte idiomas diferentes, incluindo o Português. Pode ser encontrado no *website* http://freeware.odlican.net;
- Image2pdf OCR 3.2 - Trata-se de um programa *freeware*, desenvolvido pela SoftSolutions, que pode ser obtido em http://products.softsolutionslimited.com;
- ABBYY FineReader 11 Professional - Trata-se de um programa comercial, desenvolvido pela ABBYY USA Software House, que pode ser encontrado em http://www.abbyy.com.

A Figura 16 mostra o resultado da comparação da qualidade destes seis programas e também do algoritmo apresentado neste trabalho. A Figura 17 mostra o resultado da comparação dos tempos de execução destes seis programas e também do algoritmo apresentado neste trabalho. A imagem utilizada para a classificação contém 3718 caracteres.

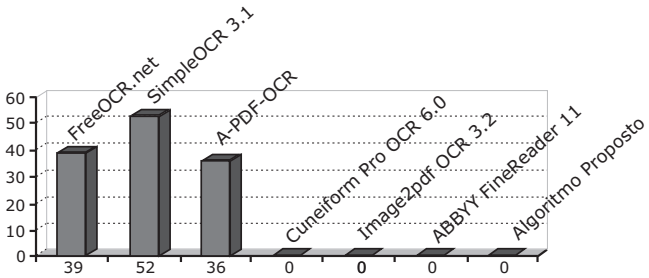


Figura 16. Contagem de erro de classificação.

4.3 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, três e quatro *pixels*, com diferentes resoluções de malha. A máquina usada possui um processador Intel Core i3 M330 de 2.13GHz e 4GB de RAM.

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, três *pixels* e quatro *pixels*, com a resolução de malha de 40×32.

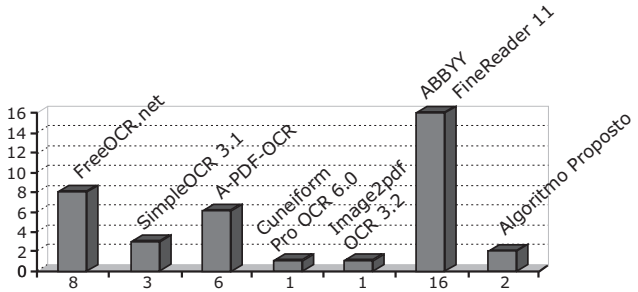


Figura 17. Tempos de excecução (segundos).

Estes resultados mostram que o algoritmo proposto consegue identificar caracteres em uma taxa média de 316,19 caracteres por segundo, usando a transição de 4 *pixels* e 923,36 caracteres por segundo, usando a transição de 3 *pixels*, com a malha 40×32 (melhores resultados). Usando a transição de 2 *pixels*, com uma malha 10×8, 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.

A Figura 18 mostra como o tempo de processamento e a quantidade de erros obtidos estão relacionados com as quantidades de *pixels* usadas nas transições. Foram utilizados os dados obtidos com o experimento usando o conjunto de dados *binaryalphadigs* (malha 10×8) para construir os gráficos.

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" (Figura 4), por exemplo, apresenta uma grande variação, que tem desafiado a maioria dos algoritmos conhecidos de reconhecimento de caracteres. A descrição dos caracteres a partir do comportamento das tran-

Tabela 2. Tempos médios de processamento (segundos) para um caractere usando o algoritmo proposto (conjunto de dados *binaryalphadigs* – Figura 4).

	Resoluções		
Transições	10×8	20×16	40×32
2 *pixels*	0,0000344	0,0001511	0,0006261
3 *pixels*	0,0000782	0,0002851	0,0010766
4 *pixels*	0,0001614	0,0005958	0,0031556

Tabela 3. Tempos médios de processamento (em segundos) para um caractere usando o algoritmo proposto (caracteres da Figura 11).

	Resolução
Transições	40×32
2 *pixels*	0,000683
3 *pixels*	0,001083
4 *pixels*	0,001849

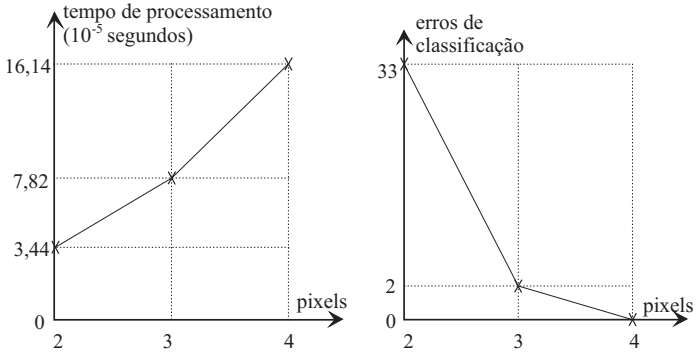


Figura 18. a) Tempos de processamento × quantidade de *pixels* usadas nas transições; b) Erros de classificação × quantidade de *pixels* usadas nas transições.

sições de níveis entre os *pixels* vizinhos se mostrou uma estratégia muito simples, rápida e eficiente, que pode ser facilmente implementada em hardware e usada em aplicações em tempo real. Este é o caso dos sistemas de visão computacional que obtêm suas imagens a partir de câmeras de vídeo, como robôs autônomos e veículos rápidos não tripulados. 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. Observando que um aumento excessivo na quantidade de transições gera uma perda da capacidade de generalização da classificação. 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.

Referências

Dimond, T.L., Devices for reading handwritten characters. In: *Proceedings of Eastern Joint Computer Conference*. p. 232–237, 1957.

Duda, R.O.; Hart, P.E. & Stork, D.G., *Pattern Classification*. 2a edição. New York, USA: John Wiley & Sons, 2001.

Eden, M., On the formalization of handwriting. In: *Proceedings of the Fourth London Symposium on Information Theory*. 1961.

Eden, M. & Halle, M., The characterization of cursive writing. In: *Proceedings of the Fourth London Symposium on Information Theory*. p. 287–299, 1961.

Frank, A. & Asuncion, A., UCI machine learning repository. 2010. http://archive.ics.uci.edu/ml/.

Frishkopf, L.S. & Harmon, L.D., Machine reading of cursive script. In: *Proceedings of the Fourth London Symposium on Information Theory*. p. 300–316, 1961.

Gonzalez, R.C. & Woods, R.E., *Digital Image Processing*. 2a edição. Reading, USA: Addison-Wesley, 2001.

Handel, P.W., *Statistical Machine*. 1933. U.S. Patent 1 915 993.

Inselberg, A., The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.

Jain, A.K.; Duin, R.P.W. & Mao, J., Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

Jain, G. & Ko, J., *Handwritten Digits Recognition*. Multimedia Systems Project Report, University of Toronto, 2008.

Lin, H.; Ou, W. & Zhu, T., The research of algorithm for handwritten character recognition in correcting assignment system. In: *Proceedings of the Sixth International Conference on Image and Graphics*. p. 456–460, 2011.

Marques Filho, O. & Vieira Neto, H., *Processamento Digital de Imagens*. Rio de Janeiro, RJ: Brasport, 1999.

Montaña, E.G., Digits recognition via neural networks. 2007. http://ociotec.com/wp-content/uploads/2007/07/DigitsRecognition.pdf.

Neisser, U. & Weene, P., A note on human recognition of hand-print characters. *Information and Control*, 3(2):191–196, 1960.

Pavlidis, T., *Structural Pattern Recognition*. Heidelberg, Germany: Springer-Verlag, 1980.

Pereira, J.F.; Alves, V.M.O.; Cavalcanti, G.D.C. & Ren, T.I., Modular image principal component analysis for handwritten digits recognition. In: *Proceedings of the Seventeenth International Conference on Systems, Signals and Image Processing*. p. 356–359, 2010.

Schalkoff, R.J., *Pattern Recognition: Statistical, Structural and Neural Approaches*. New York, USA: John Wiley & Sons, 1992.

Shrivastava, S.K. & Gharde, S.S., Support vector machine for handwritten devanagari numeral recognition. *International Journal of Computer Applications*, 7(11):9–14, 2010.

Tauschek, G., *Reading Machine*. 1935. U.S. Patent 2 026 329.

Trentini, V.B.; Godoy, L.A.T. & Marana, A.N., Reconhecimento automático de placas de veículos. In: *Anais do VI Workshop de Visão Computacional*. p. 267–272, 2010.

Zhu, X.; Shi, Y. & MA, S., Research on handwritten character recognition. *Pattern Recognition and Artificial Intelligence*, 13(2):172–180, 2000.

Notas Biográficas

Francisco Assis da Silva possui graduação em Ciência da Computação (Universidade do Oeste Paulista – UNOESTE, 1998), mestrado em Ciência da Computação (Universidade Federal do Rio Grande do Sul – UFRGS, 2002) e doutorado em Engenharia Elétrica na área de Visão Computacional (Universidade de São Paulo – USP/São Carlos, 2012). Atualmente é docente da UNOESTE/Presidente Prudente.

Almir Olivette Artero possui graduação em Matemática (Universidade Estadual Paulista – UNESP, 1990), especialização em Sistemas de Informação (Universidade do Oeste Paulista – UNOESTE, 1995), mestrado em Ciências Cartográficas (UNESP, 1999) e doutorado em Ciência da Computação (Universidade de São Paulo – USP, 2005). Atualmente é docente da UNESP/Presidente Prudente).

Maria Stela Veludo de Paiva possui graduação em Engenharia Elétrica/Eletrônica (Universidade de São Paulo – USP, 1979), mestrado e doutorado em Física Aplicada (USP/São Carlos, 1984 e 1990), tendo realizado Pós-Doutorado na University of Southampton (1992). Atualmente é docente no Departamento de Engenharia Elétrica da Escola de Engenharia de São Carlos (USP) e desenvolve pesquisas na área de Visão Computacional.

Ricardo Luís Barbosa possui graduação e especialização em Matemática (Universidade Estadual Paulista – UNESP, 1990 e 1993), mestrado e doutorado em Ciências Cartográficas (UNESP, 1999 e 2006). Atualmente é docente da UNESP/Sorocaba e desenvolve pesquisas na empresa Cartovias Engenharia Cartográfica.

Capítulo 15

Localização, Segmentação e Reconhecimento de Caracteres em Placas de Automóveis

Leonardo Augusto Oliveira e Adilson Gonzaga*

Resumo: O objetivo deste capítulo é propor uma técnica de baixa complexidade computacional para localização de placas de automóveis em imagens de cenas reais, segmentação desta placa e extração dos caracteres para reconhecimento. A etapa de Localização é baseada na busca de regiões da imagem de maior concentração de mudanças bruscas na intensidade. A etapa de Segmentação consiste no processamento da imagem para obter uma imagem binária sem ruídos e com objetos bem definidos. A etapa de Reconhecimento consiste em examinar cada objeto, em comparação com um conjunto de máscaras que indicam a sua identificação. Cada módulo obteve taxa de precisão de cerca de 90%.

Palavras-chave: Processamento de imagens, LPR, Reconhecimento de placas.

Abstract: *The purpose of this chapter is to propose a low computational complexity technique to locate license plates in images of real scenes, in order to segment and extract the characters for recognition. The Localization step is based on searching image regions of highest concentration of abrupt changes of intensity. The Segmentation step consists in processing the image in order to obtain a binary image without noise and well defined objects. The Recognition step consists in examining each object in comparison with a set of masks that indicate his identification. Each single module achieved an accuracy rate of about 90%.*

Keywords: *Image processing, LPR, Plate recognition.*

*Autor para contato: agonzaga@sc.usp.br

1. Introdução

O crescente interesse pela automatização de processos e introdução de novas tecnologias são fatores motivadores para diversas pesquisas no ramo de reconhecimento de padrões em imagens. 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 (Anagnostopoulos et al., 2008).

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 (Gonzalez & Woods, 2008). Atualmente, métodos computacionais inteligentes permitem a abordagem do problema utilizando lógica *fuzzy*, redes neurais artificiais e algoritmos genéticos (de Campos, 2001).

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 abordagem proposta, descrita na Seção 3. Na Seção 4 são apresentados os resultados obtidos com o algoritmo desenvolvido. Na Seção 5 conclui-se sobre a metotodologia proposta.

2. Técnicas de LPR

São três etapas básicas no processo de reconhecimento de uma placa de automóvel: **(a)** Localização e segmentação da placa, **(b)** Segmentação dos caracteres e **(c)** Reconhecimento dos caracteres.

2.1 Localização

A Localização e segmentação da placa consiste em realizar um rastreamento em toda a imagem a fim de identificar a exata região que contém somente a placa do automóvel e segmentá-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 et al. (2004) propõem um algoritmo "baseado em operações morfológicas multi-estágio". A busca por candidatos a 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 obtidos pelos autores 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.

Anagnostopoulos et al. (2006) 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. Discutem o uso da transformada de Hough (HT, do inglês *Hough Transform*), afirmando que ela exige muito esforço computacional. É proposta a combinação de duas técnicas: aplicação de algoritmos de detecção de bordas aliados à HT. O uso daquela técnica torna esta menos dispendiosa computacionalmente.

Em seguida, os autores propõem um método de análise estatística. A técnica "Janelas Concêntricas Deslizantes" (do inglês *Sliding Concentric Windows*) 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.

Khalifa et al. (2006) calculam a projeção horizontal (vetor com soma dos valores de intensidade dos píxeis em cada linha) e os picos do gráfico indicam a provável posição vertical da placa (a altura em que ela se encontra na imagem). Janelas deslizantes buscam pelas regiões de maior densidade de bordas em projeções verticais. Desta forma, tem-se a posição vertical e horizontal da placa na imagem. Os resultados obtidos para localização das placas tiveram precisão de 92,1%.

Martinsky (2007) 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 quatro 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 dos caracteres

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 et al. (2004) realizam o agrupamento de pixeis pela rotulagem 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, aplica-se a HT aos 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 al-

goritmo complexo, mas com taxa de 95,6% de sucesso para a identificação de caracteres com uso de imagens de entrada de alta complexidade.

Conci & Monteiro (2004) fazem uma análise mais simples: após a rotulagem dos píxeis e obtenção dos objetos contidos na imagem, estimam-se os limites superior e inferior dos caracteres na placa. Desta forma, os objetos que não obedeçam a esse critério são eliminados e apenas os 7 caracteres interessantes restarão.

Anagnostopoulos et al. (2006) propõe um método interessante de localização de cadidatos a placas: SCW, ou "Janelas Concêntricas Deslizantes", já apresentado na seção 2.1. Aos candidatos a placa, aplica-se um método de binarização utilizando um valor de limiar localmente adaptivo chamado de método de Sauvola, que calcula um valor de limiar para cada pixel considerando média e variância locais. O objetivo é eliminar problemas relacionados a iluminação não-homogênea. Em seguida, os autores propõem o método de Análise de Componentes Conectados (CCA, do ingês *Connected Component Analysis*), que nada mais é do que o método de rotulagem. Com os píxeis rotulados e agrupados, aplica-se um algoritmo de seleção dos objetos de interesse. Os autores provocam, propositalmente, que a placa seja um único objeto com diversos orifícios (os caracteres), para então fazer a seguinte análise: o objeto desejado deve ter orientação horizontal (inclinação menor do que 35 graus), *aspect ratio* entre 2 e 6 e mais do que 3 orifícios presentes na imagem. Considera-se, ainda, o caso de a binarização resultar em caracteres como objetos e a placa como plano de fundo. Desta forma, a condição do número de orifícios será desobedecida, a figura terá seus valores de píxeis invertidos e o processo de seleção será realizado novamente.

Com a imagem invertida, tem-se agora os caracteres como objetos. De forma semelhante à realizada anteriormente, aqueles objetos que não satisfizerem condições de orientação angular e altura são eliminados e o processo de segmentação é finalizado. Os autores indicam taxa de sucesso de 96,5%, sendo que utilizaram um banco de dados extenso (1334 entradas) e composto de imagens de características complexas.

Draghici (2007) faz inicialmente uma projeção horizontal de uma imagem binária, de modo a identificar grupos de objetos que estejam em uma mesma reta horizontal. Neste grupo, é feita a projeção vertical da imagem para identificar cada objeto separadamente. Feita a segmentação dos caracteres, os resultados são validados. Caso um erro seja reportado, o algoritmo retorna ao ponto de binarização da imagem, buscando outro valor de limiar e repetindo o processo (projeção horizontal $\rightarrow$ projeção vertical $\rightarrow$ segmentação) até que sejam encontrados objetos que sejam validados como caracteres de uma placa de automóvel.

2.3 Reconhecimento

O Reconhecimento de caracteres é a etapa em que um objeto segmentado 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 et al. (1999). 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 por meio de um limite máximo percentual de pontos diferentes. Os autores não especificam um valor, mas assinalam uma alta taxa de acerto do sistema.

Conci & Monteiro (2004) 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.

Sancho (2006) utiliza a técnica de correlação-cruzada (*cross-correlation*). Basicamente, a imagem de um caractere $f(x, y)$ é comparada a uma máscara padrão $w(x, y)$ usando correlação 2D.

O valor de saída da função de correlação-cruzada será maior de acordo com a maior semelhança entre a máscara e o objeto analisado. O autor admite que símbolos semelhantes (por exemplo, as letras **O** e **D**) não podem ser avaliados com segurança pelo método proposto. Para este problema, aplica-se um Solucionador de Problemas com Restrições Otimizado (COPS, do inglês *Constrained Optimization Problem Solver*). O COPS têm o objetivo de atingir os seguintes objetivos: **(a)** os candidatos devem ser placas de automóveis válidas na Espanha, **(b)** apenas os três símbolos com maior valor de correlação-cruzada são considerados, **(c)** sequências maiores de caracteres têm precedência em relação às menores e **(d)** a soma dos valores de correlação-cruzada é considerada. O autor indica uma taxa de 90% de acerto com menos de um segundo de tempo de processamento.

3. Material e Método

A abordagem proposta se baseia na morfologia e estatíticas das imagens para realizar a Localização e Segmentação das placas, e na Correlação 2D para o Reconhecimento. 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.

Para a implementação do algoritmo proposto neste trabalho, foram utilizadas imagens próprias e outras disponíveis na Internet. São cinco bancos de imagens: BD1, BD2, BD3, BD4 e BD5. 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 (LPDSI, 2011). As imagens são consideradas de má qualidade. Os bancos BD2 e BD3 são compostos por, respectivamente, 79 e 127 imagens, capturadas no estacionamento do câmpus de São Carlos da Universidade de São Paulo. 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. Os bancos BD4 e BD5 são compostos por 17 imagens cada um, também obtidas no estacionamento do câmpus de São Carlos da Universidade de São Paulo e de boa qualidade. Diferenciam-se dos outros pela posição de obtenção das fotos, de tal forma que estas estivessem inclinadas, gerando a necessidade de se trabalhar com correção de rotação nos algoritmos.

As Figuras 1a, 1b e 1c são exemplos de imagens utilizadas dos diferentes bancos. As placas foram parcialmente cobertas apenas para fins de divulgação das imagens.

(a) Imagem do BD1

(b) Imagem do BD2

(c) Imagem do BD4

Figura 1. Exemplos de imagens utilizadas neste trabalho.

Para obtenção das imagens dos bancos BD2, BD3, BD4 e BD5 foi utilizada uma câmera do modelo Sony Cyber-shot DSC-H10, em modo automático, sem flash e com resolução mínima (640×480 pixeis).

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 pixeis vizinhos na direção horizontal. Uma imagem resultante desta 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 pixeis 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 2 ilustra o procedimento.

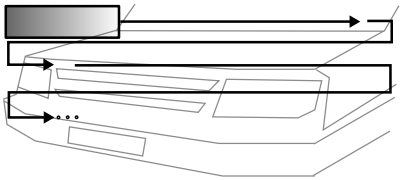


Figura 2. Varredura da imagem por janelas.

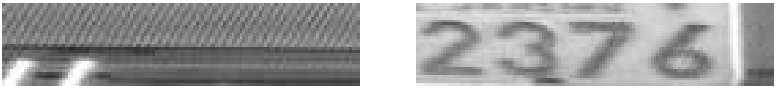
O processo desenvolvido exige uma filtragem de suavização para retirar ruído e eliminar a possibilidade de detecção de elementos indesejados. A grade frontal presente em alguns modelos de carros é um exemplo de elemento indesejado. Estes, porém, costumam apresentar frequência de mudança de intensidade maior do que os caracteres da placa. O filtro de suavização aliado à operação morfológica de abertura promovem a eliminação destas 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 seguida de abertura morfológica em escala de cinza, ambos 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 pixeis serão eliminados enquanto elementos do plano de fundo serão preenchidos. As Figuras 3a, 3b e 3c exemplificam o processamento da imagem (suavização e abertura morfológica) e as figuras seguintes mostram a Localização falha de uma região de interesse sem prévia filtragem (Figura 4a) e Localização com sucesso desta região após a filtragem proposta (Figura 4b).

(a) Imagem original (b) Suavização (c) Abertura

Figura 3. Processo de filtragem da imagem para Localização.



(a) Sem a filtragem (b) Com a filtragem

Figura 4. Resultados de Localização.

Para os casos de imagens dos bancos BD4 e BD5 é preciso considerar a possibilidade de corrigir a rotação da imagem, o que ocorre quando a fotografia é obtida de maneira inclinada. A correção da rotação será feita com o uso da HT (Transformada de Hough), de acordo com o seguinte algoritmo:

i) Aplicar detector de bordas Sobel na imagem;

ii) Aplicar a HT ($\rho_j = x_j \cos\theta_j + y_j \sin\theta_j$) em cada pixel da imagem;

iii) Obter uma matriz H cujos índices indicam o par (ρ_j,θ_j) e o valor de cada elemento indica a quantidade de pontos do plano xy residem na reta indicada em $\rho\theta$;

iv) Elementos grandes de H indicam as linhas retas da imagem, pois vários pontos residem naquele par (ρ_j,θ_j). Selecionar o maior elemento de H;

v) Encontrar o θ correspondente ao H_{max} no gráfico $\rho = x\cos\theta + y\sin\theta$;

vi) O valor do ângulo é tal que $0 \leq \theta \leq \pi$. Caso $\theta > \pi/2$, definir $\theta' = \theta - \pi$, que indica a mudança no sentido de rotação. Caso contrário, $\theta' = \theta$. Armazenar valor de θ';

vii) Rotacionar a imagem em θ' radianos.

Após esta correção de rotação, a imagem pode ser analisada nas etapas já descritas como se a fotografia tivesse sido obtida frontalmente à placa do automóvel.

3.2 Processamento da imagem da placa

Com a localização e segmentação 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 (Gonzalez & Woods, 2008). O algoritmo é composto por 5 passos:

i) Definir um valor estimado inicial T_0.

ii) Dividir os pixeis 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 pixeis 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 pixeis 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 atribuídos valores a uma variável em cada teste. Quanto mais distante do valor esperado o objeto estiver, maior será o valor desta 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* (razão de largura por 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, considerando a 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 esta 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. Desta forma, esta 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 $X = 80$.
- **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* (AR) e testar os objetos quanto a esta 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 pela equação 1:

$$H_C = \begin{cases} X\frac{|AR_{min} - AR|}{AR} & \text{para } AR < AR_{min} \\ X\frac{|AR_{max} - AR|}{AR} & \text{para } AR > AR_{max} \\ 0 & \text{caso contrário} \end{cases} \quad (1)$$

  na qual H_C é o valor atribuído ao teste e X é o peso vinculado a este valor, AR, AR_{min} e AR_{max} são os valores de *aspect ratio* do objeto, mínimo e máximo esperados, respectivamente.

  O valor do peso X é utilizado para maximizar os valores daqueles objetos que apresentem *aspect ratio* muito distante do esperado. Estabeleceu-se:

$$X = \begin{cases} 60 & \text{para } AR \leq \frac{AR_{min}}{1,5} \\ 20 & \text{para } \frac{AR_{min}}{1,5} < AR \leq \frac{AR_{min}}{1,2} \\ 10 & \text{para } \frac{AR_{min}}{1,2} < AR \leq AR_{min} \\ 90 & \text{para } AR \geq 1,5AR_{max} \\ 30 & \text{para } 1,2AR_{max} \geq AR < 1,5AR_{max} \\ 15 & \text{para } AR_{max} \geq AR < 1,2AR_{max} \end{cases} \quad (2)$$

- **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:

$$Área = AspectRatio \cdot AlturaMediana^2 \quad (3)$$

Portanto, pode-se considerar $A_{min} = AR_{min} \cdot AM^2$ e $A_{max} = AR_{max} \cdot AM^2$. Assim, H_D é calculada por:

$$H_D = \begin{cases} X\frac{|A_{min}-A|}{A} & \text{para } A < A_{min} \\ X\frac{|A_{max}-A|}{A} & \text{para } A > A_{max} \\ 0 & \text{caso contrário} \end{cases} \quad (4)$$

na qual H_D é o valor atribuído ao teste e X é o peso vinculado a este valor, A, A_{min} e A_{max} são os valores de área do objeto, mínimo e máximo esperados, respectivamente, calculados conforme a equação 3.

Neste teste, estabeleceu-se, empiricamente:

$$X = \begin{cases} 70 & \text{para } A \leq \frac{A_{min}}{2,0} \\ 10 & \text{para } \frac{A_{min}}{2,0} < A \leq \frac{A_{min}}{1,5} \\ 5 & \text{para } \frac{A_{min}}{1,5} < A \leq A_{min} \\ & \\ 70 & \text{para } A \geq 2,0A_{max} \\ 10 & \text{para } 2,0A_{max} > A \geq 1,5A_{max} \\ 5 & \text{para } 1,5A_{max} > A \geq A_{max} \end{cases} \quad (5)$$

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.

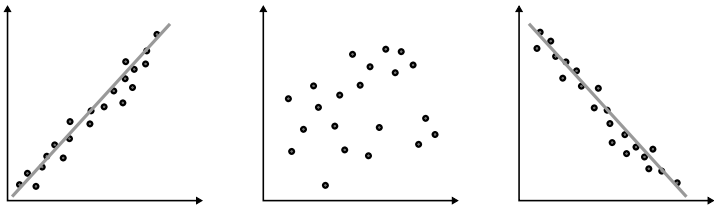
A comparação entre o caractere e as máscaras é feita por Correlação 2D, por meio do Coeficiente de Correlação de Pearson (CCP), uma medida estatística para comparação linear entre dois conjuntos de dados. Seu valor, em módulo, varia de 0 a 1. Quanto mais perto de 1 maior é o grau de dependência estatística linear entre as variáveis. No outro oposto, quanto

mais próximo de zero, menor é a força desta relação (Figueiredo Filho & Silva Júnior, 1999). A Equação 6 representa o CCP entre dois vetores.

$$\rho = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2} \tag{6}$$

na qual ρ é o CCP, x_i e y_i são os valores da posição i dos vetores x e y, e $\bar{x}$ e $\bar{y}$ representam a média dos valores dos vetores x e y, respectivamente.

Os gráficos das Figuras 5a, 5b e 5c são exemplos de correlações positiva, nula e negativa (respectivamente, da esquerda para a direita) entre dois vetores de valores. Nota-se que valores próximos de x e y levam a uma correlação positiva dos dados. É o caso de uma análise entre um caractere e sua máscara correspondente.



(a) Positiva (b) Nula (c) Negativa

Figura 5. Gráficos de Correlação.

O cálculo do CCP entre as imagens exige que elas tenham o mesmo tamanho e o resultado apresentado não é invariante a escala ou rotação. O desenvolvimento de um algoritmo que obedeça a essas condições de invariância é complexo e computacionalmente mais dispendioso. Neste trabalho houve um esforço de correção de rotação, inclinação e escala de tal forma que não se fez necessário considerar tais condições. A simples aplicação da correlação é suficiente para definir a melhor máscara para dado objeto.

A imagem do caractere desconhecido é comparada, por meio do cálculo do CCP, com as 72 máscaras. A máscara mais semelhante ao 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.

Apesar da má qualidade das imagens do BD1, foram obtidos ótimos resultados. A taxa de 97,3% alcançada neste trabalho é extremamente

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%
BD4	17	14	82,3%
BD5	17	13	76,5%

positiva, visto que foi utilizado um método simples e rápido. As taxas de sucesso dos bancos BD2 e BD3 são superiores a 91%.

As taxas de sucesso mais baixas dos bancos BD4 e BD5 são consequências de falhas do processo de correção de rotação, extremamente importante para essas imagens.

A Figura 6 mostra exemplos de resultados da etapa de Localização em todos os bancos de dados utilizados.

(a) Exemplo 1 - BD1 (b) Exemplo 2 - BD1

(c) Exemplo 1 - BD2 (d) Exemplo 2 - BD2

(e) Exemplo 1 - BD3 (f) Exemplo 2 - BD3

(g) Exemplo 1 - BD4 (h) Exemplo 2 - BD4

(i) Exemplo 1 - BD5 (j) Exemplo 2 - BD5

Figura 6. Exemplos de resultados de Localização.

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%
BD4	98	85	86,7%
BD5	91	78	85,7%

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.

Pode-se observar que as taxas de acerto dos bancos BD4 e BD5 se aproximam das taxas dos bancos BD2 e BD3, como esperado, já que os resultados negativos do processo de correção de rotação não têm influência nesta etapa.

A Figura 7 mostra exemplos de resultados da etapa de Segmentação em todos os bancos de dados utilizados. As imagens foram selecionadas em uma etapa intermediária para mostrar a presença de alguns elementos indesejados, principalmente nas laterais das figuras, que serão posteriormente retirados para o Reconhecimento dos caracteres. Os procedimentos mais complicados, no entanto, ocorreram até obter os resultados exemplificados nestas figuras.

A taxa de acerto na etapa de Reconhecimento é mostrada 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%
BD4	85	77	90,6%
BD5	78	70	89,7%

As altas taxas de acerto do algoritmo indicam, mais do que a qualidade do processo de reconhecimento, a qualidade das etapas anteriores ao reconhecimento. O método de Correlação 2D não é invariante a escala de rotação e altamente sensível a pequenos desvios da reta indicativa da correlação. Figueiredo Filho & Silva Júnior (1999) afirmam que o CCP é fortemente afetado pela presença de *outliers*, que são pontos aleatórios fora da curva esperada. Veja o exemplo de um *outlier* na Figura 8. O segmento

(a) Exemplo 1 - BD1

(b) Exemplo 2 - BD1

(c) Exemplo 1 - BD2

(d) Exemplo 2 - BD2

(e) Exemplo 1 - BD3

(f) Exemplo 2 - BD3

(g) Exemplo 1 - BD4

(h) Exemplo 2 - BD4

(i) Exemplo 1 - BD5

(j) Exemplo 2 - BD5

Figura 7. Exemplos de resultados de Segmentação.

sólido indica a provável reta obtida com a influência do *outlier* e o segmento tracejado indica a provável reta sem a sua influência. Analisando graficamente, diz-se que quanto maior é a distância da reta em relação às regiões mais densas de pontos, menor é o coeficiente de correlação entre os eles.

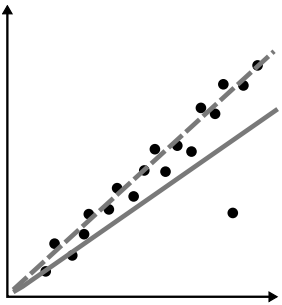


Figura 8. Representação simples de um outlier: o ponto fora da curva

Portanto, um caractere pode ter seu Reconhecimento prejudicado por mais que ocorram apenas pequenos desvios em relação à sua máscara correspondente. As altas taxas de acerto do processo de reconhecimento indicam, portanto, que as etapas de Localização e Segmentação foram bem executadas (salvo os problemas ocorridos com a correção de rotação).

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%
BD4	119	77	64,7%
BD5	119	70	58,8%
Todos	**2205**	**1754**	**79,5%**

Para se analisar os tempos de processamento do programa desenvolvido com a metodologia proposta, realizou-se cinco execuções do mesmo código e foram medidos os tempos de execução. Os resultados estão na Tabela 5.

Tabela 5. Tempos de execução, em segundos.

Etapas	A	B	C	D	Todas
	0,628	1,413	0,067	0,178	
	0,681	1,358	0,038	0,148	
	0,644	1,442	0,030	0,146	
	0,615	1,408	0,031	0,134	
	0,634	1,343	0,029	0,141	
$t_{médio}$	**0,641**	**1,393**	**0,039**	**0,149**	**2,222**

Na Tabela 5, a etapa **A** corresponde ao carregamento da imagem, a etapa **B** à Localização, a etapa **C** à Segmentação e a etapa **D** ao Reconhecimento. O tempo médio total para carregar a imagem, localizar a ROI, segmentar os objetos e reconhecer os caracteres de uma placa em uma imagem foi de aproximadamente 2,2 segundos.

5. Conclusão

A taxa de acerto de cerca 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 & Woods (2008) afirmam que o reconhecimento com sucesso será fortemente influenciado por uma segmentação precisa. Acredita-se 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 melhores, 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.

Os resultados dos bancos BD4 e BD5 merecem destaque negativo, mas com ressalvas. Apesar do baixo desempenho global, percebe-se que o desempenho abaixo dos demais bancos se deve unicamente a problemas de correção de rotação, identificados na etapa de Localização. A saída muitas vezes encontradas em sistemas práticos de reconhecimento automático de placas de automóveis é evitar tais imagens, buscando obter placas de forma que não necessitem de rotação ou que tenham um ângulo de inclinação conhecido.

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.

6. Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo suporte financeiro a esta pesquisa.

Referências

Anagnostopoulos, C.N.; Anagnostopoulos, I.; Loumos, V. & Kayafas, E., A license plate recognition algorithm for intelligent transportation system applications. *IEEE Transactions on Intelligent Transportation Systems*, 7(3):377–392, 2006.

Anagnostopoulos, C.N.; Anagnostopoulos, I.; Psoroulas, I.D.; Loumos, V. & Kayafas, E., License plate recognition from still images and video sequences: a survey. *IEEE Transactions on Intelligent Transportation Systems*, 9(3):377–391, 2008.

Belvisi, R.; Freitas, R.; Figueiredo, R. & Marcovitch, D., Um sistema de reconhecimento automático de placas de automóveis. In: *Anais*

do XIX Congresso Nacional da Sociedade Brasileira de Computação - Encontro Nacional de Inteligência Artificial. v. 4, p. 537–539, 1999.

Brandão, T.; Sequeira, M.M. & Albuquerque, M., Multistage morphology-based license-plate location algorithm. In: *Proceedings of 5th International Workshop on Image Analysis for Multimedia Interactive Services*. Lisboa, Portugal, 2004.

de Campos, T.J., *Reconhecimento de Caracteres Alfanuméricos de Placas em Imagens de Veículos*. Dissertação de mestrado, Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2001.

Chang, S.L.; Chen, L.S.; Chung, Y.C. & Chen, S.W., Automatic license plate recognition. *IEEE Transactions on Intelligent Transportation Systems*, 5(1):42–53, 2004.

Conci, A. & Monteiro, L.H., *Reconhecimento de Placas de Veículos por Imagens*. Dissertação de mestrado, Universidade Federal Fluminense, Niterói, RJ, 2004.

Draghici, S., *A Neural Network Based ArtificialVvision System for Licence Plate Recognition*. Tese de doutorado, Department of Computer Science, Wayne State University, Detroit, USA, 2007.

Figueiredo Filho, D.B. & Silva Júnior, J.A., Desvendando os mistérios do coeficiente de correlação de Pearson. *Revista Política Hoje*, 19(1):115–146, 1999.

Gonzalez, R.C. & Woods, R.E., *Digital Image Processing*. 3a edição. New Jersey: Pearson Prentice Hall, 2008.

Khalifa, O.; Khan, S.; Islam, R. & Suleiman, A., *Malaysian Vehicle License Plate Recognition*. Tese de doutorado, Kulliyyah of Engineering, International Islamic University, Malaysia, 2006.

LPDSI, , LPR image database sample. Disponível em: <http://www.cbpf.br/cat/pdsi/downloads/LPR_ImageDataBase_Sample.zip>, Acesso em 08 de novembro, 2011.

Martinsky, O., *Algorithmic and Mathematical Principles of Automatic Number Plate Recognition Systems*. B.Sc. Thesis, Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic, 2007.

Sancho, X.G., *A Simple License Plate Recognition System for Spanish License Plates*. Tese de doutorado, Universitat Rovira i Virgili, Tarragona, Spain, 2006.

Notas Biográficas

Leonardo Augusto Oliveira é graduado em Engenharia Elétrica com ênfase em Eletrônica e habilitação em Sistemas Digitais (Universidade de São Paulo/câmpus São Carlos, 2010).

Adilson Gonzaga é graduado e mestre em Engenharia Elétrica (Universidade de São Paulo/São Carlos, 1977 e 1982, respectivamente) e doutor em Física (Universidade de São Paulo/São Carlos,1991). Atualmente é professor Associado da Universidade de São Paulo. Tem experiência na área de engenharia elétrica, sistemas digitais, microprocessadores e processamento de imagens, atuando principalmente nos seguintes temas: visão computacional, processamento de imagens, biometria e reconhecimento de padrões.

Capítulo 16

Segmentação de Gestos e Camundongos por Subtração de Fundo, Aprendizagem Supervisionada e *Watershed*

Bruno Brandoli Machado, Wesley Nunes Gonçalves, Hemerson Pistori*,
Jonathan de Andrade Silva, Kleber Padovani de Souza,
Bruno Toledo e Wesley Tessaro

Resumo: Este capítulo apresenta uma comparação entre técnicas de segmentação baseadas em subtração de fundo, aprendizagem supervisionada e *watershed* aplicadas a dois problemas reais: (1) detecção de pele humana e (2) rastreamento de camundongos. Os resultados obtidos utilizando cinco métricas de comparação diferentes indicam que a abordagem de subtração de fundo é mais eficiente na segmentação de imagens com camundongos e a abordagem de aprendizagem supervisionada com máquinas de vetores de suporte é indicada para detecção de pele.

Palavras-chave: Segmentação, Subtração de fundo, Aprendizagem supervisionada, *Watershed*.

Abstract: *This chapter presents a comparison of some image segmentation techniques based on background subtraction, supervised learning and watershed applied to two real applications: (1) human skin detection and (2) mouse tracking. The results using five different comparison metrics indicate that background subtraction performs better in mice images and that the supervised learning approach, using support vector machines, achieves better results for human skin detection.*

Keywords: *Segmentation, Background subtraction, Supervised learning, Watershed,*

*Autor para contato: pistori@ucdb.br

1. Introdução

A segmentação de imagens é uma etapa essencial para diversos sistemas de visão computacional. A ideia principal é particionar imagens em regiões que sejam homogêneas de acordo com um determinado critério de avaliação. Porém, o particionamento da imagem é altamente dependente do domínio da aplicação. Assim, a avaliação de técnicas de segmentação para um domínio diferente é uma questão que cabe ao pesquisador testá-las experimentalmente. Existe uma grande quantidade de aplicações em segmentação de imagens. Alguns dos exemplos típicos incluem: rastreamento de pessoas em sistemas de segurança (Fuentes & Velastin, 2006; Lu & Manduchi, 2011), detecção e reconhecimento de face (Yang et al., 2002; Li & Ngan, 2008), rastreamento de objetos (Spagnolo et al., 2006; Bugeau & Pérez, 2009), reconhecimento de gestos (Kim et al., 2007), segmentação de imagens médicas (Nie et al., 2009; Harati et al., 2011), segmentação de células (Ko et al., 2011) e sensoreamento remoto (Wang et al., 2010).

Neste trabalho é investigado o desempenho de técnicas de segmentação sobre imagens provenientes de duas aplicações: a detecção de pele para o reconhecimento de gestos usadas em língua de sinais e a identificação de camundongos em estudos etológicos usados na classificação de comportamento animal. A detecção de pele é fundamental em diversos sistemas de interação homem-máquina baseados em visão e de comunicação homem-homem mediada por computadores (Murthy & Jadon, 2009). O mesmo pode se dizer da segmentação de pele humana para o reconhecimento de gestos baseado em visão, principalmente quando o sinalizador – indivíduo que transmite a mensagem por gestos – não utiliza recursos adicionais de apoio, tais como luvas coloridas ou marcadores. Outro problema estudado por nosso grupo é a identificação automática de comportamentos de camundongos para avaliar o efeito de novos fármacos (Pistori et al., 2010). Embora as aplicações não possuam relação direta do ponto de vista prático, o resultado da segmentação é útil para o tratamento de imagens em etapas posteriores, por exemplo, no rastreamento de animais.

Neste estudo, os experimentos foram realizados sobre um conjunto de 40 imagens de estudos etológicos e um conjunto com 240 imagens de gestos de Língua Brasileira de Sinais – LIBRAS. Além disto, as imagens de referência, chamadas ao longo do texto de *ground-truth*, foram produzidas para todas as amostras com a ajuda de especialistas de ambas as áreas. Tais imagens são tipicamente usadas para avaliar a qualidade dos algoritmos, reduzindo assim a subjetividade de avaliações baseadas apenas em análise visual. Aqui foram usadas cinco métricas estudadas em Sezgin & Sankur (2004), são elas: (1) taxa de acurácia ou porcentagem de classificação correta, (2) coeficiente de Jaccard, (3) coeficiente de Yule, (4) área relativa ao erro e (5) erro de classificação.

As principais contribuições do nosso trabalho são o estudo comparativo de três grupos de técnicas de segmentação para duas aplicações reais. O valor da pesquisa está em orientar pesquisadores que trabalham em aplicações semelhantes. Referente ao texto, este trabalho está organizado em quatro seções. A Seção 2 apresenta as técnicas de segmentação relacionadas ao fundo da imagem. Na Seção 3 são descritas as medidas de desempenho aqui utilizadas. Os experimentos e os resultados obtidos são discutidos na Seção 4. Finalmente, as conclusões e trabalhos futuros são apresentados na Seção 5.

2. Técnicas de Segmentação

Esta seção apresenta uma breve revisão das técnicas de segmentação utilizadas neste trabalho: subtração de fundo, subtração por fundo adaptativo, modelo Gaussiano, árvores de decisão, redes neurais artificiais e máquinas de vetores de suporte. Basicamente, tais técnicas fazem o uso das propriedades das imagens, por exemplo, a informação de cor (intensidades dos pixels) ou de um conjunto de características, por exemplo, informação de textura, para segmentar as imagens (Solomon & Breckon, 2010). Neste trabalho, as propriedades de cor serão utilizadas para discriminar os objetos de interesse (*foreground*) dos objetos de fundo da imagem (*background*).

Considere uma imagem $\mathbf{I} = \{\mathbf{p}_1, ..., \mathbf{p}_n\}$ descrita por vetores de pixels $\mathbf{p}_i \in \Re^3, i \in [1, ..., n]$ — sendo n a quantidade de pixel da imagem. Cada pixel $\mathbf{p}_i$ é descrito pelo espaço de cor RGB, respectivamente as componentes de cor: vermelho (R), verde (G) e azul (B). Cada componente do espaço de cores RGB é descrito por um vetor de valores que varia de 0 a 255, por exemplo, a componente R é descrita pelo vetor $R = [0, ..., 255]$, sendo que 0 é o tom mais escuro e 255 o tom mais claro. Portanto, $\mathbf{p}_i = [r, g, b]$, sendo que $r \in R$, $g \in G$ e $b \in B$. Para imagens em tons de cinza em que os valores de intensidade de seus pixels variam dentro do intervalo $[0, ..., 255]$, cada pixel $\mathbf{p}_i$ é descrito por um único valor $q \in [0, ..., 255]$, nesse caso, $\mathbf{p}_i = q$.

2.1 Subtração de fundo e fundo adaptativo

A subtração de fundo está entre as técnicas mais usadas em visão computacional, devido sua escalabilidade e simplicidade (Cheung & Kamath, 2005; Khan et al., 2009). Este grupo de técnicas é utilizado para diversas tarefas, por exemplo, monitoramento de trânsito (Cheung & Kamath, 2005) e aplicações em multimídia (Baf et al., 2007).

Basicamente, a técnica de subtração calcula a diferença entre duas imagens, uma imagem de referência (que representa o fundo – *background*) e uma imagem atual, conforme apresentado na Equação 1. Seja um *pixel* $\mathbf{p}_i^{\mathrm{c}}$ da imagem atual $\mathbf{I}_{\mathrm{c}}$, este é subtraído do pixel $\mathbf{p}_i^{\mathrm{r}}$ da imagem de referência $\mathbf{I}_{\mathrm{r}}$. Quando essa diferença é maior que um limiar τ – cujo valor é determi-

nado pelo usuário – o pixel $\mathbf{p}_i^{\mathrm{c}}$ é considerado como pixel de interesse. Caso contrário, esse pixel é considerado como fundo.

$$\mathbf{I} = \{\forall \mathbf{p}_i^{\mathrm{c}} \in \mathbf{I}_{\mathrm{c}} : |\mathbf{p}_i^{\mathrm{c}} - \mathbf{p}_i^{\mathrm{r}}| > \tau\} \tag{1}$$

No caso em que $\mathbf{p}_i$ é m-dimensional, por exemplo, $\mathbf{p}_i = [r, g, b]$ com $m = 3$, pode-se utilizar a distância Euclidiana entre os pixels, conforme apresentado na Equação 2, onde $dist_E(\cdot)$ é a distância Euclidiana entre os pixels. O objetivo é realçar apenas os objetos de interesse da imagem atual, removendo os objetos que fazem parte do fundo.

$$\mathbf{I} = \{\forall \mathbf{p}_i^{\mathrm{c}} \in \mathbf{I}_{\mathrm{c}} : dist_E(\mathbf{p}_i^{\mathrm{c}}, \mathbf{p}_i^{\mathrm{r}}) > \tau\} \tag{2}$$

No entanto, a subtração de fundo não atualiza as informações de fundo. Consequentemente, ela não é capaz de lidar com mudanças de cenas nas imagens e com as diferentes condições de iluminação ou objetos irrelevantes, que ocasionalmente aparecem nas imagens e permanecem imóveis durante um período de tempo. Para superar esta limitação, uma variante adaptativa que atualiza a imagem de referência ao longo do tempo foi proposto em (Heikkilä & Silvén, 2004), tornando este método mais robusto em relação à técnica tradicional. Nesta variante a imagem de referência é ajustada iterativamente da seguinte forma:

$$\mathbf{I}_{\mathrm{r}}(z+1) = \alpha \mathbf{I}_{\mathrm{c}}(z) + (1-\alpha)\mathbf{I}_{\mathrm{r}}(z) \tag{3}$$

onde z corresponde ao valor da iteração e $\alpha \in [0, 1]$ é uma constante de atualização.

2.2 Segmentação baseada no aprendizado supervisionado

Aprendizagem de Máquina (AM) envolve o estudo de métodos computacionais que aprendem e aperfeiçoam o seu desempenho a partir de experiência (Mitchell, 1997). No aprendizado supervisionado, cada instância[1] de um conjunto de dados contém a informação do rótulo da classe. Especificamente, seja um conjunto de dados $\mathbf{X}$ com n instâncias, *i.e.,* $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n}$, sendo que cada $\mathbf{x}_i$ é descrito por um vetor de valores de atributos ou características m-dimensional $\mathbf{x}_i = [x_i^j]_{j=1}^{m}$. Além disto, para cada instância $\mathbf{x}_i$ é atribuído um rótulo de classe $c_i \in C$, sendo $C = \{c_1, ..., c_k\}$ o conjunto de k possíveis classes. Portanto, o conjunto de dados é representado pela tupla $< \mathbf{x}_i, c_j >$.

Os algoritmos de AM podem ser utilizados para a segmentação de imagens por meio da classificação de um conjunto de pixels. Basicamente, a classificação de dados é realizada em por um procedimento descrito em

[1] O termo instância pode ser também compreendido como uma tupla, um exemplo ou um registro de um conjunto de dados.

duas etapas (Han & Kamber, 2000): i) construção de um modelo (classificador) a partir de um conjunto de treinamento e ii) inferência dos rótulo das instâncias, cuja a informação da classe é desconhecida, de um conjunto de *teste* por meio do modelo construído.

No contexto de segmentação de imagens, considere um conjunto de pixels de um objeto de interesse O_F de tamanho l. Os pixels de O_F são extraídos para compor o conjunto de treinamento $\mathbf{S} = \{\mathbf{p}_i\}_{i=1}^{l}$. Cada pixel de $\mathbf{S}$ receberá o rótulo para classe F (pertencente ao objeto de interesse), ou seja, $< \mathbf{p}_i, c >$, sendo $c \in C = \{F\}$. Daí um classificador é treinado utilizando as instâncias do conjunto de treinamento, rotuladas com a classe F. Este problema de classificação é conhecido como *one-class classification* (Moya et al., 1993), no qual um classificador é treinado apenas com instâncias de uma classe. Para a construção de um modelo a partir de $\mathbf{S}$, aqui foi utilizado o algoritmo de aprendizagem estatística Gaussiana (Seção 2.2.1). Para o problema de classificação de k-classes foram adotados três algoritmos: árvores de decisão (Seção 2.2.2), redes neurais artificiais (Seção 2.2.3) e o algoritmo de máquinas de vetores de sporte (Seção 2.2.4), sendo que as instâncias do conjunto de treinamento são rotuladas como pertencente ao objeto de interesse F ou ao fundo B, isto é, $< \mathbf{p}_i, c_j >$, em que $c_j \in C = \{F, B\}$ e $k = 2$.

2.2.1 Aprendizagem estatística Gaussiana

Nesta estratégia (Vapnik, 1998), os parâmetros de uma distribuição Gaussiana multivariada $\mathcal{N}(\mu, \Sigma)$ são estimados a partir dos pixels de $\mathbf{S}$, definido por:

$$\mu = \frac{1}{|\mathbf{S}|} \sum_{\mathbf{p} \in \mathbf{S}} \mathbf{p}, \quad \Sigma = \frac{1}{|\mathbf{S}|} \sum_{\mathbf{p} \in \mathbf{S}} [\mathbf{p} - \mu][\mathbf{p} - \mu]^t \tag{4}$$

onde μ é a média, Σ corresponde à matriz de covariância, $\mathbf{p}$ é um vetor (pixel) do conjunto de treinamento. Portanto, para o conjunto de treinamento, que é composto de instâncias rotuladas com a classe F, é estimado os parâmetros da distribuição Gaussiana (construção do modelo). Nesse caso, temos uma distribuição $\mathcal{N}(\mu, \Sigma)$.

Após a etapa de treinamento, um novo pixel $\mathbf{p}'$, cujo rótulo da classe é desconhecido, é atribuído o rótulo de classe F, se a sua distância de Mahalanobis com relação a $\mathcal{N}(\mu, \Sigma)$ é maior do que um limiar α, cujo valor é definido pelo usuário, ou seja, $dist_M(\mathbf{p}', \Phi) > \alpha$, sendo Φ uma representação para a distribuição normal $\mathcal{N}(\mu, \Sigma)$. Caso contrário, este pixel é considerado como pertencente ao fundo e descartado. A distância de Mahalanobis entre $\mathbf{p}'$ e uma distribuição normal $\mathcal{N}(\mu, \Sigma)$ é apresentada na Equação 5.

$$dist_M(\mathbf{p}', \Phi) = [\mathbf{p}' - \mu]^t \Sigma^{-1} [\mathbf{p}' - \mu] \tag{5}$$

2.2.2 Árvores de decisão

A indução de árvores de decisão (Quinlan, 1986) é baseada no uso da estratégia de divisão e conquista para tarefas de classificação. Originalmente, o algoritmo divide o espaço de atributos em regiões de decisão, gerando nós internos (não-folhas) ou nós de teste. A ideia é construir uma estrutura hierárquica de maneira que cada nó não-folha corresponde a um atributo e os nós folhas correspondam aos rótulos das classes. As conexões entre os nós correspondem aos diferentes valores para os atributos. Um exemplo de árvore de decisão é apresentado na Figura 1.

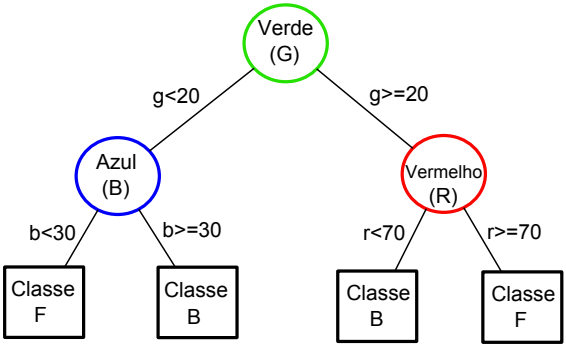


Figura 1. Exemplo de uma árvore de decisão para classificação de pixels $\mathbf{p}_i = [r, g, b]$ descritos pelos componentes R, G e B do espaço de cores.

Assim, os atributos de uma instância são testados ou avaliados em cada nó não folha da árvore desde a raiz via procedimento de busca *top-down* para inferir o rótulo da classe. Dentre as diversas implementações deste método, usamos o algoritmo C4.5 que divide cada nó sobre baseado no ganho de informação. Seja uma atributo A com v valores distintos $\{a_1, ..., a_v\}$ observados do conjunto de treinamento. Pode-se obter uma partição de $\mathbf{S}$ em v subconjuntos $\{\mathbf{S}_1, ..., \mathbf{S}_v\}$, onde $\mathbf{S}_j$ contém as instâncias de $\mathbf{S}$ cujo valor do atributo A é a_j. Dessa maneira, o cálculo do ganho para uma atributo A é apresentado na Equação 7.

$$Entropia(\mathbf{S}) = -\sum_{i=1}^{k} u_i \log_2(u_i) \tag{6}$$

$$Ganho(\mathbf{S}, A) = Entropia(\mathbf{S}) - \sum_{j=1}^{v} \frac{|\mathbf{S}_j|}{|\mathbf{S}|} Entropia(\mathbf{S}_j) \tag{7}$$

onde u_i é a probabilidade de uma instância de $\mathbf{S}$ pertencer a classe c_i, cujo valor é estimado pela proporção de instâncias em $\mathbf{S}$ pertencente a classe c_i. O atributo com maior valor de ganho é escolhido para a divisão.

2.2.3 Redes neurais artificiais

Redes Neurais Artificiais (RNAs) são modelos matemáticos inspirados nas estruturas neurais biológicas (Haykin, 1998). Existem diversos modelos para implementação de uma RNA, por exemplo, SOM (*Self Organizing Map*) e MLP (*Multi-Layer Perceptron*). Neste trabalho, serão utilizadas as estruturas do tipo MLP. Maiores detalhes sobre as arquiteturas de redes neurais artificiais podem ser encontrados em (Haykin, 1998).

MLP é uma arquitetura particular de rede neural *feedforward* para dados não-linearmente separáveis (Rumelhart et al., 1986). Uma rede MLP consiste em uma ou mais camadas ocultas entre uma entrada e a uma camada de saída dos neurônios, conforme apresentado na Figura 2. Cada neurônio é totalmente conectado aos neurônios da camada seguinte e pode ser descrito como um elemento de processamento, que é ativado por uma função de ativação não-linear. A ativação da rede neural é o produto interno do vetor de entrada ($\mathbf{p}$), com os pesos (conexões) em camadas ocultas. Para o treinamento do MLP, foi adotado o algoritmo *backpropagation* (Haykin, 1998).

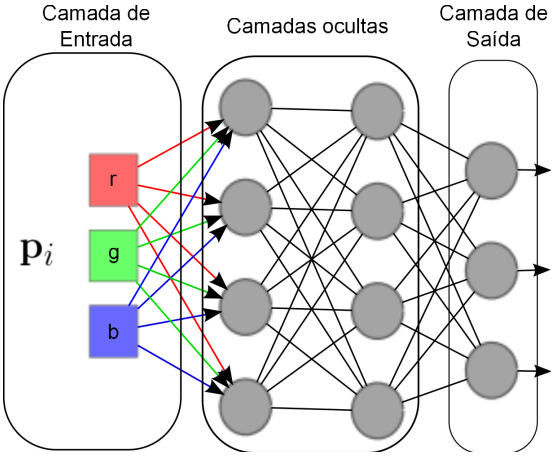


Figura 2. Exemplo de MLP com duas camadas ocultas cada qual com quatro neurônios e uma camada de saída com três neurônios. para classificação do pixel $\mathbf{p}_i = [r, g, b]$ descritos pelas componentes R, G e B do espaço de cores.

2.2.4 Máquina de vetores de suporte

Máquinas de Vetores de Suporte (*Support Vector Machines* – SVM) (Vapnik, 1979) é uma técnica amplamente utilizada para classificação de dados. Para o problema de classificação de pixels, considere um conjunto de treinamento $\mathbf{S}$ com instâncias rotuladas a uma das classe $C = \{F, B\}$. A

SVM encontra um hiperplano que separa as duas classes cuja margem de separação é máxima. SVM é caso particular de algoritmos de classificação baseados em *kernels* que mapeiam as instâncias de um conjunto de treinamento para um espaço de alta dimensão para encontrar o hiperplano de separação. Entretanto, o treinamento de uma SVM para encontrar a solução ótima, isto é, o hiperplano com separação máxima, tem custo computacional quadrático (Platt, 1999). Alternativamente, o algoritmo SMO (*Sequential Minimal Optimization*) (Platt, 1999) pode ser utilizado para solucionar esse problema de otimização de maneira mais eficiente. A estratégia é dividir o problema em um subconjunto de problemas menores (divisão-e-conquista). Desta maneira, optou-se pelo uso do algoritmo SMO nesse trabalho.

3. Métodos de Avaliação

A avaliação visual de imagens não é um critério confiável para analisar o resultado de algoritmos de segmentação. A maneira mais conhecida de se medir o desempenho é pela comparação quantitativa do resultado de um algoritmo em relação às imagens de *ground-truth*. A Figura 3 mostra quatro exemplos de imagens de *ground-truth* para avaliar os algoritmos. A avaliação pode ser obtida pelo acerto em termos de pixels ou em regiões da imagem. Neste trabalho são computadas cinco medidas, três delas baseadas em pixels e duas baseadas em propriedades da região. Em seguida, elas são descritas.

Figura 3. Amostra de quatro imagens de gestos (linha 1) e as imagens de referência geradas manualmente para cada uma (linha 2).

3.1 Avaliação baseada em pixel

Nesta abordagem, os resultados são baseados na comparação entre o *ground-truth* e as imagens de saída. Cada pixel analisado pode ser classificados em quatro rótulos, nomeadamente: (1) verdadeiro positivo, (2)

falso positivo (3), negativo verdadeiro, e (4) falso negativo. Um verdadeiro positivo (VP) ocorre quando o resultado da predição corresponde ao *ground-truth*, caso contrário, um falso positivo (FP) é encontrado. Por outro lado, um verdadeiro negativo (VN) ocorre quando o resultado da predição é diferente da imagem de *ground-truth*, ou então, ele é considerado um falso negativo (FN).

Seguindo o esquema de classificação dos pixels, é possível calcular três medidas quantitativas sobre duas imagens binárias: a porcentagem de classificação correta (PCC), o coeficiente de Jaccard (CJ) e o coeficiente de Yule (CY). Tais medidas são formalizadas nas Equações 8a, 8b e 8c, respectivamente.

$$PCC = \frac{VP + FP}{VP + FP + VN + FN} \tag{8a}$$

$$CJ = \frac{VP}{VP + FP + FN} \tag{8b}$$

$$CY = \left| \frac{VP}{VP + FP} + \frac{VN}{VN + FN} - 1 \right| \tag{8c}$$

3.2 Avaliação baseada em região

Esta abordagem mede a qualidade da segmentação usando regiões. Além disto, as medidas de desempenho são marcadas variando de 0, indicando uma segmentação perfeita, até 1, para um caso totalmente incorreto. A área relativa ao erro (*Relative Area Error* – RAE) do objeto segmentado é obtido pela formas e áreas a partir da imagem segmentada contra o *ground-truth* (Sezgin & Sankur, 2004). A medida RAE é 0 se ocorrer a combinação perfeita entre a imagem de saída e a imagem de *ground-truth*, enquanto a combinação mínima é 1. A Equação 9 define medida RAE, onde A_0 é a área da imagem de *ground-truth* e A_t é a área da imagem segmentada.

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

Outra métrica abordada é o erro de classificação (do inglês ME). Ele é obtido calculando a porcentagem de pixels de fundo segmentado como objeto, ou frente estabelecido como fundo. A métrica ME é formalizada na Equação 10, em que B_0 e F_0 referem-se ao fundo e os objetos da imagem de *ground-truth*; B_t e F_t indicam o fundo e os objetos em primeiro plano da imagem segmentada; e $|.|$ representa a cardinalidade do conjunto.

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

4. Resultados Experimentais

Para avaliar os algoritmos descritos neste trabalho, os experimentos foram realizados sobre dois conjuntos de imagens. Primeiro, são descritos ambos os conjuntos e, em seguida, detalhes da configuração dos experimentos. Por fim, são discutidos o desempenho de cada grupo de técnica.

4.1 Conjunto de imagens

4.1.1 Conjunto de análise de comportamento animal

Este conjunto consiste de dois comportamentos amplamente estudados em áreas que envolvem a Etologia: locomoção espacial e exploração vertical. Exemplos de cada comportamento são mostrados na Figura 4. A locomoção espacial (coluna à esquerda da Figura 4) corresponde à caminhada do animal na arena, enquanto a exploração vertical é caracterizada quando o roedor deixa as patas posteriores do piso (coluna à direita da Figura 4), em outras palavras, quando ele fica em pé. Pesquisadores da Universidade Católica Dom Bosco (UCDB) têm analisado o comportamento de espécies de camundongos. Geralmente os experimentos etológicos ocorrem usando a arena campo aberto (Walsh & Cummins, 1976; Prut & Belzung, 2003). O campo aberto é um dos procedimentos mais populares em ciências biológicas (fisiologia, farmacologia e psicologia) por ser um modelo confiável para quantificar comportamentos animais. As medidas da área da arena seguem as medidas tradicionais encontradas em Hall (1934), neste caso 1,2 m de diâmetro com uma parede de 0,45 m de altura.

A arena permite que pesquisadores testem os efeitos de novos fármacos medidos pelo número de vezes em que o animal realizou um determinado comportamento. Embora esta tarefa aparenta ser simples, diversos problemas relacionados com fadiga e diferença na quantificação dos comportamentos tem sido reportados na literatura (Carroll & Hughes, 1978; Stanford, 2007). Com o objetivo de apoiar a quantificação dos comportamentos, o grupo de pesquisa INOVISAO tem trabalhado em parceria com pesquisadores da área da saúde da UCDB. Um conjunto de 40 imagens de imagens foi separado pelos especialistas usando duas espécies de camundongos, conhecidas por Suíço e C57. As imagens capturadas possuem dimensão de 640×480 pixels e são coloridas. Vale ressaltar que camundongos da espécie Suíço são animais com pelagem branca, enquanto animais da espécie C57 apresentam pelagem preta. Normalmente, os experimentos em áreas biológicas acontecem usando o contraste entre a cor do animal e a arena para facilitar a análise dos comportamentos. Aqui o interesse está em avaliar a capacidade dos algoritmos de segmentação, considerado uma etapa de pré-processamento para que depois o animal seja rastreado na arena usando técnicas de visão computacional (Pistori et al., 2010). A maior dificuldade nos experimentos está em segmentar animais em que a

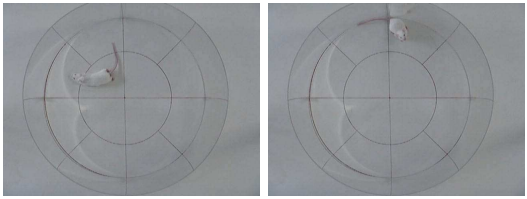
sua cor está fortemente correlacionada com a cor do fundo, por exemplo, um animal da espécie Suíço, com a pelagem branca, juntamente com o fundo branco da arena. Combinando dois tipos de fundo e duas espécies de animais obtém-se quatro configurações diferentes. Duas imagens para cada combinação é mostrado na Figura 4.

4.1.2 Conjunto de língua brasileira de sinais – LIBRAS

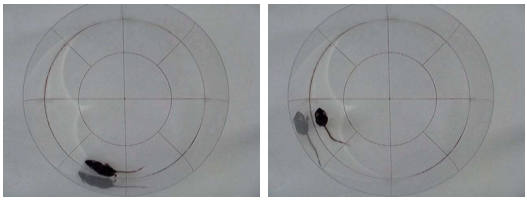
Este conjunto consiste de imagens tiradas de quatro configurações de ambiente. As configurações são capturadas sob duas condições de iluminação e dois tipos de fundo, cada qual com 6 atores que realizam 10 posturas cada, resultando um total de 240 imagens. Exemplo de cada postura é mostrado na Figura 5. O grupo de pesquisa INOVISAO tem avaliado técnicas de segmentação para detecção de pele em imagens. A detecção de pele por segmentação pode ser considerada um passo de pré-processamento para análise de imagens ou vídeo. O objetivo é então utilizar o resultado da segmentação para reconhecer e rastrear gestos e vários outros domínios de interação humano-computador (Murthy & Jadon, 2009). Aqui, a informação da cor dos pixels (modelo RGB) foi a característica utilizada nos experimentos. Porém, a representação por meio da cor se torna uma tarefa desafiadora pela sensibilidade de três fatores: iluminação, etnicidade e características individuais de aparência. Para criar um conjunto de imagens com tais variações, as posturas presentes neste conjunto foram tiradas de seis pessoas com tons de pele diferentes. O tamanho das imagens é de 800 $\times$ 600 pixels. Para cada pose foi tirada duas configurações de iluminação: natural e artificial. Um total de 120 imagens são capturadas usando iluminação artificial, neste caso dentro de salas, e outras 120 imagens são tiradas sob iluminação natural. Ainda nesse conjunto, foram capturadas duas configurações diferentes em termos de fundo: plano (simples) e complexo. O fundo plano ou simples é caracterizado por não possuir nenhum objeto próximo ao corpo, enquanto o fundo complexo corresponde às imagens tiradas em cenários com vários objetos ou detalhes. As combinações dos tipos de iluminação e fundo, assim como os seis sinalizadores, são mostradas na Figura 5.

4.2 Configuração dos experimentos

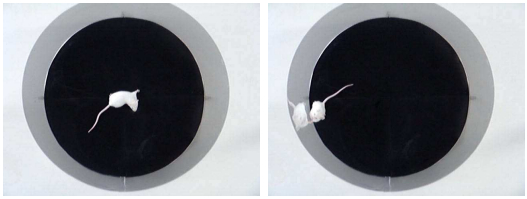
Os experimentos foram realizados com três grupos de técnicas de segmentação (veja a Seção 2). A comparação é feita pelas medidas de desempenho (veja a Seção 3). Primeiro, criou-se as imagens de *ground-truth* de ambos os conjuntos de imagens, um processo que demandou muito tempo. Em relação ao grupo de técnicas, para as técnicas de subtração foi necessário capturar a imagem referência. Além disto, amostras com tamanho de 40 $\times$40 pixels foram coletadas com o objetivo de treinar os modelos de aprendizagem supervisionada. Para gestos, amostras de diferentes partes



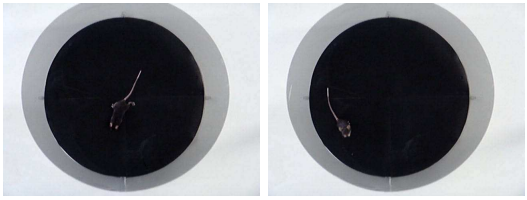
(a) Fundo branco e camundongo Suíço



(b) Fundo branco e camundongo C57



(c) Fundo preto e camundongo Suiço



(d) Fundo preto e camundongo C57

Figura 4. Dois exemplos de cada configuração dos experimentos [a-d]. A arena consiste em uma base de madeira revestida por fórmica e uma parede de acrílico que evita a fuga do animal. Os animais são submetidos a explorar o ambiente desconhecido e, ao longo do tempo, especialistas observam dois tipos de comportamentos: locomoção espacial (coluna à esquerda) e exploração vertical (coluna à direita).

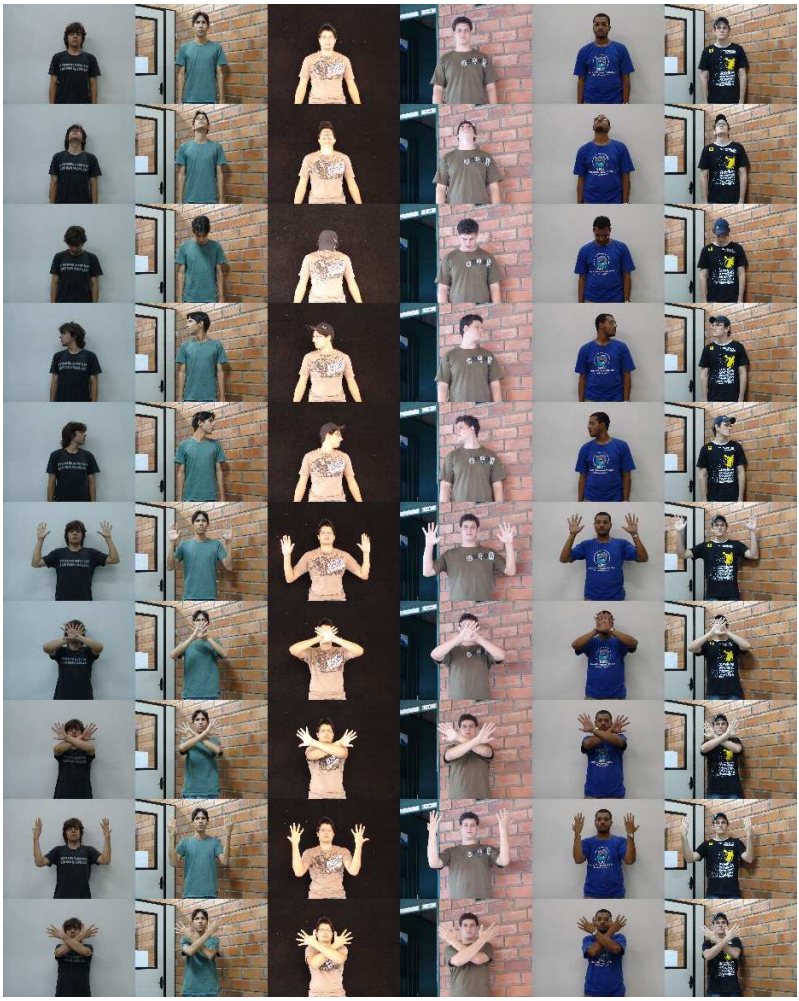


Figura 5. Amostra dos dez gestos existentes no conjunto de imagens LIBRAS, realizados por 6 diferentes sinalizadores com diferentes tonalidades de pele. Além disto, as quatros primeiras colunas correspondem às combinações de condições de iluminação e fundo, as duas primeira com iluminação artificial e as duas seguintes iluminação natural.

do corpo são extraídas, tais como partes de face, braço direito, braço esquerdo e pescoço. Para imagens do comportamento do camundongo, cinco amostras do animal e cinco amostras da arena foram selecionadas. Parâ-

metros para ambas as abordagens foram testados empiricamente para cada algoritmo, resultando em mais de 70.000 imagens segmentadas. Aqui, os resultados de ambos os conjuntos de imagens foi alcançado pela precisão média para determinar a precisão de cada algoritmo de segmentação.

4.3 Resultados

Experimento 1: Primeiro é feita a comparação entre as combinações de fundo e cor da pelagem dos animais, mostrada no gráfico da Figura 6. Os animais da espécie Suíço possuem pelagem branca e os animais C57 apresentam pelagem preta. As combinações de espécie e cor de fundo são mostradas no eixo-y, enquanto o desempenho são indicados no eixo-x. Os maiores desempenhos são obtidos pelo contraste entre o fundo e a cor do pelo do animal, por exemplo, fundo preto e camundongo Suíço. Para técnicas de aprendizagem supervisionada, a combinação fundo branco obteve melhores resultados.

Experimento 2: As Figuras 7 e 8 mostram o desempenho dos métodos de segmentação para as combinações de fundo e cor da pele humana. As combinações são mostradas no eixo-y do gráfico, enquanto os desempenhos são indicados no eixo-x. Pode-se observar que as técnicas de subtração de fundo com a presença de iluminação artificial não atinge um bom desempenho para as quatros combinações. Há uma clara diferença em desempenho para fundo simples ao grupo de técnicas de subtração, assim como para aquelas baseadas em aprendizagem supervisionada. Os modelos árvores de decisão, redes neurais artificiais e máquinas de vetores de suporte atingem um desempenho muito superior quando comparados as técnicas de subtração. Focando em iluminação natural, é possível observar que a combinação para ambos tipos de fundos e peles atingem melhores resultados em abordagens de subtração de fundo em relação à iluminação artificial. Estes resultados mostram-se consistentes devido à sombra em que os sinalizadores sofrem ao realizar as posturas. Olhando para os gráficos de abordagem supervisionada, os resultados tem o mesmo comportamento em termos de acerto, independente da condição de iluminação, porém obtendo desempenho um pouco superior em ambientes de luz natural.

Experimento 3: Experimentos preliminares tem sido realizados para avaliar a técnica de segmentação por *watershed*, especificamente a *watershed* baseada em imersão (Vincent & Soille, 1991). Foi usado o conjunto de imagens de camundongos composto de 40 imagens separadas em quatro categorias, de acordo com a combinação de fundo e cor da pelagem dos animais. A Tabela 1 apresenta os resultados para as quatro combinações. A comparação foi feita usando as métricas baseadas em pixels, pois algumas vezes a segmentação por *watershed* pode gerar regiões desconectadas

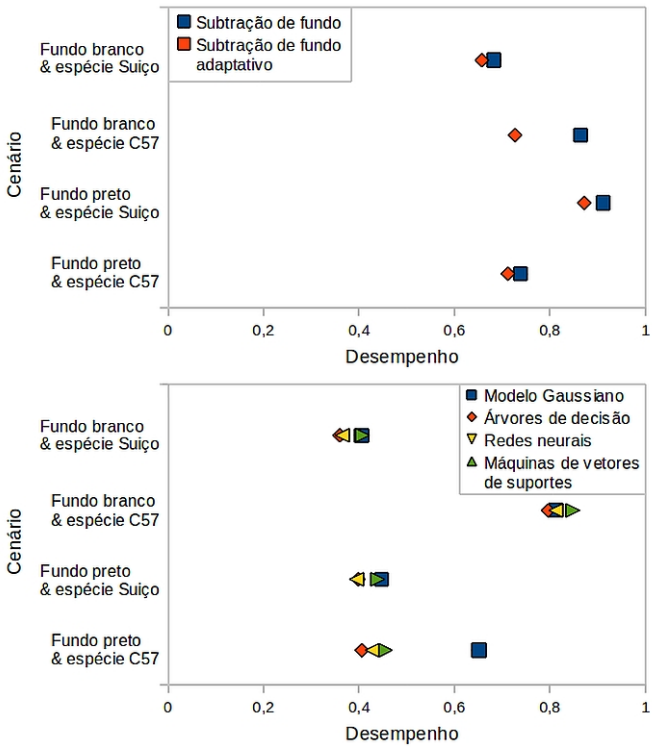


Figura 6. Avaliação de desempenho do conjunto de imagens de camundongos. Neste experimento foi avaliado apenas iluminação artificial, pois experimentos etológicos devem ocorrer em ambientes fechados.

e impede o uso da avaliação por regiões. Para este experimento, três medidas foram usadas: porcentagem de classificação correta (PCC), coeficiente de Jaccard (CJ), (3) coeficiente de Yule (CY). Além disto, foi calculado a média e o desvio padrão para cada medida. Os maiores valores para tais medidas foram atingidos em casos que o contraste entre a arena e o animal, por exemplo, fundo preto e camundongo branco. Vale ressaltar que o resultado do *watershed* pode variar dependendo do pré-processamento das imagens, pois a entrada deste método requer uma imagem binarizada. Além disto, vale ressaltar que mesmo com resultados não satisfatórios em casos que não haja contraste entre fundo e roedor, é possível ainda localizar o animal na cena, porém muitas vez com corpo do animal dividido em várias regiões.

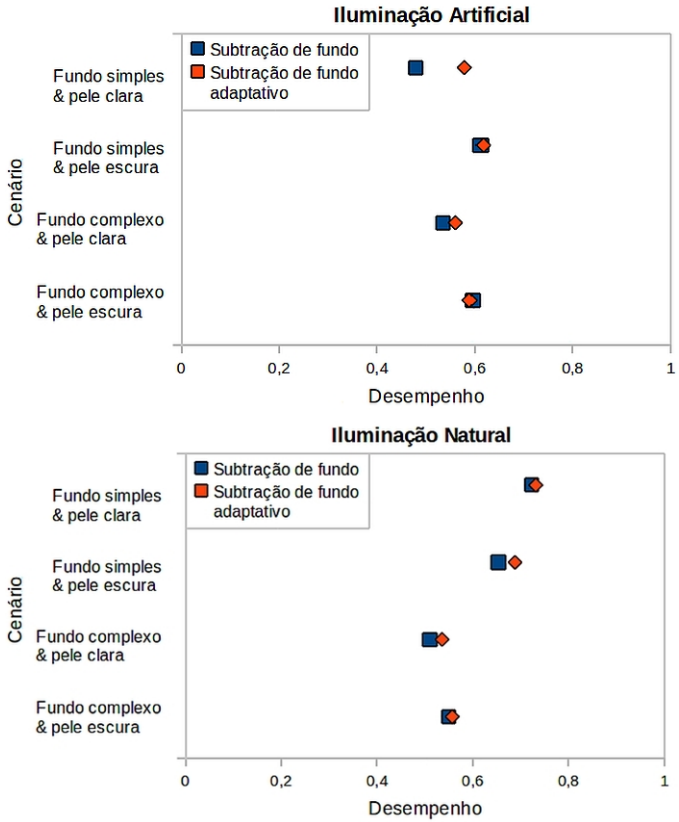


Figura 7. Avaliação de desempenho das técnicas baseadas em subtração de fundo do conjunto LIBRAS. Neste experimento foi avaliado as combinações de fundo e cor da pele humana. Diferente dos experimentos anteriores, as técnicas de segmentação foram avaliadas sob duas condições de iluminação: artificial e natural.

5. Conclusão

Neste trabalho foram avaliados dois grupos de técnicas de segmentação aplicados em duas importantes aplicações do mundo real: a detecção de pele para reconhecimento de gestos em Língua Brasileira de Sinais e a análise do comportamento de camundongos. Foram examinados os desafios de tais aplicações, incluindo a mudança de iluminação e a forte semelhança entre objetos relevantes e irrelevantes na imagem. A avaliação foi baseada em métricas de desempenho conhecidas cientificamente comparando-se ima-

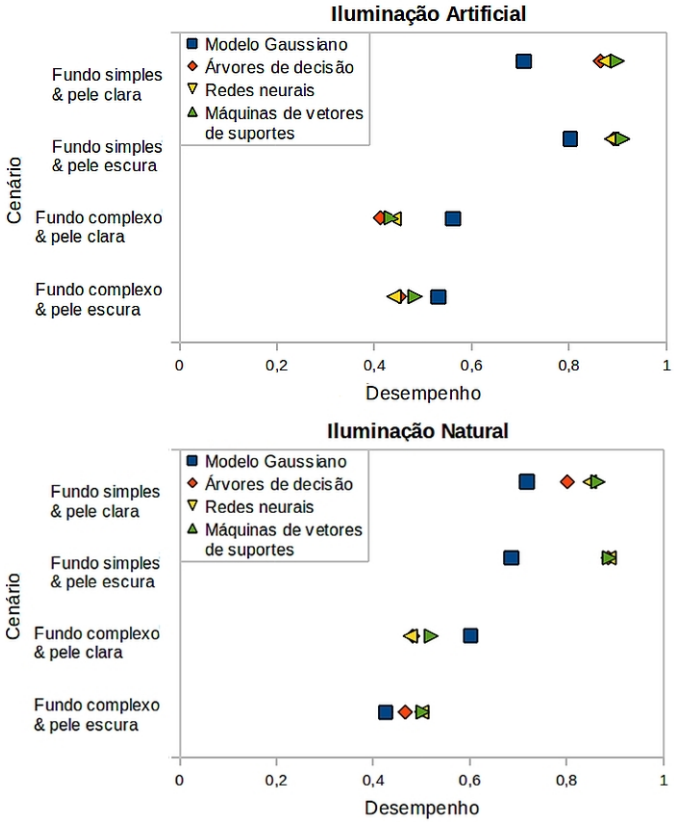


Figura 8. Avaliação de desempenho das técnicas baseadas em aprendizagem supervisionada do conjunto LIBRAS. Neste experimento foi avaliado as combinações de fundo e cor da pele humana. Também avaliadas sob duas condições de iluminação: artificial e natural.

gens segmentadas com as respectivas imagens de referência, segmentadas manualmente. Os resultados aqui apresentados apontam que os algoritmos utilizados apresentam bons desempenhos em um banco de dados de imagem específico. Os resultados dos experimentos sugerem fortemente que a abordagem de subtração de fundo é melhor para as imagens dos camundongos, principalmente em imagens com grande contraste. No entanto, também é preciso dar atenção a características visuais básicas em alguns casos, como, por exemplo, quando se lida com camundongos de espécie C57, que são animais de pelagem preta que, quando novos, apresentam estruturas epidérmicas bem visíveis. Com base nos resultados obtidos em

Tabela 1. Avaliação de desempenho do conjunto de imagens de camundongos usando a técnica *watershed* baseada em imersão.

		Camundongos	
Cenário	**Medida**	**Média**	**Desvio Padrão**
Fundo branco e espécie Suíço	PCC	0,98	0,0100
	CJ	0,39	0,2000
	CY	0,40	0,2200
Fundo branco e espécie C57	PCC	**0,99**	0,0005
	CJ	**0,69**	0,1100
	CY	**0,92**	0,0800
Fundo preto e espécie Suíço	PCC	**0,99**	0,0100
	CJ	0,66	0,1300
	CY	0,84	0,1900
Fundo preto e espécie C57	PCC	0,74	0,0100
	CJ	0,02	0,0070
	CY	0,02	0,0090

imagens de detecção de pele, observou-se que a abordagem de aprendizado supervisionado alcançou melhores resultados para ambas as condições de iluminação, em particular, as máquinas de vetor de suporte. Pesquisas futuras podem se concentrar na análise de diferentes técnicas de segmentação de imagem para contribuir com este trabalho. O nosso grupo busca avaliar o potencial de abordagens baseada em regiões, um caso particular a ser explorado tem sido a técnica *watershed*.

Agradecimentos

Este trabalho recebeu apoio financeiro da Universidade Católica Dom Bosco, UCDB, da Agência Financiadora de Estudos e Projetos, FINEP e da Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul, FUNDECT. Os autores também contaram com bolsas do Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPQ, nas modalidades ITI-A, PIBIC e Produtividade em Desenvolvimento Tecnológico e Extensão Inovadora.

Referências

Baf, F.E.; Bouwmans, T. & Vachon, B., Comparison of background subtraction methods for a multimedia learning space. In: *Proceedings of the Second International Conference on Signal Processing and Multimedia Applications*. Setúbal, Portugal: INSTICC Press, p. 153–158, 2007.

Bugeau, A. & Pérez, P., Detection and segmentation of moving objects in complex scenes. *Computer Vision and Image Understanding*, 113(4):459–476, 2009.

Carroll, W. & Hughes, , Observer influence on automated open field activity. *Physiology & Behavior*, 20(4):481–485, 1978.

Cheung, S.C.S. & Kamath, C., Robust background subtraction with foreground validation for urban traffic video. *EURASIP Journal on Applied Signal Processing*, 2005(14):2330–2340, 2005.

Fuentes, L.M. & Velastin, S.A., People tracking in surveillance applications. *Image and Vision Computing*, 24(11):1165–1171, 2006.

Hall, C., Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *Journal of Comparative Psychology*, 18(3):385–403, 1934.

Han, J. & Kamber, M., *Data Mining: Concepts and Techniques*. San Francisco, USA: Morgan Kaufmann, 2000.

Harati, V.; Khayati, R. & Farzan, A., Fully automated tumor segmentation based on improved fuzzy connectedness algorithm in brain MR images. *Computers in Biology and Medicine*, 41(7):483–492, 2011.

Haykin, S., *Neural Networks: A Comprehensive Foundation*. 2a edição. Upper Saddle River, USA: Prentice Hall PTR, 1998.

Heikkilä, J. & Silvén, O., A real-time system for monitoring of cyclists and pedestrians. *Image and Vision Computing*, 22(7):563–570, 2004.

Khan, M.H.; Kypraios, I. & Khan, U., 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*. New York, USA: ACM Press, p. 31:1–31:6, 2009.

Kim, D.; Song, J. & Kim, D., Simultaneous gesture segmentation and recognition based on forward spotting accumulative HMMs. *Pattern Recognition*, 40:3012–3026, 2007.

Ko, B.C.; Gim, J.W. & Nam, J.Y., Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake. *Micron*, 42(7):695–705, 2011.

Li, H. & Ngan, K.N., Saliency model-based face segmentation and tracking in head-and-shoulder video sequences. *Journal of Visual Communication and Image Representation*, 19(5):320–333, 2008.

Lu, X. & Manduchi, R., Fast image motion segmentation for surveillance applications. *Image and Vision Computing*, 29(2-3):104–116, 2011.

Mitchell, T.M., *Machine Learning*. New York, USA: McGraw-Hill, 1997.

Moya, M.; Koch, M. & Hostetler, L., One-class classifier networks for target recognition applications. In: *Proceedings on World Congress on Neural Networks*. p. 797–801, 1993.

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

Nie, J.; Xue, Z.; Liu, T.; Young, G.S.; Setayesh, K.; Guo, L. & Wong, S.T., Automated brain tumor segmentation using spatial accuracy-weighted hidden Markov random field. *Computerized Medical Imaging and Graphics*, 33(6):431–441, 2009.

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

Platt, J.C., Fast training of support vector machines using sequential minimal optimization, Cambridge, USA: MIT Press. p. 185–208.

Prut, L. & Belzung, C., The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review. *European Journal of Pharmacology*, 463(1–3):3–33, 2003.

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

Rumelhart, D.E.; Hinton, G.E. & Williams, R.J., Learning internal representations by error propagation, Cambridge, USA: MIT Press, v. 1. p. 318–362.

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

Solomon, C. & Breckon, T., *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. New York, USA: Wiley-Blackwell, 2010.

Spagnolo, P.; Orazio, T.; Leo, M. & Distante, A., Moving object segmentation by background subtraction and temporal analysis. *Image and Vision Computing*, 24(5):411–423, 2006.

Stanford, S.C., The open field test: reinventing the wheel. *Journal of Psychopharmacology*, 21(2):134–135, 2007.

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

Vapnik, V.N., *Statistical Learning Theory*. New York, USA: Wiley-Interscience, 1998.

Vincent, L. & Soille, P., Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:583–598, 1991.

Walsh, R.N. & Cummins, R.A., The open-field test: a critical review. *Psychological Bulletin*, 83(3):482–504, 1976.

Wang, Z.; Jensen, J.R. & Im, J., An automatic region-based image segmentation algorithm for remote sensing applications. *Environmental Modelling & Software*, 25(10):1149–1165, 2010.

Yang, M.H.; Kriegman, D.J. & Ahuja, N., Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:34–58, 2002.

Notas Biográficas

Bruno Brandoli Machado é graduado em Engenharia de Computação (Universidade Católica Dom Bosco – UCDB, 2007) e tem mestrado em Ciência da Computação (Universidade de São Paulo – USP, 2010). Atualmente é doutorando no Instituto de Ciências Matemáticas e de Computação (ICMC/USP) e participa do GCI – Grupo de Computação Interdisciplinar do Instituto de Física de São Carlos. Seus interesses atuais em pesquisa incluem projeção de dados e redução de dimensionalidade, descrição por características locais, sistemas multiagentes, redes complexas e equações diferenciais parciais não lineares.

Wesley Nunes Gonçalves é graduado em Engenharia de Computação (UCDB, 2007) e tem mestrado em Ciência de Computação (USP, 2010). Atualmente é doutorando no programa de pós-graduação em Física Computacional da USP. Seus tópicos de interesses em visão computacional incluem modelagem de texturas estáticas e dinâmicas, segmentação e rastreamento de objetos, sistemas multiagentes e redes complexas.

Hemerson Pistori é bacharel e mestre em Ciência da Computação (UFMS e UNICAMP, respectivamente) e doutor em Engenharia da Computação (USP). O Professor Hemerson foi um dos fundadores do departamento de Engenharia de Computação da UCDB, tendo ocupado o cargo de chefe de departamento entre 1998 e 2001. Também ocupou as funções de presidente do comitê científico, diretor de pesquisa e desde 2009 é o Pró-Reitor de Pesquisa e Pós-Graduação desta instituição, além de docente permanente do programa de pós-graduação em Biotecnologia. Atualmente coordena o grupo de pesquisa, desenvolvimento e inovação em visão computacional, INOVISAO, da UCDB, onde trabalha como professor e pesquisador desde 1993.

Jonathan de Andrade Silva é graduado em Engenharia de Computação (UCDB, 2007) e mestre em Ciência da Computação (USP, 2010). Atualmente é doutorando no Instituto de Ciências Matemáticas e de Computação da USP. Seus tópicos de interesse incluem computação evolutiva, processamento de imagens e mineração de dados.

Kleber Padovani de Souza é graduado em Engenharia de Computação (2006, UCDB) e em Tecnologia em Processamento de Dados (2004, UNIDERP) e mestre em Ciência da Computação (2010, UFMS). Atualmente é professor universitário em cursos de graduação (UCDB) e pós-graduação (FESCG, UCDB-TNT) e realiza pesquisas em visão computacional como membro do Grupo de Pesquisa em Engenharia e Computação (GPEC/UCDB).

Bruno Toledo é aluno de graduação em Engenharia de Computação UCDB e membro do grupo INOVISAO. Seus tópicos de interesse incluem segmentação de imagens e reconhecimento de gestos.

Wesley Tessaro é aluno de graduação em Engenharia de Computação UCDB e membro do grupo INOVISAO. Seus tópicos de interesse incluem segmentação de imagens e análise de textura.

Capítulo 17

Arcabouço Computacional para Segmentação e Restauração Digital de Artefatos em Imagens Frontais de Face

André Sobiecki*, Gilson Antonio Giraldi, Luiz Antonio Pereira Neves, Gilka Jorge Figaro Gattás e Carlos Eduardo Thomaz

Resumo: Há uma variedade de aplicações que usa a imagem de face como uma informação relevante. Dependendo da aplicação, como, por exemplo, no problema de identificação de pessoas desaparecidas com base em fotos, as imagens faciais podem tornar-se ruidosas para digitalização devido à baixa qualidade do papel fotográfico e a presença de artefatos. Este capítulo propõe um arcabouço computacional para a verificação, segmentação e restauração de artefatos em imagens digitais frontais de face. Para a verificação, utiliza-se o método de índice de qualidade estrutural, e para a segmentação e a restauração são utilizados respectivamente métodos de decisão estatística e de *inpainting* tradicionais. Os resultados mostram que todas as imagens processadas pelo arcabouço computacional proposto apresentam uma melhora significativa de qualidade..

Palavras-chave: Processamento de imagens, Decisão estatística, *Inpainting*, Reconhecimento de Padrões.

Abstract: *There is a variety of applications that has used the face image as a relevant information. 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 artifacts. In this work, we propose a computational framework for verification, segmentation and automatic restoration of these frontal face image artifacts. For verification, an index of image quality has been used, for segmentation statistical decision methods and for restoration inpainting techniques. Our results show that all the images processed by the framework proposed have had a significant improvement in their digitalization quality.*

Keywords: *Image processing, Statistical Decision, Inpainting, Pattern Recognition.*

*Autor para contato: andresobiecki-sbs.sc@hotmail.com

Neves et al. (Eds.), Avanços em Visão Computacional (2012) DOI: 10.7436/2012.avc.17 ISBN 978-85-64619-09-8

1. Introdução

O reconhecimento facial é o processo de obtenção da identidade de uma pessoa com base em informações obtidas a partir da aparência facial (Ayinde & Yang, 2002). Na área da segurança, a maioria dos sistemas de fácil utilização pode garantir o nosso patrimônio e proteger a nossa privacidade por senhas numéricas ou alfanuméricas. Há outras formas de proteção baseadas em métodos de idenficação biométrica por impressão digital, análise da retina ou iris dos olhos, e sistemas de reconhecimento de voz (Zhao et al., 2003; Hjelmas & Low, 2001; Zhao et al., 2003). Porém estes métodos exigem a colaboração do usuário enquanto que os sistemas de reconhecimento facial possuem a vantagem de ser um método transparente para o usuário, onde não é necessário a aprovação e nem a colaboração explícita do indivíduo. Na verdade, o reconhecimento facial é a forma mais comum para as pessoas se identificarem entre si.

Nos últimos anos, o reconhecimento facial tem recebido atenção considerável das comunidades científicas e comerciais, mas ainda há muitos desafios no desenvolvimento destas aplicações. Um dos desafios é a questão da imagem facial estar em baixa qualidade afetando o desempenho do sistema, especialmente nas etapas automáticas de detecção e identificação faciais. Os problemas de qualidade observados, tais como sombras, artefatos, ofuscamento e borrões são comuns e afetam este tipo de reconhecimento (Zamani et al., 2008; Zhao et al., 2003; Castillo, 2006).

Há na atualidade diversas 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 (REDESAP (2010)) disponibiliza publicamente imagens faciais de pessoas sendo que muitas destas fotos são antigas e possuem rasuras como dobraduras, arranhões, luminância irregular, bolor, carimbos e escritos diversos sobre a imagem digitalizada. Na prática, estas rasuras dificultam o reconhecimento facial automático, pois uma rasura sobre uma imagem facial pode estar em diversas cores, tons, formatos e tamanhos. Uma vez que 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 (Jun et al., 2011) é de fundamental importância que a imagem a ser utilizada em um processo automático de reconhecimento facial apresente as características faciais discriminantes com o mínimo de artefatos possível.

Este capítulo propõe e implementa um arcabouço computacional envolvendo segmentação e restauração automática de rasuras baseado em um modelo estatístico construído a partir de imagens faciais frontais e em técnicas de *inpainting*. 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 pelos métodos de *inpainting* propostos em Oliveira et al. (2001), Bertalmio et al. (2001) e Telea (2004).

2. Arcabouço Computacional

A Figura 1 apresenta o diagrama do arcabouço computacional proposto neste artigo para identificação e eliminação de rasuras automaticamente.

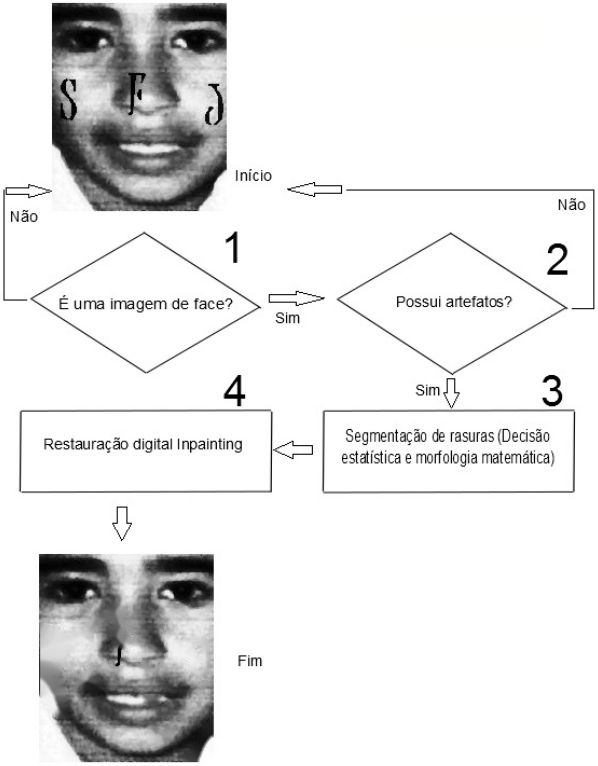


Figura 1. Diagrama da metodologia proposta.

Todas as imagens são previamente equalizadas e pré-processadas (Amaral et al., 2009). Em seguida, de acordo com o diagrama da Figura 1, é verificado se a imagem de entrada é uma face e se possui artefatos. Para isto, é utilizado um método de índice de qualidade estrutural. Uma vez identificado que trata-se de uma imagem de face com artefatos, aplica-se a etapa seguinte onde é utilizado o método de decisão estatística para a segmentação das rasuras. Finalmente, segue-se a ultima etapa do fluxo de dados da Figura 1 onde é executada a restauração digital via métodos de

inpainting. Todas estas etapas do arcabouço proposto são explicadas nas seções seguintes.

2.1 Índice de qualidade estrutural

O índice de medida de qualidade proposto por Wang et al. (2004) possui medidas referentes à luminância $l(\mathbf{x},\mathbf{y})$, contraste $c(\mathbf{x},\mathbf{y})$ e estrutural $s(\mathbf{x},\mathbf{y})$, conforme as equações 1, 2 e 3, respectivamente:

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2}, \tag{1}$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}, \tag{2}$$

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy}}{\sigma_x\sigma_y}. \tag{3}$$

onde $\mathbf{x} = \{x_i | i = 1,2,...,n\}$ e $\mathbf{y} = \{y_i | i = 1,2,...,n\}$ são sinais n-dimensionais referentes às imagens. As variáveis estatísticas são calculadas como:

$$\mu_x = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \mu_y = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{4}$$

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2, \qquad \sigma_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\overline{y})^2 \tag{5}$$

$$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}), \tag{6}$$

Os índices $l(\mathbf{x},\mathbf{y})$ e $c(\mathbf{x},\mathbf{y})$ podem apresentar valores entre 0 e 1 e o índice $s(\mathbf{x},\mathbf{y})$ entre -1 e 1. Para imagens normalizadas e equalizadas, tais como as utilizadas neste trabalho (Amaral et al., 2009) os índices de qualidade de luminância e de contraste apresentam o valor máximo igual a 1, pois todas têm valores de média e desvio muito similares. No entanto, os valores de covariância variam entre imagens normalizadas e equalizadas, e esta medida de qualidade pode ser usada. Embora $s(\mathbf{x},\mathbf{y})$ não utilize uma representação descritiva explícita da estrutura de uma imagem, este índice será igual a 1 se e somente se as duas imagens de comparação forem exatamente iguais (informações estruturais são todos os atributos da imagem que não sejam referentes às informações de luminância e de contraste) (Wang et al., 2004).

Para descrever características de interesse entre as imagens, utiliza-se a medida de qualidade estrutural definida pela equação (3) pois, além de investigar a qualidade digital de cada imagem, também há interesse em implementar uma medida que informe a existência de artefatos sobre uma imagem facial frontal utilizando como referência uma imagem padrão.

Há também interesse em aplicar um método que informe se a imagem de entrada é ou não uma imagem de face.

2.2 Decisão estatística

Para a segmentação de artefatos utiliza-se um método de decisão estatística baseado na teoria de inferência estatística (Spiegel & Stephens, 2008; Bussab & Morrettin, 2002), onde amostras populacionais podem gerar informações a priori e a partir destas informações é possível tomar decisões dado um nível α de significância estatística.

A partir de uma população amostral de 385 imagens frontais com expressão facial neutra e normalizadas espacialmente (Amaral et al., 2009), calcula-se a imagem média destas amostras através da seguinte equação:

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x_i}, \tag{7}$$

onde $\mathbf{x_i}$ é o vetor n-dimensional que representa a concatenação de todos os pixeis da imagem de face i, e N o número total de amostras. Neste trabalho, $N = 385$. Para verificar quão válida seria a afirmação de que a média amostral se aproximaria da média da população de faces frontais com expressão facial neutra, calcula-se o intervalo de 99,9% de confiança (IC) desta estimativa, ou seja:

$$99,9\%IC = \overline{\mathbf{x}} \pm 3,291\frac{\sigma}{\sqrt{N}}, \tag{8}$$

onde σ é o desvio-padrão das amostras.

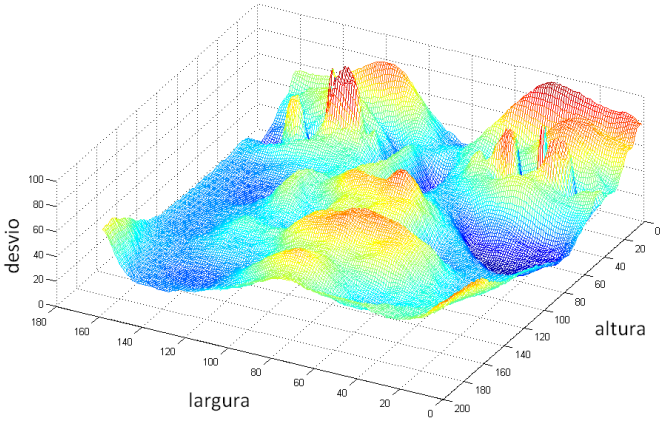
As Figuras 2(c) e 2(d) ilustram as imagens que correspondem aos limites inferior e superior do intervalo de confiança descritos na equação (8). Observa-se que visualmente as imagens são muito parecidas, pois possuem correlação de 0,998942 e 0,998962 com a imagem média amostral.

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 $\mathbf{z}$ da diferença entre uma imagem de face $\mathbf{x}$ e a imagem da média amostral $\overline{\mathbf{x}}$, supondo que a distribuição das tonalidades dos pixeis segue uma densidade de probabilidade Gaussiana com média nula e variância 1, isto é:

$$\mathbf{z} = \frac{\mathbf{x} - \overline{\mathbf{x}}}{\sigma} \tag{9}$$

Quanto maior for o valor absoluto de z, maior será a significância estatística desta diferença, pixel a pixel. As Figuras 2(a) e 2(e) mostram a imagem média e o gráfico dos valores de desvio-padrão de todos os pixeis utilizados na equação (9) para a identificação das rasuras.

(a) (b) (c) (d)



(e)

Figura 2. Ilustração das seguintes imagens: (a) Imagem média, (b) Imagem mediana, (c) Inferência estatística positiva, (d) Inferência estatística negativa e (e) Gráfico dos valores de desvio.

2.3 Operações morfológicas

Para corrigir as imperfeições das segmentações baseadas na decisão estatística, 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 (Aptoula & Lefèfre, 2007; Santiago et al., 2009; Gonzalez & Woods, 2000). As informações extraídas de uma imagem são obtidas pela sua transformação a partir de um conjunto bem definido chamado elemento estruturante, computacionalmente determinado por uma matriz e um ponto central.

Uma operação morfológica é determinada pela iteração de um dado elemento estruturante B em uma imagem I. A cada iteração é realizada uma operação morfológica local na imagem com região de suporte dada pelas dimensões do elemento estruturante. Ao final de todo este processo é obtida uma nova imagem R.

Computacionalmente, a imagem R de saída é inicializada com *pixels* inativos, ou seja, pretos. Os resultados das operações morfológicas locais são guardados na imagem R, enquanto a imagem I permanece intacta. Caso contrário, uma operação morfológica local seria comprometida pelas anteriores.

Existem duas operações básicas na morfologia matemática denominadas dilatação e erosão. Seja z a posição de um pixel na imagem I e B_z a translação do elemento estruturante B para o ponto z, pode-se então definir:

- Dilatação da Imagem binária I pelo Elemento Estruturante B:

$$R = I \oplus B = \{z | I \cap B_z \neq \varnothing\}. \tag{10}$$

- Erosão da Imagem binária I pelo Elemento Estruturante B:

$$R = I \ominus B = \{z | B_z \subset I\}. \tag{11}$$

2.4 Restauração digital

Para restauração digital das imagens de face são utilizados métodos de *inpainting* digital que são métodos computacionais que permitem restaurar regiões da imagem onde ocorreu a perda de informação decorrente de rasuras, borrões e carimbos, por exemplo. Na primeira metade dos anos 90 já se encontravam técnicas deste tipo, utilizando modelos do tipo AR (*Auto-Regressive models*) e Markovianos para restauração de sequências de vídeo (Kokaram et al., 1995)). Mais recentemente, foram encontrados métodos de *inpainting* baseados em métodos de interpolação e difusão, modelos variacionais e equações diferenciais parciais (Chan & Shen, 2005).

A escolha da técnica adequada depende do tipo da imagem e das características da rasura. Por exemplo, para a restauração de regiões pequenas, pode-se aplicar a técnica proposta por Oliveira et al. (2001), baseada em técnicas de difusão. O método proposto em Telea (2004), utiliza também a idéia de difundir o campo de intensidades da fronteira da região afetada em direção ao interior da rasura. Bugeau & Bertalmio (2009) utilizaram também um método de difusão, mas combinando síntese de textura, para realizar o *inpainting*.

Métodos baseados em equações diferenciais parciais foram extensamente aplicados em processamento de imagens para restauração, filtragem e métodos multiscala (Chan & Shen, 2005). No caso do *inpainting*, um tipo particular de restauração, foram desenvolvidas técnicas baseadas nas

equações de Laplace e Poisson (Perez et al., 2003; Jeschke et al., 2009), difusão anisotrópica e difusão de curvatura (Chan & Shen, 2001), síntese de textura e equações diferenciais parciais (Bugeau et al. (2009), equações de Navier-Stokes (Bertalmio et al., 2001), dentre outros.

Neste trabalho, optou-se pelos métodos descritos nas seções seguintes. O primeiro deles, proposto em Oliveira et al. (2001), foi escolhido por sua eficiência computacional e facilidade de implementação. O segundo método, apresentado em Telea (2004), é um pouco mais custoso e sua implementação é mais elaborada. Porém, sua aplicação não esta limitada a regiões pequenas. Finalmente, o método descrito em Bertalmio et al. (2001) é computacionalmente o mais custoso entre os três, porém está melhor fundamentado do ponto de vista teórico.

2.5 *Inpainting* via difusão

No caso de regiões pequenas, a alternativa mais direta para realizar o *inpainting* é aplicar um método para difundir (borrar) o campo de intensidades sobre a fronteira da região rasurada de tal forma a recuperar a imagem no seu interior, como proposto em Oliveira et al. (2001).

O processo de difusão pode ser realizado pela convolução da imagem original com um filtro passa-baixas, como o filtro Gaussiano, por exemplo. No caso discreto, este processo é definido pela seguinte equação:

$$v\left(m,n\right) = I\left(m,n\right) \circledast h\left(m,n\right) = \sum_{(s,w)\in V} I\left(s,w\right) h\left(m-s,n-w\right), \quad (12)$$

onde I representa a imagem original, h o núcleo do filtro, V uma vizinhanca da origem (por exemplo: $V = \{(s,w)\}\,; \quad -1 \leq s,w \leq 1$) e v a imagem filtrada. Na Tabela 1 são apresentados dois filtros usados em (Oliveira et al., 2001)).

a	b	a		c	c	c
b	0	b	e	c	0	c
a	b	a		c	c	c

Tabela 1. Filtros passa-baixa utilizados na região a ser feito o *inpainting*, onde a = 0,073235, b = 0,176765, c = 0,125.

Na aplicação deste procedimento, deve-se ter o cuidado para não borrar a imagem no exterior da região de interesse. Neste sentido, a determinação da fronteira da região de *inpainting* Ω é fundamental para definir barreiras para o processo de difusão (*diffusion barriers*, como denominado em Oliveira et al. (2001). Basicamente, os pixeis da fronteira sao marcados e uma vez atingido um destes pixeis, o processo de difusão é interrompido.

Como o sistema visual humano tolera quantidades moderadas de borramento em áreas distantes de bordas com alto contraste, pode ser necessário aplicar mais de uma vez a convolução definida na equação (12) para correção de regiões pequenas e obtenção da qualidade desejada na restauração. Por exemplo, em Oliveira et al. (2001) há testes onde aplica-se 100 vezes este processo.

2.6 Interpolação e *fast marching* para *inpainting*

Telea (2004) propõe um algoritmo de *inpainting* que parte também da idéia de difundir o campo de intensidades na fronteira da região afetada. Contudo, neste caso o processo é baseado em uma aproximação de primeira ordem $I_q(p)$ do valor da imagem em um ponto p da região Ω a ser corrigida:

$$I_q(p) = I(q) + \nabla I(q) \cdot (p - q), \tag{13}$$

onde q é um pixel em uma vizinhanca de p de raio $\in (B_\in(p))$, e ∇ representa o gradiente da imagem em q, definido por:

$$\nabla I(q) = \left(\frac{\partial I}{\partial x}(q), \frac{\partial I}{\partial y}(q) \right). \tag{14}$$

O *inpainting* feito no ponto p considera todos os pontos q presentes na vizinhanca $B_\in(p)$, tomando como a intensidade final $I(p)$ uma média ponderada dos valores $I_q(p)$ obtidos pela expressão (13), ou seja:

$$I(p) = \frac{\sum_{q \in B_\in(p)} w(p,q)[I(q) + \nabla I(q) \cdot (p - q)]}{\sum_{q \in B_\in(p)} w(p,q)}, \tag{15}$$

onde os pesos $w(p,q)$ são dependentes da aplicação e devem ser normalizados de acordo com a equação $\sum_q w(p,q) = 1$.

O primeiro passo é a extração da fronteira da região onde será aplicada o *inpainting*, a qual não necessita ser pequena neste caso. Um ponto que fica mais evidente agora é que o procedimento usado para avançar da fronteira para o interior da região de *inpainting* pode interferir na qualidade do resultado final. Em Telea (2004) este avanço é controlado pelo método de *fast marching* (FMM) o qual é baseado na solução da equação Eikonal:

$$|\nabla T| = 1, \quad T = 0 \quad em \quad \partial\Omega, \tag{16}$$

onde T é um campo auxiliar. A solução da equação (16) gera um mapa de distâncias dos pixeis de Ω em relação a fronteira $\partial\Omega$ (Sethian, 1996). Uma curva de nível deste campo, definida por:

$$S_k = \{p \in \Omega; \quad T(p) = k\}$$

é o lugar geométrico dos pontos $p \in \Omega$ cuja distância da fronteira é k.

Desta forma, o método de *inpainting* proposto em Telea (2004) avança aplicando a expressão (15) para cada curva de nível do campo T, partindo da fronteira $\partial\Omega = S_0$. Assim, pode-se garantir que áreas mais próximas da fronteira são corrigidas antes das áreas mais distantes, evitando a geração de artefatos e melhorando a aparência do resultado final.

2.7 Navier-Stokes *framework* para *inpainting*

O método proposto em Bertalmio et al. (2001) tem inspiração em modelos $2D$ da mecânica de fluídos incompressiveis, onde é possivel converter as equações de Navier-Stokes em uma equação diferencial envolvendo um campo escalar $\Psi = \Psi(x, y)$, denominado função de corrente.

A idéia básica do método de *inpainting* apresentado em Bertalmio et al. (2001) é considerar a intensidade de imagem I como a função de corrente. Seja então a velocidade $\mathbf{v} = (v_1(x, y), v_2(x, y))$, a viscosidade μ e a pressão $p = p(x, y)$ de um fluído incompressível. Neste caso, os campos envolvidos devem satisfazer a equação de Navier-Stokes:

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\nabla \mathbf{p} + \mu \Delta \mathbf{v}, \tag{17}$$

sujeito à condição de incompressibilidade:

$$\nabla \cdot \mathbf{v} = \mathbf{0}, \tag{18}$$

onde o operador gradiente (∇) está definido pela expressão (14) e os operadores divergente ($\nabla\cdot$) e Laplaciano (Δ) são definidos pelas equações:

$$(\nabla\cdot) = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}, \quad \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \tag{19}$$

Uma vez que o fluido é bidimensional e incompressivel, pode-se mostrar que existe uma campo escalar Ψ, a função de corrente associada, que satisfaz:

$$\nabla^{\perp}\Psi = \left(-\frac{\partial \Psi}{\partial y}, \frac{\partial \Psi}{\partial x}\right) = \mathbf{v}, \quad \nabla \times \mathbf{v} = \left(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}\right)\mathbf{z} = (\Delta\Psi)\,\mathbf{z}, \tag{20}$$

onde o operador ($\nabla\times$) é denominado rotacional do campo e $\mathbf{z}$ representa o terceiro vetor da base canônica $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. O campo vetorial obtido pelo rotacional da velocidade é denominado vorticidade $\omega = \nabla \times \mathbf{v}$, e se reduz à expressão acima para o caso $2D$. Tomando agora o rotacional da equação (17), observando que $\nabla\times(\nabla\mathbf{p}) = 0$, e aplicando as propriedades do cálculo vetorial, obtem-se a equação para o vorticidade do fluído, dada por:

$$\frac{\partial \omega}{\partial t} + \mathbf{v} \cdot \nabla\omega = \mu\Delta\omega. \tag{21}$$

Usando-se as expressões definidas em (20), fica evidente que a solução estacionária ($\frac{\partial \omega}{\partial t} = 0$) da equação (21) para viscosidade nula ($\mu = 0$), satisfaz:

$$\nabla^{\perp}\Psi \cdot \nabla\left(\boldsymbol{\Delta\Psi}\right) = 0. \tag{22}$$

Considerando-se a intensidade de imagem como a função de corrente, ou seja $I\left(x,y\right) = \Psi\left(x,y\right)$, então a expressão (22) fornece a condição para a solução estacionária da equação:

$$I_t = \nabla^{\perp} I \cdot \nabla \Delta I. \tag{23}$$

A expressão (23) foi utilizada em Bertalmio et al. (2000) e sua solução estacionária indica que as isolinhas do campo de intensidades (*isophotes*) são paralelas às curvas de nivel do campo ΔI. Adicionando à expressão (23) um termo de difusão anisotrópica da imagem geramos a equação diferencial parcial:

$$I_t = \nabla^{\perp} I \cdot \nabla \Delta I + \mu \nabla \cdot (g(|\nabla I|)\nabla I), \tag{24}$$

onde o parâmetro μ controla agora a influência do termo difusão. Em Bertalmio et al. (2001) a equação (24) é usada como inspiração para criar um modelo de *inpainting* onde o termo de difusão da vorticidade ($\Delta\omega$) na equação (21) é substituido pela difusão anisotrópica, gerando uma expressão analoga a equação (24), mas para a vorticidade:

$$\frac{\partial \omega}{\partial t} + \mathbf{v} \cdot \nabla \omega = \mu \nabla \cdot (\mathbf{g}(|\nabla \omega|)\nabla \omega), \tag{25}$$

onde:

$$\mathbf{v} = \nabla^{\perp} I. \tag{26}$$

Uma vez resolvida a equação (25), recupera-se o campo de intensidades desejado via:

$$\omega = \Delta I, \quad I\,|_{\partial\Omega} = I_0. \tag{27}$$

O estudo das equações de Navier-Stokes dadas pelas expressões (17) e (21) é bem fundamentado na literatura, tanto do ponto de vista numérico quanto teórico (existência e unicidade de soluções). Por outro lado, a relação com elementos da dinâmica de fluídos pode tornar mais intuitiva a análise do resultado do *inpainting*. Estas são vantagens do método exposto acima, embora o efeito do termo de difusão anisotrópica na equação (25) para valores moderados de μ seja um ponto ainda em aberto na teoria.

Para resolver numericamente as equações (25)-(27) é necessário impor condições de fronteira e iniciais para o campo de intensidades em Ω. Neste trabalho, as condições de fronteira são obtidas pelos pixeis vizinhos das áreas de rasura.

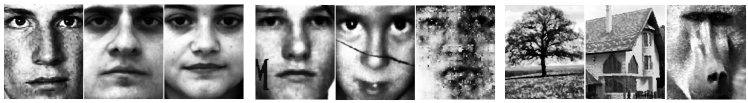
3. Experimentos e Resultados

Para gerar a imagem média e os valores correspondentes de desvio, foram utilizadas 385 imagens frontais com expressão neutra e sem artefatos (óculos, rasuras, etc.) da base FEI (Thomaz & Giraldi, 2010) e da base FERET (Phillips et al., 2000). A Figura 3 ilustra quatro imagens de cada base. Na base FEI, a maioria dos indivíduos são caucasianos, enquanto que na base FERET há uma variedade de raças.

(a) (b)

Figura 3. Exemplos de imagens de face utilizadas:(a) FEI; (b) FERET.

Para a validação do índice de qualidade estrutural de imagens, foram utilizadas três bases contendo 30 imagens cada uma compostas por imagens frontais de face sem artefatos, com artefatos e imagens diversas (animal, casa, árvore, carro, entre outros). Para a base de imagens de face com artefatos, foram obtidas imagens do site de crianças e adolescentes desaparecidos do Brasil, disponibilizadas publicamente, e de voluntários pertencentes ou não a base FEI. Há rasuras como borrão e dobras, cabelo sobre o rosto, armação de óculos, sorriso, chupeta, má qualidade de resolução e carimbo. As imagens são normalizadas e equalizadas pelo sistema desenvolvido por Amaral et al. (2009). A Figura 4 mostra exemplos das três bases.



(a) (b) (c)

Figura 4. Ilustração de imagens das bases de validação: (a) Exemplos de imagens de face sem artefatos; (b) Exemplos de imagens de face com artefatos; e (c) Exemplos de imagens não-face.

Para os métodos de *inpainting*, são utilizadas as funções disponíveis na biblioteca do OpenCV onde as condições iniciais e de fronteira são informadas por meio das imagens de entrada. Por simplicidade, o método de *inpainting* através de interpolação foi implementado usando o resultado da tranformada de distância para percorrer os pixeis da região de interesse. Utilizamos também a tranformada de distância para orientar a aplicação do método de *inpainting* via difusão.

A Figura 5 apresenta a dispersão do índice de qualidade estrutural em relação à imagem média. Pode-se observar que o *boxplot* da base de

imagens de face sem artefatos apresenta um *outlier* que corresponde a uma imagem de face de uma pessoa de raça negra com labios mais largos e também com a posição dos labios mais acima do que a maioria das outras imagens. Para o teste de índice de qualidade estrutural na base de **imagens com artefatos**, a imagem da criança com a chupeta apresenta o menor valor de índice de qualidade estrutural, 0,131. A imagem rasurada que apresenta o maior valor é uma imagem da base FEI que possui como artefato um par de óculos com uma armação de metal fina e que cobre apenas uma pequena região do nariz, 0,837. Nos valores de índice de qualidade estrutural para a base de *imagens não-face*, a imagem do galo e a imagem do macaco apresentam os maiores valores, 0,363 e 0,329, respectivamente. No caso da imagem do galo, a razão pode estar em possuir um contraste similar às imagens de face utilizadas, entre o centro da imagem (mais claro) e as demais regiões (mais escuras). Para a imagem do macaco, a justificativa é mais evidente e se baseia no fato da mesma possuir características estruturais de face semelhantes a de uma imagem de face humana.

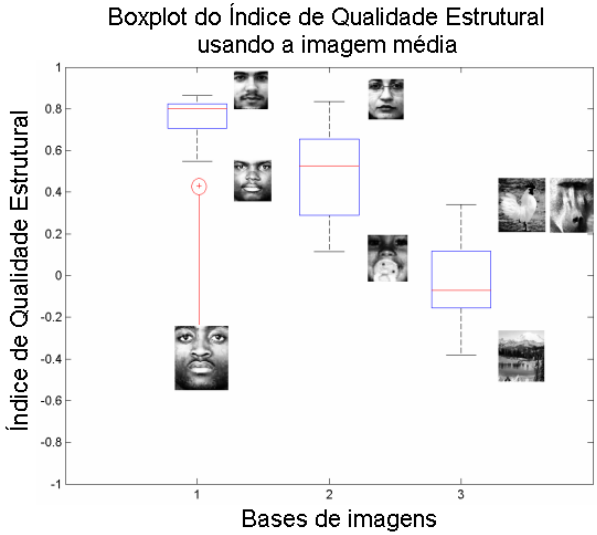


Figura 5. Gráfico *boxplot* de dispersão estatística das imagens de face sem artefatos, com artefatos, e não-faces com relação à imagem média.

De acordo com os resultados da Figura 5, caso o valor do índice de qualidade estrutural esteja abaixo do primeiro quartil da base de imagens de face sem artefatos e acima do terceiro quartil da base de imagens não-faces, podemos considerar que este valor se refere a uma imagem de face com artefatos e de interesse para segmentação e restauração automática.

A Figura 6 ilustra cinco dos 30 resultados obtidos. Nestas cinco imagens há artefatos como carimbo, pintas, óculos e chupeta, conforme mostra a coluna (a). A coluna (b) ilustra a segmentação destes artefatos para quatro níveis distintos de significância estatística, variando de 90% (1,645) com segmentação em cor verde, 95% (1,96) com segmentação em cor amarela, 99% (2,58) com segmentação em cor azul, até 99,9% (3,291) com segmentação em cor vermelha. Pode-se notar que artefatos como óculos e carimbo, e características incomuns, são segmentadas mesmo com um alto nível de significância. A coluna (c) da Figura 6 mostra a máscara binária necessária para processamento dos métodos de *inpainting*, apresentados nas colunas (d), (e) e (f). É importante ressaltar a última linha da Figura 6 que mostra um caso de insucesso, onde o artefato é uma chupeta. Este artefato ocupa uma região muito grande da face de forma que os métodos de *inpainting* não conseguem obter um resultado plausível.

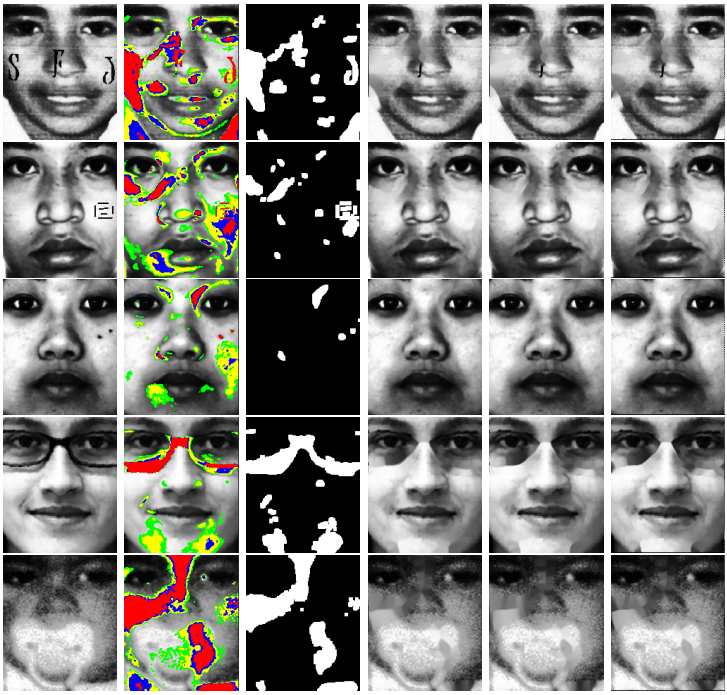
Por fim, a Figura 7 mostra a qualidade estrutural das imagens restauradas em comparação com as imagens originais. O valor médio e o valor de desvio-padrão dos índices de qualidade estrutural das imagens originais são 0,4605 e 0,2099. Para as imagens restauradas pelo método proposto por Bertalmio et al. (2001), tem-se 0,5484 e 0,1662, para as imagens restauradas pelo método proposto por Oliveira et al. (2001) tem-se 0,5444 e 0,1664, e para as imagens restauradas pelo método proposto por Telea (2004) tem-se 0,5521 e 0,1630. Na maioria das imagens testadas, o método de *inpainting* proposto por Telea (2004) obteve os maiores valores de índice de qualidade estrutural.

4. Conclusão e Trabalhos Futuros

Cada imagem com artefatos foi comparada com uma imagem de referência para um nível de significância estatística que pode variar de 90% a 99,9% com relação aos defeitos na imagem. Para resultados menos conservadores, pode-se assumir valores de significância estatística menores. No entanto, nesta situação mais regiões serão necessariamente identificadas como críticas podendo descaracterizar a singularidade de cada imagem de face.

Na maioria dos testes realizados, os métodos de *inpainting* estudados obtiveram resultados promissores, principalmente se a região de artefato for relativamente pequena em relação ao tamanho da imagem. O método de *inpainting* proposto por Telea (2004) obteve os maiores valores de índice de qualidade estrutural, portanto, quantitativamente, podemos considerar este método o mais adequado entre os três para as imagens das bases utilizadas, embora os resultados dos métodos de *inpainting* não apresentem diferenças visíveis.

Como trabalhos futuros, devido à diversidade biométrica facial da população brasileira e considerando os resultados promissores deste estudo, pode-se propor a criação de modelos estatísticos de face que sejam mais es-



(a) (b) (c) (d) (e) (f)

Figura 6. Exemplos de cinco imagens utilizadas nos testes. (a) imagem original, (b) segmentação estatística, (c) Imagem binária, (d) *Inpainting* com método proposto por Telea (2004) (Seção 2.6), (e) *Inpainting* via equação de Navier Stokes (Seção 2.7) e (f) *Inpainting* usando método proposto por Oliveira et al. (2001)(seção 2.5).

pecíficos 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*.

Para imagens frontais de face que contenham sorrisos largos ou chupetas talvez fosse válido a investigação de métodos de edição de imagens, tais como o método de Poisson proposto por Perez et al. (2003), que considera informações a priori sobre o padrão de interesse a ser restaurado ao invés somente de informações extraídas de intensidades de pixeis vizinhos.

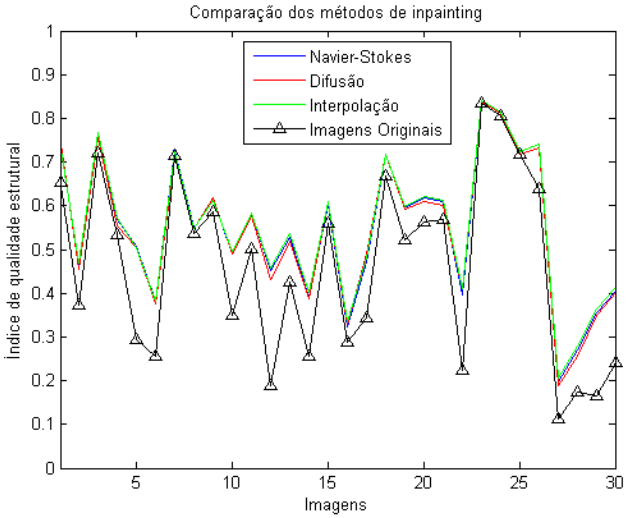


Figura 7. Índice de qualidade estrutural dos resultados em comparação com as imagens de entrada.

Agradecimentos

Os autores agradecem à CAPES (modalidade PROCAD, processo 094/2007), ao CNPq (bolsa de mestrado, processo 148531/2010-5), à LIM-40-FMUSP e à FAPESP (processo 09/53556-0) pelo apoio financeiro.

Referências

Amaral, V.; Figaro-Garcia, C.; Gattas, G.J.F. & Thomas, C.E., Normalização espacial de imagens frontais de face. *FaSCi-Tech*, 1(1), 2009.

Aptoula, E. & Lefèfre, S., A comparative study on multivariate mathematical morphology. *Pattern Recognition*, 40(11):2914–2929, 2007.

Ayinde, O. & Yang, Y.H., Face recognition approach based on rank correlation of Gabor-filtered images. *Pattern Recognition*, 35(6):1275–1289, 2002.

Bertalmio, M.; Sapiro, G. & Bertozzi, A.L., Navier-Stokes, fluid dynamics, and image and video inpainting. In: *Proceedings of IEEE Computer Vision and Pattern Recognition*. v. 1, p. I355 – I362, 2001.

Bertalmio, M.; Sapiro, G.; Caselles, V. & Ballester, C., Image inpainting. In: *Proceedings 27th International Conference on Computer Graphics and Interactive Techniques*. ACM, p. 61–69, 2000.

Bugeau, A. & Bertalmio, M., Combining texture synthesis and diffusion for image inpainting. In: *International Conference on Computer Vision Theory and Applications*. p. 28–33, 2009.

Bugeau, A.; Bertalmio, M.; Caselles, V. & Sapiro, G., *A Unifying Framework for Image Inpainting*. IMA Preprint series 2262, IMA Press, Minneapolis, USA, 2009.

Bussab, W.O. & Morrettin, P.A., *Estatística Básica*. São Paulo, SP: Editora Saraiva, 2002.

Castillo, O.Y.G., *Survey About Facial Image Quality*. Report, Fraunhofer Institute for Computer Graphics Research, Darmstadt, Germany, 2006.

Chan, T.F. & Shen, J., Non-texture inpainting by curvature-driven diffusions (CDD). *Journal of Visual Communication and Image Representation*, 12(4):436–449, 2001.

Chan, T.F. & Shen, J., *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. Philadelphia, USA: SIAM, 2005.

Gonzalez, R.C. & Woods, R.E., *Processamento de Imagens Digitais*. São Paulo, SP: Editora Edgard Blucher Ltda, 2000.

Hjelmas, E. & Low, B.K., Face detection: A survey. *Computer Vision and Image Understanding*, 83(3):236–274, 2001.

Jeschke, S.; Cline, D. & Wonka, P., A GPU Laplacian solver for diffusion curves and Poisson image editing. *Transactions on Graphics*, 28(5):1–8, 2009.

Jun, B.; Lee, J. & Kim, D., A novel illumination-robust face recognition using statistical and non-statistical method. *Pattern Recognition Letters*, 32(2):329–336, 2011.

Kokaram, A.C.; Morris, R.D.; Fitzgerald, W.J. & Rayner, P.J.W., Interpolation of missing data in image sequences. *IEEE Transactions on Image Processing*, 4(11):1509–1519, 1995.

Oliveira, M.M.; Bowen, B.; McKenna, R. & Chang, Y., Fast digital image inpainting. In: *Proceedings of the International Conference on Visualization, Imaging and Image Processing*. ACTA Press, p. 261–266, 2001.

Perez, P.; Gangnet, M. & Blake, A., Poisson image editing. *ACM Transactions on Graphics*, 22(3):313 – 318, 2003.

Phillips, P.J.; Moon, H.; Rizvi, S.A. & Rauss, P.J., The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

REDESAP, , Ministério da Justiça; pessoas desaparecidas. Disponível em: <http://www.desaparecidos.mj.gov.br/>, 2010. Acessado em: 8 Nov. 2010.

Santiago, J.D.L.; Garmino, A.; Salgado, J.; Trujillo, V. & Ortiz, A., Operadores k-estadísticos para morfología matemática de conjuntos. *Revista Faculdad de Ingeniería Universidad de Antioquia*, 48:216–227, 2009.

Sethian, J.A., *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Materials Sciences*. Cambridge, UK: Cambridge University Press, 1996.

Spiegel, M.R. & Stephens, L.R., *Statistics*. Lisboa, Portugal: Schaum's Outlines, 2008.

Telea, A., An image inpainting technique based on the fast marching method. *Journal of Graphics, GPU, and Game Tools*, 9(1):23–34, 2004.

Thomaz, C.E. & Giraldi, G.A., A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, 28(6):902–913, 2010.

Wang, Z.; Lu, L. & Bovik, A.C., Video quality assessment based on structural distortion measurement. *Signal Processing: Image Communication*, 19(2):121–132, 2004.

Zamani, A.N.; Awang, M.K.; Omar, N. & Nazeer, S.A., Image quality assessments and restoration for face detection and recognition system images. In: *Proceedings of Second Asia International Conference on Modelling & Simulation*. IEEE Press, p. 505–510, 2008.

Zhao, W.; Chellappa, R.; Phillips, P.J. & Rosenfeld, A., Face recognition: a literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

Notas Biográficas

André Sobiecki é graduado em Tecnologia de Sistemas de Informação (UDESC, 2009), mestre em Engenharia Elétrica (Centro Universitário da FEI, 2012, com estágio sanduíche de 3 meses no Laboratório Nacional de Computação Científica), atualmente é doutorando em Ciência da Computação (Departamento de Visualização Científica e Computação Gráfica, Universidade de Groningen, Holanda). Tem interesse de pesquisa na área de processamento de imagens, reconhecimento de padrões, restauração digital *inpainting* e métodos de esqueletização 2D e 3D.

Gilson Antonio Giraldi é graduado e mestre em Matemática (Pontifícia Universidade Católica de Campinas, 1986 e Universidade Estadual de Campinas, 1993, respectivamente) e doutor em Engenharia de Sistemas e Computação (Universidade Federal do Rio de Janeiro, 2000). Atualmente é pesquisador adjunto III no Laboratório Nacional de Computação Científica. Tem experiência na área de Ciência da Computação, com ênfase em Matemática da Computação, atuando principalmente nos seguintes temas: segmentação, modelos deformáveis, processamento de imagens, *snakes* e animação de fluidos.

Luiz Antonio Pereira Neves é bacharel em Informática (Universidade Positivo, 1993), e em Letras (Universidade Tuiuti do Paraná, 1990), mestre em Informática Aplicada (Pontifícia Universidade Católica do Paraná, 1999) e doutor em Engenharia Elétrica (Universidade Federal de Campina Grande, 2006). Atualmente é membro do Instituto Nacional de Ciência e Tecnologia Medicina Assistida por Computação Científica e professor adjunto da Universidade Federal do Paraná, do Setor de Educação Profissional Tecnológica. Atua em Ciência da Computação, com ênfase em reconhecimento de padrões, visão computacional, robótica, gestão da informação, classificação e interpretação des imagens digitais, processamento de imagens médicas, odontológicas e biológicas.

Gilka Jorge Figaro Gattás é graduada em Biomédicas (Universidade de Santo Amaro, 1979), mestre e doutora em Ciências Biológicas (Universidade de São Paulo, 1986 e 1994, respectivamente) e tem pós-doutorado (Harvard Medical School, 1997-1998). atualmente é professora associada - MS5 da Faculdade de Medicina da Universidade de São Paulo, e coordenadora do projeto caminho de volta: busca de crianças desaparecidas no estado de São Paulo, é membro da Sociedade Brasileira de Mutagenese Ambiental e membro da Sociedade Brasileira de Genética. Atua na área de genética humana, com ênfase em mutagenese, ciências forenses e epidemiologia genética molecular.

Carlos Eduardo Thomaz é graduado e mestre em Engenharia Elétrica (Pontifícia Universidade Católica do Rio de Janeiro, 1993 e 1999, respectivamente), doutor em Ciência da Computação (Imperial College London, 2005) e tem pós-doutorado na mesma instituição. Atualmente é professor adjunto do Centro Universitário da FEI. Tem experiência na área de Ciência da Computação, com ênfase em reconhecimento de padrões em estatística, atuando principalmente nos seguintes temas: visão computacional, computação em imagens médicas e biometria.

Capítulo 18

Detecção de Manchas Solares Utilizando Morfologia Matemática

Adílson Eduardo Spagiari, Israel Florentino dos Santos, Wander Lairson Costa, Adriana Válio e Maurício Marengoni *

Resumo: Este capítulo descreve o algoritmo Mack para detecção de manchas no disco solar, o qual foi elaborado por meio de modificações no algoritmo de Curto. Além da apresentação da importância da detecção de manchas solares e das dificuldades relacionadas ao processo de detecção automática, são descritos os dois algoritmos, apontando as modificações realizadas. O algoritmo Mack foi aplicado às imagens do observatório SOHO e os resultados foram comparados com os dados publicados pelo NOAA *Institute* e pelo SIDC. A alta correlação entre os resultados obtidos pelos autores e os provenientes destas instituições apontou consistência dos resultados. As alterações realizadas no algoritmo original sugerem um aperfeiçoamento da robustez do método.

Palavras-chave: Processamento de imagens, Visão Computacional, Manchas Solares.

Abstract: *This chapter describes the Mack algorithm for sunspots detection on the solar disk. This algorithm was created based on the Curto's algorithm. The importance of sunspots detection and its related drawbacks are presented, followed by a description of the Curto's algorithm and the changes performed in it. The proposed algorithm was applied to SOHO images and results were compared with data published by the NOAA Institute and SIDC. This comparison showed the results consistency, having reached a high data correlation with the observations. Furthermore, the changes done to the original algorithm suggest an improvement in the robustness of the method.*

Keywords: *Image processing, Computer Vision, Sunspots.*

*Autor para contato: mmarengoni@mackenzie.br

1. Introdução

A fragilidade da vida humana é evidenciada principalmente em ambientes desfavoráveis e, neste sentido, a previsão do clima espacial busca elucidar diferentes processos físicos solares e suas consequências para a vida na Terra. Alterações no campo magnético global, 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 (Colak & Qahwaji, 2007), os quais podem colocar em risco a saúde e vidas humanas.

As manchas solares são áreas escuras com intenso campo magnético 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 (Silva, 2006) gerando explosões solares e ejeções de massa que são responsáveis pelos distúrbios causados na Terra. 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 e, neste cenário, apresenta-se a razão para a contagem total de manchas solares e a classificação destas em grupos (Curto et al., 2008).

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 (Zharkov et al., 2005). Logo, o aumento substancial de dados contendo informações referentes às imagens solares, à detecção automatizada e à verificação de diferentes características de interesse, dentre outras aplicações, permitem uma previsão confiável da atividade solar e, portanto, da previsão do clima espacial (Zharkov et al., 2004).

Em física solar 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 atividade solar, assim como nos estudos das relações solares terrestres. O número relativo de manchas solares foi definido por Rudolf Wolf (Wolf, 1856), 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 (Svalgaard, 2010).

Zharkov et al. (2005) apresentaram técnicas para reconhecimento automático de manchas presentes no disco em imagens solares. Esta técnica se resume nos seguintes passos:

1. Aplicar uma suavização gaussiana com uma vizinhança 5×5 seguida por um operador de Sobel para uma cópia Δ da imagem.
2. Utilizando um valor de limiar inicial, T_0, binarizar o mapa de bordas e aplicar um filtro de mediana 5×5 ao resultado. Em seguida, contar o número de componentes conexos, N_c, e a taxa do número de *pixels* de borda para o número total de *pixels* do disco solar, R. Se N_c for maior que 250 ou R maior que 0,7, incrementar T_0 e repetir o passo 2.
3. Binarizar iterativamente a imagem original para definir menos de 100 regiões escuras. Combinar as duas imagens binárias dentro de um mapa binário de caractecterísticas candidatas.
4. Remover a borda correspondente ao limbo do mapa de manchas candidatas e preencher as possíveis lacunas nas características de contorno utilizando os operadores morfológicos de fechamento e *watershed*;
5. Utilizar uma coloração por *blobs* para definir uma região de interesse, F_i, como um conjunto de *pixels* representando um componente conexo na imagem binária resultante, $\overline{B}_\Delta$.
6. Criar um mapa de manchas solares candidatas, B_Δ, uma máscara de byte a qual conterá os resultados da detecção com os *pixels* pertencentes à umbra marcados como 2, e à penumbra marcados como 1.
7. Para cada F_i uma imagem cortada contendo F_i definir T_s e T_u:

   a. se $|F_i| \leq 5$ *pixels* associar os limiares:
   para penumbra: $T_s = 0,91 I_{QSun}$;
   para umbra $T_u = 0,6 I_{QSun}$

   b. Se $|F_i| \geq 5$ *pixels* associar os limiares:
   para penumbra: $T_s = max\{0,93 I_{QSun}; (< F_i > -0,5\Delta F_I)\}$;
   para umbra: $T_u = max\{0,55 I_{QSun}; (< F_i > -\Delta F_i)\}$

   onde: $< F_i >$ é a intensidade média e ΔF_i é o desvio padrão para F_i;

8. Binarizar a imagem cortada neste valor para definir os *pixels* candidatos (umbral/penumbral) e inserir os resultados novamente em B_Δ. Em seguida, utilizar a coloração de *blob* para definir uma mancha solar candidata, S_i com os conjuntos de *pixels* representando um componente conexo em B_Δ.
9. Verificar os resultados de detecção, sobrepondo o B_Δ com o magnetograma sincronizado, M, como segue. Para cada mancha candidata S_i de B_Δ extrair:
   a. $B_{max}(S_i) = max(M(p)|p \in S_i)$
   b. $B_{min}(S_i) = min(M(p)|p \in S_i)$
   c. se $max(abs(B_{max}(S_i)), abs(B_{min}(S_i))) < 100$ então descarte S_i como ruído;
10. Para cada S_i extrair e armazenar os seguintes parâmetros: coordenadas do centro de gravidade, área, diâmetro, tamanho da umbra, número de umbras detectadas, intensidades fotométricas: máxima, mínima e média, fluxos magnéticos: máximo, mínimo e médio, fluxo magnético total e fluxo umbral total.

Colak & Qahwaji (2007) apresentaram uma classificação automática das manchas solares em tempo real para previsão de atividade solar. O método descrito por eles foi divido em três etapas:

1. Aplicação de um processo de filtragem e cálculo das coordenadas solares.
2. Detecção inicial de manchas solares a partir de imagens contínuas, e detecção de regiões ativas a partir do magnetograma.
3. Agrupamento de manchas solares utilizando aprendizado de máquina por meio de Redes Neurais Artificais.

Posteriormente, Curto et al. (2008) 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 seus resultados aos dados disponibilizados pelo SFC (*Solar Feature Catalog*), instituto europeu de controle de atividades solares.

O SOHO (*Solar & Heliospheric Observatory*), lançado em 2 de Dezembro de 1995, é um projeto de colaboração internacional entre a ESA e a NASA para estudo do Sol a partir do seu núcleo até sua corona mais externa, assim como o vento solar. O veículo solar SOHO foi construído na Europa por um conjunto de indústrias liderado por *Matra Marconi Space* (atualmente *EADS Astrium*) sob o gerenciamento global da ESA. Os doze instrumentos a bordo do SOHO foram fornecidos por cientistas da América

e Europa. A NASA foi a responsável pelo seu lançamento e, atualmente, pelas operações em missões. Grandes antenas de rádio espalhadas ao redor do mundo, as quais formam a Rede Espacial da NASA, são utilizadas para download de dados e comandos. O controle de missão é baseado no *Goddard Space Flight Center* em Maryland [1].

Diante deste contexto, este trabalho apresenta uma versão modificada do algoritmo de detecção de Curto et al. (2008) adaptada para as imagens solares na faixa visível do observatório SOHO (*Solar and Heliospheric Observatory*), denominada algoritmo Mack. As operações morfológicas relevantes para este artigo são brevemente discutidas na Seção 2. A técnica de Curto et al. (2008) para descrição e agrupamento e as modificações feitas em seu algoritmo são descritas nas Seções 3 e 4, respectivamente. A técnica de agrupamento utiliza o mesmo princípio de Curto et al. (2008): as manchas solares são agrupadas com base em suas distâncias heliográficas uma em relação à outra. A Seção 4 detalha o método empregado para converter as coordenadas cartesianas em coordenadas heliográficas. A Seção 5 apresenta os resultados da comparação dos dados com os apresentados por Curto et al. (2008) para detecção, assim como expõe os resultados de atividade solar com os dados obtidos ao longo dos anos de 2001 e 2009, 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 destas, por exemplo, a filtragem morfológica.

As operações morfológicas são baseadas na teoria dos conjuntos. Em imagens binárias, os conjuntos em questão são membros do espaço 2-D de números inteiros Z^2, em que cada elemento de um conjunto é um vetor bidimensional, cujas coordenadas são (x, y) de um *pixel* branco (ou preto, dependendo da convenção) de uma imagem. As imagens digitais em níveis de cinza podem ser representadas como conjuntos, cujos componentes estão em Z^3. Neste caso, dois componentes de cada elemento do conjunto referem-se às coordenadas de um pixel, e o terceiro corresponde ao seu valor discreto de intensidade. Um conceito fundamental da morfologia é o de **elemento estruturante** (ES): pequenos conjuntos ou subimagens utilizadas para examinar uma imagem, com o objetivo de identificar propriedades de interesse (Gonzalez & Woods, 2010). A Figura 1 exibe exemplos de elementos estruturantes.

[1] http://sohowww.nascom.nasa.gov/about/about.html

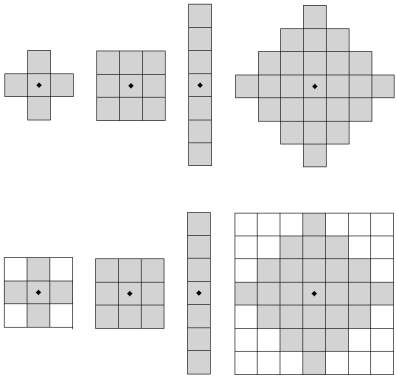


Figura 1. Primeira linha: exemplos de elementos estruturantes. Segunda linha: elementos estruturantes convertidos em arranjos retangulares (Gonzalez & Woods, 2010).

A seguir são apresentadas as operações morfológicas relevantes para este trabalho. Embora a morfologia tenha sido inicialmente aplicada em imagens binárias, as operações aqui descritas foram aplicadas em imagens em nível de cinza.

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:

$$f \ominus b(x, y) = \min\{f(x + s, y + t)\}_{(s,t)\in b_N} \tag{1}$$

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 (Gonzalez & Woods, 2010), tem o efeito de aumentar as partes mais claras da imagem. A dilatação é dada por:

$$f \oplus b(x, y) = \max\{f(x + s, y + t)\}_{(s,t)\in b_N} \tag{2}$$

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 o fechamento são dados, respectivamente, por:

$$f \circ b = (f \ominus b) \oplus b \tag{3}$$

$$f \bullet b = (f \oplus b) \ominus b \tag{4}$$

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.

A abertura geralmente suaviza o contorno de um objeto, rompe os istmos e elimina as saliências finas. O fechamento também tende a suavizar contornos, mas, ao contrário da abertura, geralmente funde as descontinuidades estreitas e alonga os golfos finos, elimina pequenos buracos e preenche as lacunas em um contorno (Gonzalez & Woods, 2010).

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. A seguir são apresentadas as descrições das transformadas *top-hat* e *bottom-hat*:

$$T(f) = f - (f \circ b) \tag{5}$$

$$T(f) = (f \bullet b) - f \tag{6}$$

Uma das principais aplicações destas transformadas é a remoção de uma imagem fazendo uso de um elemento estruturante na operação de abertura ou de fechamento que não se encaixa nos objetos a serem removidos. A operação de diferença resulta em uma imagem na qual apenas os componentes removidos permanecem (Gonzalez & Woods, 2010).

3. Detecção de Manchas Solares

A superfície do Sol apresenta uma distribuição de estruturas com diferentes níveis de intensidades e padrões regulares indefinidos, conforme se apresenta na Figura 2.

Geralmente cada estrutura escura é considerada como uma mancha solar e, sua identificação, envolve um processo de segmentação da imagem em escala de cinza do Sol, como nas imagens mostradas nas Figuras 2 e 3. Segundo Curto et al. (2008), 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

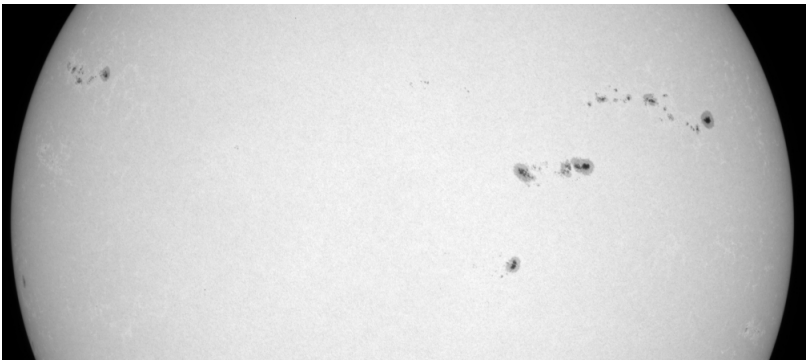


Figura 2. Imagem original com manchas, SOHO.

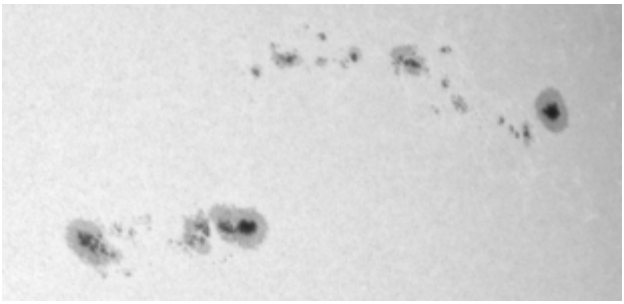


Figura 3. Manchas solares ampliadas.

binarização, este produz histogramas não particionáveis em decorrência da variação de intensidade da imagem do disco solar, ou seja, o escurecimento do limbo solar (Zharkov et al., 2005). Este fenômeno, o escurecimento do limbo solar, ocorre porque o brilho é máximo na faixa visível localizada no centro do disco e decai aproximadamente 20% nas bordas em função do gradiente positivo de temperatura da atmosfera solar. Logo, camadas mais próximas à superfície, que são mais frias, são menos brilhantes (Silva, 2006).

Estruturas escuras na imagem são consideradas manchas candidatas, uma vez que podem se tratar de ruídos e não de manchas solares. Para distinguir uma mancha solar de um ruído tem-se que considerar duas características: o tamanho da mancha candidata e o quão escura ela é. Espacialmente, quanto maior a mancha candidata, maior a probabilidade de ser uma mancha solar e, quanto menor, maior a probabilidade de ser

um ruído. No histograma de cor, quanto mais escura a mancha candidata, maior a probabilidade de ser uma mancha solar e, quanto mais clara, maior a probabilidade de ser um ruído. O algoritmo atua de forma iterativa para encontrar um tamanho de elemento estruturante que elimine espacialmente o maior número de ruídos e, ao mesmo tempo, minimize a eliminação de manchas verdadeiras. O algoritmo também atua para encontrar um ponto de corte ótimo no processo de binarização, visando sempre maximizar a eliminação de ruídos e minimizar a eliminação de manchas.

A técnica descrita por Curto et al. (2008) consiste na aplicação de um procedimento utilizando a binarização iterativa, assim como em Zharkov et al. (2005). Contudo, em Curto et al. (2008), 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. (2008) foi modificada para eliminar ruídos e produzir maior robustez contra o efeito de escurecimento do limbo. Assim, aplica-se a operação *bottom-hat* (Figura 4) em detrimento da operação *top-hat* utilizada por Curto et al. (2008). 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, o qual inicialmente detecta somente 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.

A presente proposta visa controlar os *pixels* detectados como pertencentes às manchas solares e interromper o processo quando o seu número estabilizar. O número de *pixels* com valor branco após a binarização na iteração n é P_n, assim, 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 disto, o processo finaliza quando a taxa de decaimento de *pixels* detectados tende a zero (Figura 5). 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

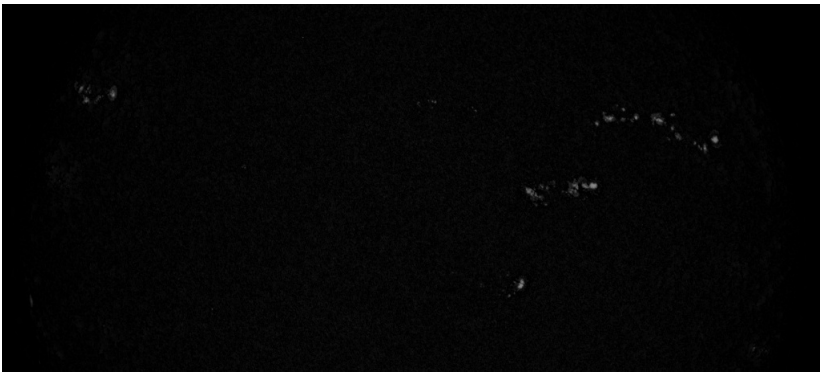


Figura 4. Imagem após operação *bottom-hat*.

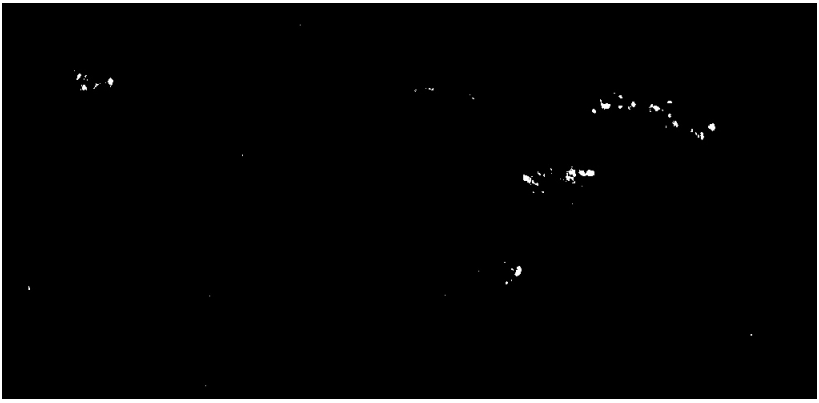


Figura 5. Imagem binarizada.

valor maior que zero. Isto 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 significativa. 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 (Curto et al., 2008) ou, para alguns casos observados, decresce. Curto et al. (2008)

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 foram feitos em passos de duas unidades. No final, conforme mencionado, aplicou-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* (Gonzalez & Woods, 2010), 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.

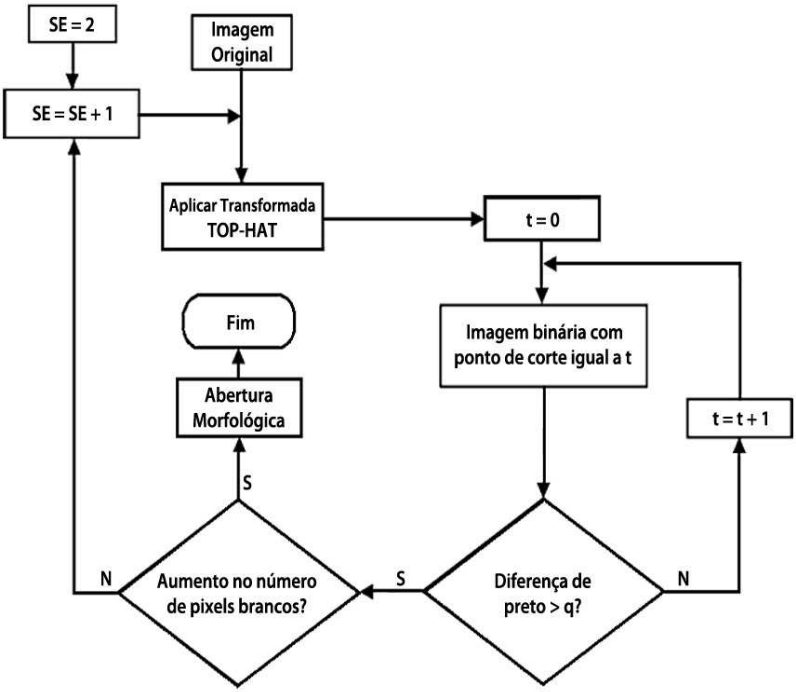


Figura 6. Algoritmo de detecção de manchas.

O valor 2 para a abertura morfológica ao final foi determinado experimentalmente, sendo observado que tamanhos estruturantes maiores tendiam a eliminar manchas de tamanhos reduzidos. Diante disto, não foi utilizado um tamanho ímpar para o elemento estruturante.

Para este trabalho, utilizou-se o valor 40 para o parâmetro q, determinado experimentalmente. A obtenção deste valor se deu observando a quantidade de falsos positivos e falsos negativos para um subconjunto das imagens de referência, em relação aos resultados de Curto et al. (2008). Para um conjunto de imagens proveniente de outra fonte, possivelmente este parâmetro tenha de ser ajustado. Uma vez detectadas as manchas, é necessário determinar sua localização na imagem. Para isto, 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, verificam-se os *pixels* localizados imediatamente à sua esquerda e na parte superior na estrutura da imagem, a fim de se determinar se este pixel pertence a um *blob* já existente. Em caso negativo, 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 (Silva, 2006).

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 (Wolf, 1856). Duas manchas solares fazem parte do mesmo grupo se estão localizadas próximas e compartilham do mesmo arco de fluxo magnético, para isto utiliza-se o critério de proximidade espacial (Curto et al., 2008). Neste contexto, estatisticamente define-se quais manchas solares fazem parte do mesmo grupo se estiverem até seis graus heliográficos equidistantes (Curto et al., 2008).

Para se determinar as distâncias entre as diferentes manchas solares, torna-se necessário converter a imagem de coordenadas cartesianas para coordenadas heliográficas (Green, 1985). 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 por efeitos de projeção como nas proximidades 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 que 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), transformam-se as coordenadas cartesianas (x, y) das manchas solares em coordenadas heliográficas. Aplicam-se as equações 7 e 8 (Green, 1985) para o centro de cada mancha, para encontrar sua localização (longitude e latitude) 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 (Green, 1985) nas manchas encontradas.

$$\text{Lat} = \arcsin\left(\frac{y}{r}\right) \tag{7}$$

$$\text{Lon} = \arcsin\left(\frac{x}{\sqrt{r^2 + y^2}}\right) \tag{8}$$

$$\text{Lamb} = \arccos(\sin(\text{lat}_1)\sin(\text{lat}_2) + \cos(\text{lat}_1)\cos(\text{lat}_2)\cos(\text{lon}_2 - \text{lon}_1)) \tag{9}$$

Na Figura 7 é apresentado o fluxograma do algoritmo de detecção de grupos de manchas.

5. Resultados

Este trabalho apresentou uma versão modificada do algoritmo de detecção de Curto et al. (2008), implementado em linguagem C++, utilizando a biblioteca de algoritmos de Visão Computacional OpenCV (Intel & Garage, 1999), adaptado para as imagens solares na faixa visível do observatório SOHO. A Figura 8 mostra resumidamente o passos no processo de detecção e agrupamento.

5.1 Detecção de manchas

Curto et al. (2008), para verificação dos resultados, aplicaram seu algoritmo às imagens do observatório Ebro e compararam 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. (2008).

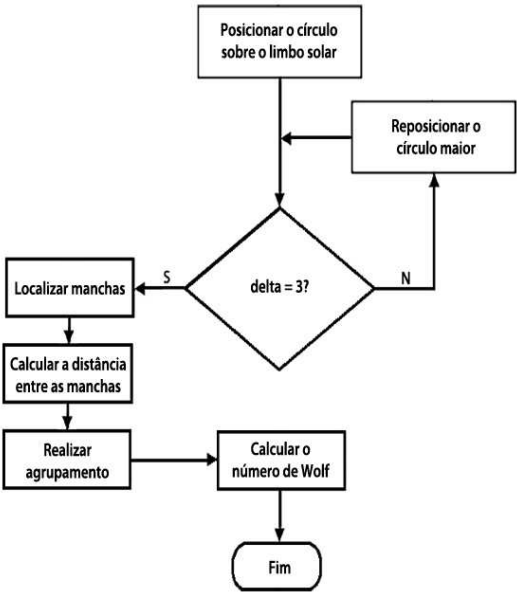


Figura 7. Algoritmo de detecção de grupos.

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. Ressalta-se que algumas imagens presentes na tabela original de Curto et al. (2008) em determinadas datas, não se encontravam disponíveis no site do laboratório SOHO.

Nota-se, pelos resultados obtidos que, de forma geral, os dados apresentados por Curto et al. (2008) e por este trabalho são consistentes. Em alguns pontos, por exemplo o dia 27/05/2004, Curto et al. (2008) obtiveram um resultado discrepante com o SFC, enquanto o presente trabalho obteve um resultado mais próximo ao do SFC. Em outros casos, Curto et al. (2008) 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. (2008). O coeficiente de correlação entre os dados do laboratório Ebro, utilizando o método original de Curto et al. (2008), e os resultados do SFC, foi de 46%. Utilizando as imagens do observatório SOHO com as alterações realizadas no algoritmo de Curto et al. (2008), foi obtida uma correlação de 59%. A correlação entre os resultados do algoritmo Mack e o de Curto et al. (2008) foi de 92%.

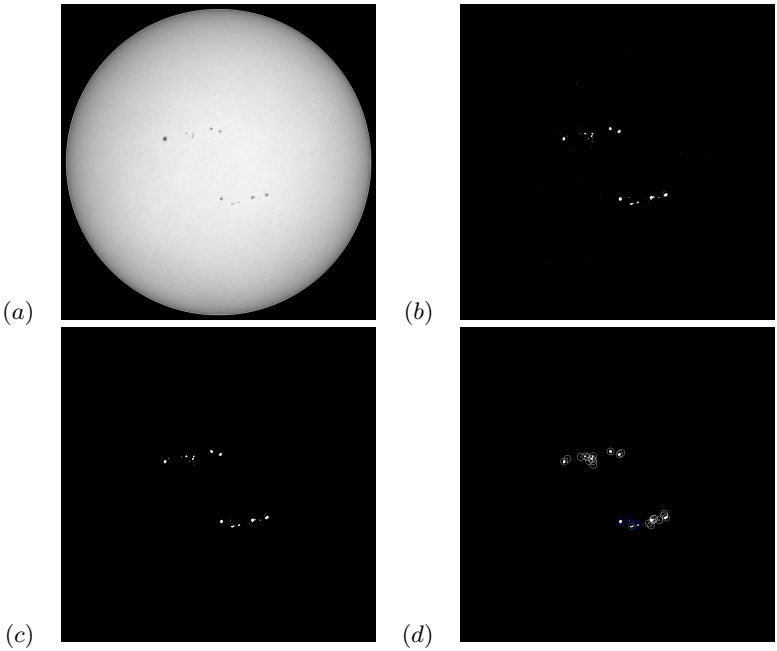


Figura 8. Processo de detecção e agrupamento das manchas solares. (a) imagem original com o círculo maior marcado (b) imagem binarizada pelo processo de detecção (c) imagem binária final (d) agrupamento.

5.2 Agrupamento de manchas

O agrupamento de manchas foi comparado com os resultados de agrupamento utilizando 350 imagens do ano de 2001 e 260 imagens do ano de 2009 do observatório SOHO com os dados divulgados pelos Institutos NOAA e o SIDC, em relação ao número de Wolf. O processamento automático de agrupamento resultou em uma correlação de aproximadamente 79% e 93%, respectivamente, para o ano de 2001 e 63% e 67%, respectivamente, para o ano de 2009. A Figura 9 apresenta o gráfico de comparação dos resultados deste trabalho com os resultados dos institutos NOAA e SIDC para o ano de 2001 e a Figura 10 mostra o mesmo gráfico para o ano de 2009.

Nota-se, pelo gráfico do ano de 2001, 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.

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	27	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*

Comparando as figuras 9 e 10, os resultados foram mais díspares para o ano de 2009 do que para o ano de 2001, uma vez que o ano de 2001 se caracterizou pela existência de poucas manchas em função do período do ciclo solar. Além disto, estas manchas foram espacialmente pequenas, o que ocasionou em um aumento da probabilidade de manchas se confundirem com ruídos. Na prática, o que houve foi que mais ruídos foram erroneamente detectados como manchas. A Figura 11 mostra uma imagem do ano de 2009 do observatório SOHO. É possível observar que as manchas solares são quase imperceptíveis. Isto se deve ao fato que o Sol estava em um período de mínima atividade enquanto em 2001, a atividade solar se encontrava próxima ao máximo do ciclo solar de 11 anos.

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 da previsão de clima espacial com a manipulação do numeroso acervo disponibilizado no Observatório SOHO.

Ferramentas morfológicas para a detecção de manchas solares foram utilizadas para analisar e extrair características geométricas das imagens, dentre elas, as operações morfológicas de pré e pós-processamento. Tais

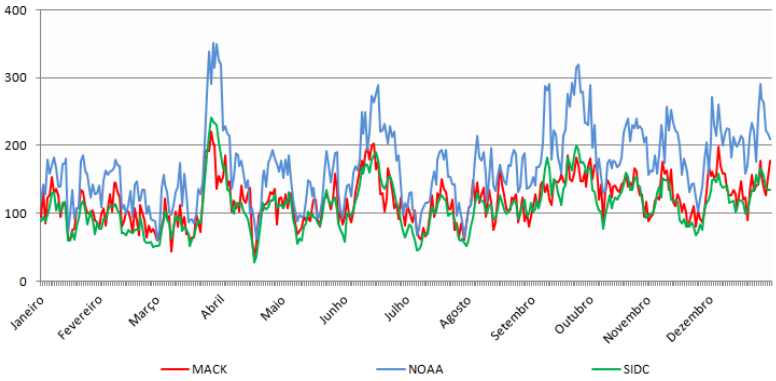


Figura 9. Comparação entre os algoritmos para o ano de 2001.

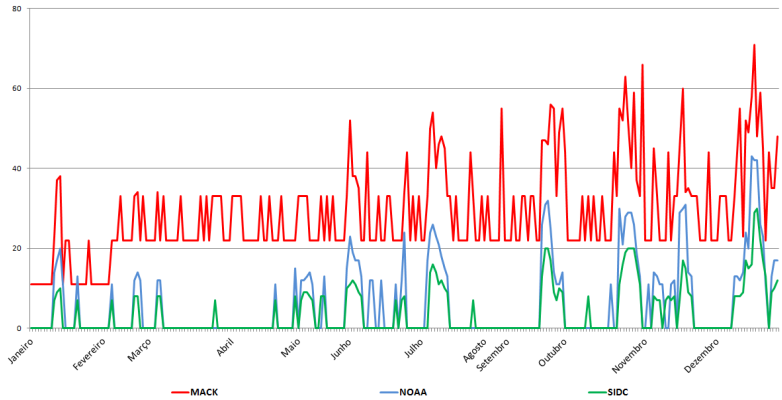


Figura 10. Comparação entre os algoritmos para o ano de 2009.

operações são baseadas na teoria dos conjuntos, sendo utilizadas neste trabalho 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. (2008). Além disto, a correlação com os dados divulgados pelos institutos NOAA e SIDC apon-

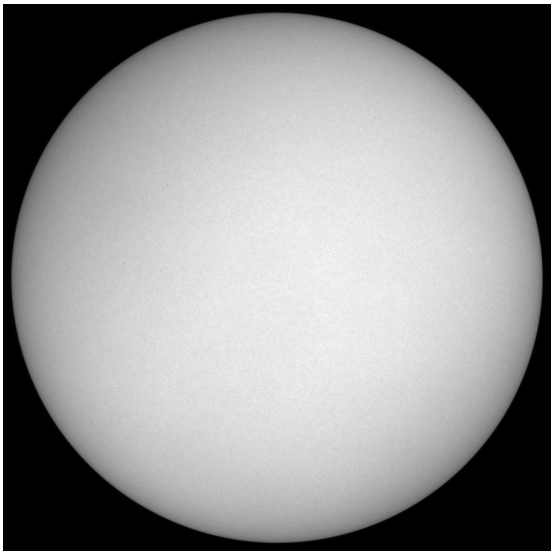


Figura 11. Imagem do observatório SOHO do ano de 2009.

taram que o algoritmo Mack pode contribuir para a detecção automática de manchas solares.

Uma dificuldade encontrada foi a determinação de falsos positivos ou negativos, tanto na detecção de manchas quanto no agrupamento. Tal fato indica a 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 Zharkov et al. (2004).

Agradecimentos

A. Válio gostaria de agradecer o suporte financeiro parcial da FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, processo no 2006/50654-3).

Referências

Colak, T. & Qahwaji, R., Automatic sunspot classification for real-time forecasting of solar activities. In: *Proceedings of International Con-*

ference on Recent Advances in Space Technologies. Piscataway, USA: IEEE Press, p. 733–738, 2007.

Curto, J.; Blanca, M. & Martínez, E., Automatic sunspots detection on full-disk solar images using mathematical morphology. *Solar Physics*, 250:411–429, 2008.

Gonzalez, R.C. & Woods, R.C., *Processamento Digital de Imagens*, São Paulo, SP: Pearson Prentice-Hall. 3a edição, p. 415–449.

Green, R.M., *Spherical Astronomy*. Cambridge, UK: Cambridge University Press, 1985.

Intel, & Garage, W., Open Source Computer Vision Library. http://opencv.willowgarage.com, 1999.

Silva, A.V.R., *Nossa estrela: o Sol*. São Paulo, SP: Editora Livraria da Física, 2006.

Svalgaard, L., Objective calibration of sunspot numbers. In: *Proceedings of American Geophysical Union Fall Meeting*. p. SH53B–03, 2010.

Wolf, R., Mittheilungen uber die Sonnenflecken. v. 1. Zurich, Switzerland: Eidegenossische Sternwarte, 1856.

Zharkov, S.; Zharkova, V.; Ipson, S. & Benkhalil, A., Automated recognition of sunspots on the SOHO/MDI white light solar images. In: Negoita, M.G.; Howlett, R.J. & Jain, L.C. (Eds.), *Knowledge-Based Intelligent Information and Engineering Systems*. Berlin, Germany: Springer-Verlag, v. 3215 de *Lecture Notes in Computer Science*, p. 446–452, 2004.

Zharkov, S.; Zharkova, V.; Ipson, S. & Benkhalil, A., Technique for automated recognition of sunspots on full-disk solar images. *EURASIP Journal on Advances in Signal Processing*, 2005(15):2573–2584, 2005.

Notas Biográficas

Adílson Eduardo Spagiari é graduado em Engenharia da Computação (Faculdades Associadas de São Paulo). Atualmente é mestrando no curso de pós-graduação em Engenharia Elétrica da Universidade Presbiteriana Mackenzie.

Israel Florentino dos Santos é graduado em Engenharia da Computação (Faculdades Associadas de São Paulo, 2006). Atualmente é mestrando no curso de pós-graduação em Engenharia Elétrica da Universidade Presbiteriana Mackenzie.

Wander Lairson Costa é graduado em Engenharia da Computação (Faculdades Associadas de São Paulo, 2006). Atualmente é mestrando no curso de pós-graduação em Engenharia Elétrica da Universidade Presbiteriana Mackenzie.

Adriana Válio é Bacharel em Física (UNICAMP, 1986), mestre em Astronomia (USP, 1989 e University of California at Berkeley, EUA, 1992), doutor em Astronomia (University of California at Berkeley, EUA, 1995) e Livre Docente (Universidade de São Paulo, 2008). Atualmente é professor adjunto da Universidade Presbiteriana Mackenzie e membro do corpo docente da pós-graduação em Engenharia da Universidade Presbiteriana Mackenzie e da pós-graduação em Astrofísica do INPE.

Maurício Marengoni é graduado em Engenharia Industrial Mecânica (Centro Universitário da FEI, 1983), mestre em Computação (Instituto de Pesquisas Espaciais – INPE, 1992 e University of Rochester, EUA, 1994) e doutor em Ciência da Computação (Universidade de Massachusetts Amherst, EUA, 2002). Desde 2004 é professor adjunto do Programa de Pós Graduação em Engenharia Elétrica da Universidade Presbiteriana Mackenzie.

Capítulo 19

Exploração de Espaços de Características para Imagens por meio de Projeções Multidimensionais

Bruno Brandoli Machado*, Danilo Medeiros Eler, Glenda Michele Botelho, Rosane Minghim e João do Espírito Santo Batista Neto

Resumo: Aplicações para imagens baseiam-se na premissa de que o conjunto de dados sob investigação é corretamente representado por características. Apesar dos avanços na área, a definição de um conjunto de características adequado para representar um conjunto de dados ainda é um desafio. A abordagem apresentada neste capítulo procura inspecionar esta etapa de definição de características por meio de representações visuais. Isto permite que usuários mudem interativamente os parâmetros dos algoritmos extratores que avaliam visualmente a representatividade da saída obtida por estes algoritmos, obtendo novas perspectivas sobre a representatividade de tais características. É mostrada a relação da melhoria da qualidade da representação com a melhoria dos resultados de classificadores. Também é apresentado o uso das mesmas abordagens visuais no apoio ao processo de seleção, mostrando que o processo apoia a seleção mantendo a capacidade de discriminação das características.

Palavras-chave: Análise visual do espaço de características, Visualização dos espaço de características, Seleção de características, Avaliação do espaço de características.

Abstract: *Imaging applications rely on suitable extraction of features. Regardless of the advances in this field, for many applications finding the suitable set of features to represent an image set still remains challenging. The approach proposed by this chapter inspects this stage of the image processing pipeline via visualization techniques. It allows users to interactively modify feature extraction parameters and visualize improvement in their decisions by checking on improvements of the visualization generated. We show the relationships between improvement of the visualizations and the performance of the classifiers based on the same features. We also extend the approach to support feature selection seeking to maintain discrimination properties of the set of features.*

Keywords: *Visual feature space analysis, Feature space visualization, Feature selection, Feature space evaluation.*

*Autor para contato: brandoli@icmc.usp.br

1. Introdução

Gerar vetores de características é uma tarefa crucial para muitas aplicações relacionadas a imagens, incluindo reconhecimento de padrões, recuperação de imagem por conteúdo ou mineração de imagens. Esta tarefa é capaz de determinar a precisão dos resultados alcançados por aplicações desta natureza (Theodoridis & Koutroumbas, 2006). Há muitos algoritmos para computar tais vetores, cada qual com seu próprio conjunto de parâmetros, refletindo em diferentes propriedades do conjunto de imagens sob investigação. Portanto, a escolha de um descritor que forneça as maiores taxas de classificação e a maior eficiência na recuperação é uma tarefa desafiadora, que requer, sobretudo, um conhecimento *a priori* do conjunto de imagens (Aggarwal, 2002). Adicionalmente, muitos experimentos de alto custo computacional podem ser necessários para criar um modelo coerente com a aplicação (Pampalk et al., 2003), demandando também tempo do usuário até convergir para uma parametrização satisfatória.

Tradicionalmente, usuários começam por selecionar um conjunto de dados rotulado e por definir os parâmetros para os algoritmos extratores de características. Em seguida, os vetores de características são computados e a classificação é realizada. Os resultados somente exibirão altas taxas de classificação correta se o conjunto de características se mostrar adequado para o conjunto de imagens em análise. Estas taxas podem ser melhoradas somente depois da execução da tarefa de classificação, isto é, os usuários não podem antecipadamente inferir quão efetivas serão as características computadas para a classificação. Nosso objetivo principal é prover uma metodologia efetiva para auxiliar usuários a encontrar um espaço de características coerente.

Neste trabalho, uma abordagem para análise visual de espaços de características por meio da aplicação de técnicas de visualização de informação é apresentada. Inicialmente, diferentes conjuntos de características são gerados por diferentes técnicas ou variações de parâmetros. Em seguida, estes conjuntos são visualmente organizados em uma representação 2D, ou uma projeção, que revela a similaridade entre as imagens do conjunto, isto é, são formados agrupamentos de imagens similares de acordo com o espaço de características. Para reduzir a subjetividade durante a análise visual, uma medida de qualidade, chamada "coeficiente de silhueta" (Tan et al., 2005), é calculada a partir da representação visual. Este coeficiente indica o grau de capacidade das características em promover o agrupamento de imagens similares. Os experimentos mostram que sempre que a qualidade da representação visual, refletida pela silhueta, melhora, a taxa de classificação aumenta.

Esta abordagem visual permite que o usuário explore espaços de características e escolha o melhor. Além disto, permite que o usuário localize itens de dados que dificultam a criação do modelo. Nem sempre é possí-

vel, entretanto, compreender quais características influenciam a formação de agrupamentos de imagens similares. Nestes casos, é necessário empregar um algoritmo seletor de características para melhorar a qualidade do espaço. Para tanto, também é apresenta uma abordagem visual para seleção de características, combinando técnicas tradicionais de seleção com o auxílio da mineração visual de dados (Wong, 1999). Nesta abordagem, ao invés de projetar n imagens como instância de dados, projetam-se m características. A projeção resultante consistirá, portanto, de m amostras, cada qual representando uma característica. A exemplo da projeção tradicional, agrupamentos de características poderão ser gerados. Se agrupamentos indicam instâncias com propriedades comuns, então podemos assumir que características pertencentes ao mesmo grupo possuem poder discriminatório semelhante. Ao selecionar apenas algumas amostras de cada agrupamento da projeção, teremos um subconjunto de características, configurando um processo de seleção. Por fim, o melhor espaço de características é aquele cuja projeção apresenta o melhor coeficiente de silhueta, conforme comentado na descrição da primeira abordagem.

As principais contribuições deste trabalho são: propiciar uma abordagem para análise visual de espaços de características com o objetivo de convergir em espaços de características representativos para o propósito de classificação de imagens; uma abordagem para análise visual do relacionamento entre características, com o objetivo apoiar o processo de seleção por eliminação de características altamente correlacionadas; um método para avaliar quantitativamente a qualidade das representações visuais assim apoiando as decisões do usuário. Adicionalmente, foi adaptada uma técnica para seleção de características baseada em projeções (Botelho & Batista Neto, 2011) para complementar a metodologia de utilização de visualização baseada em projeções multidimensionais na escolha de espaços de características adequados a conjuntos de imagens.

Este capítulo está organizado da seguinte maneira: a Seção 2 descreve os trabalhos relacionados a análise visual de espaços de características. A Seção 3 apresenta informações conceituais dos principais tópicos relacionados às abordagens apresentadas. Na Seção 4 são apresentadas as duas técnicas sugeridas para a análise visual do espaço de características. A Seção 5 apresenta os resultados dos experimentos executados. Finalmente, as conclusões e trabalhos futuros são apresentados na Seção 6.

2. Trabalhos Relacionados

Utilizar técnicas de visualização para explorar e extrair conhecimento de conjuntos de dados é uma maneira eficiente de combinar a inteligência humana com métodos computacionais (Aggarwal, 2002). Muitas técnicas e ferramentas de visualização foram desenvolvidas para possibilitar a interação com dados abstratos, algumas das quais especialmente desenvolvidas

para explorar espaços multidimensionais resultantes de processos extratores de características.

Uma destas abordagens foi proposta por Rodrigues Jr. et al. (2003), posteriormente estendida em Rodrigues Jr. et al. (2005). O objetivo desta abordagem é prover suporte para análise de características empregadas em buscas de similaridade para um sistema de recuperação de imagens por conteúdo. Uma vez que o espaço de característica é formado, uma representação visual é criada para mostrar que a melhor representação visual leva à melhor medida de precisão e revocação quando uma busca é executada sobre os dados. Entretanto, a representação visual não é usada para auxiliar o usuário a interativamente definir um melhor conjunto de características ou o melhor conjunto de parâmetros para o algoritmo extrator, mas somente para confirmar que as medidas de busca precisão e revocação correspondem à melhoria da qualidade da representação visual. Na abordagem apresentada aqui, a representação visual é como um guia interativo para explorar, definir e refinar o conjunto de características, fornecendo uma perspectiva de como os parâmetros do algoritmo extrator ou os pesos das características afetam o relacionamento de similaridade entre imagens.

PEx-Image (Eler et al., 2009) é uma ferramenta que emprega representações visuais baseadas em posicionamento de pontos no plano para auxiliar a exploração de conjuntos de imagens. Ela foi aplicada para auxiliar na escolha da melhor representação visual de espaços de características para um determinado conjunto de dados. Em contraste à abordagem apresentada neste capítulo, na *PEx-Image* a qualidade da representação visual é definida de acordo com a perspectiva do usuário, enquanto que aqui alia-se a isto a uma medida quantitativa bem conhecida, emprestada da comunidade de mineração de dados, para reduzir a subjetividade das conclusões puramente baseadas na análise visual.

3. Conceitos Básicos

3.1 Extração de características

Extração de características é o processo de capturar características quantitativas de uma imagem representando-as em um espaço vetorial. As características são calculadas por meio de métodos tradicionais de análise de textura como matrizes de co-ocorrência (Haralick et al., 1973) e filtros de Gabor (Bianconi & Fernández, 2007). Para o primeiro, são consideradas cinco medidas (energia, entropia, inércia, momento de diferença inversa e correlação), para cinco distâncias (pares de pixels com distâncias entre si variando entre 1 e 5 pixels) e quatro direções; totalizando 100 diferentes características. Para o segundo, são calculadas 16 características usando a energia sobre as respostas dos filtros de Gabor (quatro orientações e quatro escalas).

3.2 Técnicas de projeção multidimensionais

Técnicas de projeção multidimensional (Paulovich et al., 2008) são um tipo de representação visual de um conjuntos de dados multidimensional que mapeia os indivíduos num espaço de baixa dimensão (duas ou três). Cada instância de dado corresponde em tela a um elemento visual, que pode ser representado por um círculo, um ponto ou uma esfera. As posições geométricas destes elementos refletem algum tipo de relacionamento entre as instâncias de dado, tais como similaridade ou relacionamento de vizinhança (Paulovich et al., 2008). Deste modo, se os elementos são posicionados na mesma vizinhança no espaço visual, em princípio as instâncias de dado que eles representam são similares de acordo com uma certa distância.

Atualmente, há um grande número de diferentes técnicas que consideram diferentes aspectos da distribuição dos dados – ver Paulovich et al. (2008) para mais detalhes sobre tais técnicas. Neste trabalho o interesse é definir uma técnica que reflita da melhor maneira possível, no espaço visual, as relações de similaridade entre as instâncias de dados do espaço original. Uma técnica que recebe um grande interesse para esta tarefa é a *Classical Scaling* (Cox & Cox, 2000). Nesta técnica é definida uma matriz de distância duplamente centrada (*doubly-centered*) entre todos os pares de instâncias, seguida de uma decomposição espectral para recuperar as coordenadas cartesianas dos elementos no espaço visual. É possível provar que se a função de distância é Euclideana, o espaço projetado apresenta o menor desvio médio quadrático, calculado a partir do espaço original $\mathbf{X}$ e todos possíveis espaços reduzidos $\mathbf{A}$. Isto é, entre todas matrizes $\mathbf{A}_{n\times p}$, a projeção $Y_{n\times p}$ é uma que minimiza $\|\mathbf{X}-\mathbf{A}\|^2 = \sum_{i,j}(x_{i,j}-a_{i,j})^2$ (Mardia et al., 1979), onde $x_{i,j}$ é a distância entre as instâncias i e j no espaço original e $a_{i,j}$ é a distância entre elas no espaço de menor dimensão (espaço projetado).

3.3 Seleção de características

A seleção de características é um processo importante para minimizar os problemas gerados pela alta dimensionalidade e pela utilização de conjuntos de dados que contenham características correlacionadas ou irrelevantes. Na prática, a seleção de características consiste em selecionar um subconjunto de características mais relevantes, dado o conjunto original de tamanho m. Entretanto, encontrar um subconjunto ótimo só é possível por meio de uma busca exaustiva, o que é inviável computacionalmente na maioria dos casos. Diante disto, vários métodos utilizam heurísticas para realizar a seleção.

Tradicionalmente, os métodos de seleção de características seguem uma perspectiva de reconhecimento de padrões (Jain & Zongker, 1997; Kudo & Sklansky, 2000; Liu & Yu, 2005), onde alguma forma de análise estatística

dos dados é conduzida, ou adotam uma abordagem de redes neurais artificiais (Santos, 2007). Em ambos os casos, a qualidade dos subconjuntos selecionados é medida por meio de uma função critério.

4. Abordagens para Análise Visual do Espaço de Características

O uso de projeções na definição de espaços de características pode apoiar de várias formas o processo, admitindo rápida intervenção do usuário. Descrevemos a seguir o uso deste tipo de visualização para a comparação de espaços e para a seleção de características.

4.1 Apoio visual à definição de espaços de características

A primeira abordagem de análise visual apresentada neste trabalho tem o objetivo de verificar se imagens similares, de acordo com o ponto de vista do usuário, são também similares de acordo com o conjunto de características extraído. Embora usuários possam empregar habilidades visuais para determinar a qualidade de uma projeção, ela não é suficiente para identificar diferenças oriundas de pequenas modificações do conjunto de características. Para reduzir este problema de subjetividade, emprega-se uma medida chamada de "coeficiente de silhueta" (Tan et al., 2005), originalmente proposta para avaliar resultados de algoritmos de agrupamento.

O coeficiente de silhueta mede tanto a coesão quanto a separação entre as instâncias de agrupamentos (*clusters*). Dada uma instância d_i, sua coesão a_i é calculada como a média das distâncias entre d_i e todas as outras instâncias pertencentes ao mesmo grupo de d_i. A separação b_i é a distância mínima entre d_i e todas as outras instâncias pertencentes a outros agrupamentos. A silhueta de uma projeção é dada como a média da silhueta para todas instâncias, como apresentado na Equação (1).

$$S = \frac{1}{n} \sum_{i=1}^{n} \frac{(b_i - a_i)}{max(a_i, b_i)} \tag{1}$$

A silhueta pode variar entre $-1 \leq S \leq 1$. Valores maiores indicam melhor coesão e separação entre agrupamentos. Na presente abordagem, agrupamentos são compostos levando em consideração as etiquetas pré-determinadas da imagem, e a silhueta indica se imagens pertencentes a uma mesma classe são mais similares entre elas mesmas do que imagens pertencentes a outras classes. Deste modo, o melhor conjunto de características é aquele que resulta em projeções que geram os maiores coeficientes de silhueta.

O diagrama da Figura 1 ilustra esta abordagem. Primeiramente, um conjunto de características de imagem é extraído e então projetado no espaço 2D. Se o usuário julgar satisfatório o resultado projeção, o conjunto

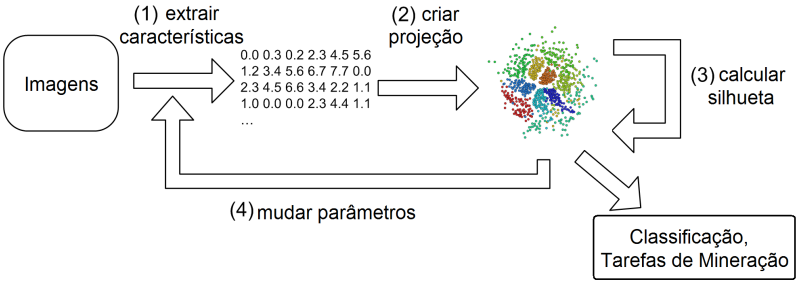


Figura 1. Abordagem para exploração visual do espaço de características.

de características pode ser empregado em outras tarefas como, por exemplo, classificação. No entanto, quando erros significativos de posicionamento são identificados, os parâmetros usados para calcular as características deverão ser modificados e o processo é refeito até se obter uma versão satisfatória do *layout* visual. No caso de pequenas diferenças de *layout*, o uso do coeficiente de silhueta pode ajudar no processo de decisão.

Esta abordagem auxilia na compreensão do espaço de características, permitindo que o usuário decida sobre a configuração dos parâmetros dos algoritmos extratores e até mesmo sobre quais características utilizar. No entanto, quando o espaço de características é de alta dimensão ou o entendimento das características é não trivial, é necessário executar uma etapa de seleção. Após a aplicação de um seletor, esta abordagem pode ser empregada para verificar se o espaço reduzido é melhor do que o original, ou até mesmo comparar diferentes espaços reduzidos (i.e., espaços resultantes de uma seleção de características).

4.2 Apoio das projeções multidimensionais ao processo de seleção de características

Esta abordagem combina projeções multidimensionais com elementos tradicionais de seleção de características. Além do usuário participar do processo de seleção através da representação visual, a abordagem também explora a seleção de características de forma automática, combinando projeções com algoritmos de agrupamento. A Figura 2 apresenta a metodologia usada nesta abordagem de seleção de características.

O conjunto original de dados é representado por uma matriz $M_{n\times m}$, composta por n instâncias (imagens) com m características cada. Esta matriz, que é o espaço de características, é projetada para uma futura comparação entre as projeções dos espaços reduzidos. Em seguida, o conjunto de dados original é transposto, obtendo-se agora uma matriz $M_{m\times n}$, ou seja, uma matriz com m instâncias (características) e n dimensões (imagens). A partir deste conjunto transposto, três processos de seleção são efetuados:

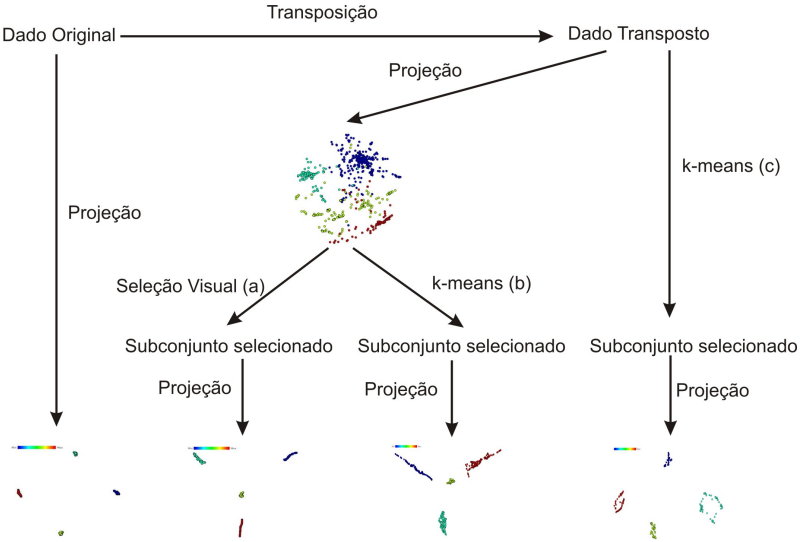


Figura 2. Metodologia. (a) Seleção Visual, com auxílio do usuário. (b) Seleção por meio da aplicação do algoritmo *k-means* sobre os dados projetados. (c) Seleção por meio da aplicação do algoritmo *k-means* diretamente no conjunto transposto.

a) seleção visual, com auxílio do usuário; b) seleção por meio da aplicação de *k-means* sobre os dados projetados; e c) seleção por meio da aplicação de *k-means* diretamente no conjunto transposto.

No primeiro processo (a), o conjunto transposto é projetado, podendo revelar agrupamentos. Cada um dos m pontos representa uma característica. A partir destes agrupamentos, um subconjunto de características (uma ou mais amostras de cada agrupamento) pode ser selecionado manualmente pelo usuário.

O segundo processo (b) consiste em substituir a seleção manual de características pela seleção automática, por meio da aplicação do algoritmo *k-means* sobre as características projetadas. Esta solução é normalmente empregada quando a projeção de características não revela agrupamentos bem definidos.

No terceiro processo de seleção (c), o algoritmo de agrupamento não supervisionado *k-means* é aplicado diretamente sobre o conjunto de dados transposto, gerando grupos de características. A partir destes, seleciona-se um subconjunto de características, geralmente aquelas mais próximas do centroide de cada agrupamento.

Os subconjuntos de características selecionados pelos três processos são, então, utilizados para gerar novas projeções do conjunto de imagens. Fi-

nalmente, as projeções obtidas das seleções e do espaço original podem ser comparadas tanto subjetiva, por meio da análise visual de um observador; quanto numericamente, por meio do coeficiente de silhueta, como no caso da comparação entre espaços de características descrito acima.

5. Resultados

Nesta seção são apresentados resultados da aplicação das abordagens descritas na Seção 4. A primeira abordagem é utilizada para auxiliar o usuário na configuração de um espaço de características e também para verificar qual espaço produz o melhor resultado visual. Esta abordagem é validada comparando-se o valor do coeficiente de silhueta com o valor da taxa de classificação de classificadores. A silhueta aqui é sempre calculada com base da projeção usando *Classical Scaling*. Uma vez que esta abordagem foi validada, ela é utilizada para verificar qual é o melhor espaço de características após a execução das etapas de seleção da segunda abordagem.

5.1 Análise visual do espaço de características

No experimento a seguir foi utilizado um conjunto de 70 imagens de texturas de Brodatz (Brodatz, 1966), que consiste em sete classes com dez imagens cada. São empregadas 16 características de Gabor, para quatro diferentes orientações (0°, 45°, 90° e 135°) e quatro escalas (numeradas de 1 a 4). A visualização do conjunto para este espaço de características é apresentada na Figura 3. As imagens são representadas como círculos e são coloridas de acordo com a classe à qual pertencem. É importante salientar que a informação de classe disponível para cada amostra do conjunto não é utilizada para gerar a projeção, mas somente para propósitos de exibição. Para um melhor entendimento, uma amostra de imagem de cada classe foi colocada perto de seu agrupamento correspondente. Nesta projeção, cinco classes estão bem separadas e duas misturadas (canto superior direito da figura), evidenciando uma provável limitação do poder discriminativo do conjunto de características escolhido. Uma observação mais aprofundada revela que em termos de padrão de textura, as duas classes de imagens são muito similares. Portanto, cabe ao usuário decidir se ambas devem ser consideradas como uma única classe ou qual deve ser a evolução do conjunto de características para realizar a separação.

Este experimento é estendido com a adição de três novas classes de textura (mármore e duas classes de arame). A projeção resultante é apresentada na Figura 4(a), destacando uma amostra (imagem) de cada nova classe adicionada. O mesmo conjunto de parâmetros foi utilizado neste experimento. Observe que a classe de imagens de mármore, cuja amostra é destacada em vermelho, aparece espalhada por toda a projeção. Um usuário sem muita experiência poderia facilmente concluir que o conjunto de características escolhido não é apropriado para representar a classe mármore,

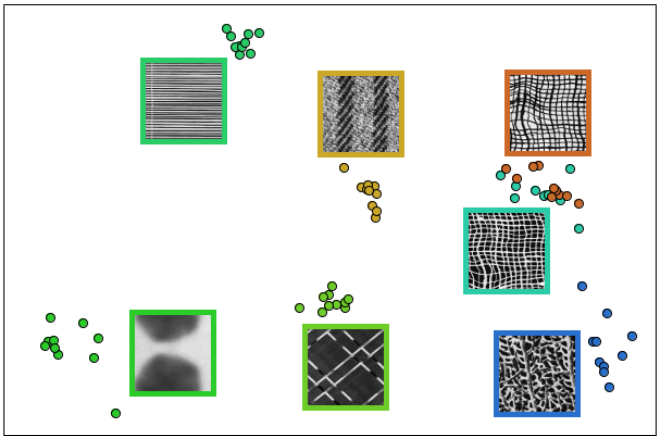
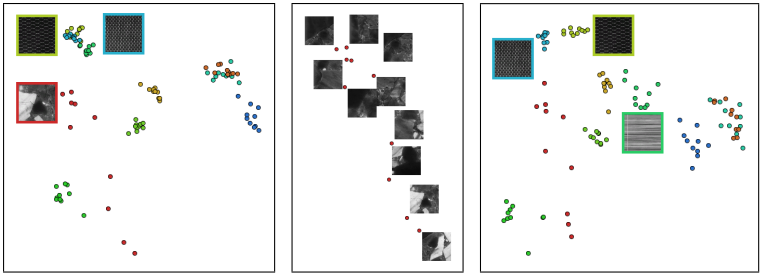


Figura 3. Análise Visual do uso de características de Gabor para o subconjunto de Brodatz.

o que provavelmente coincidiria com a opinião de um usuário especialista. As imagens desta classe apresentam elementos de textura não uniforme, os quais variam em tamanho e em orientação. Este aspecto é ilustrado na Figura 4(b) que mostra em detalhes (*zoom-in*) uma área da classe de imagens de mármore apresentada na Figura 4(a). Uma vez que um conjunto fixo de parâmetros (para orientação e escala) é usado, é improvável que todas as sutilezas das texturas (e.g., variação de tamanho e orientação) das imagens de mármore possam eventualmente ser capturadas.



(a) (b) (c)

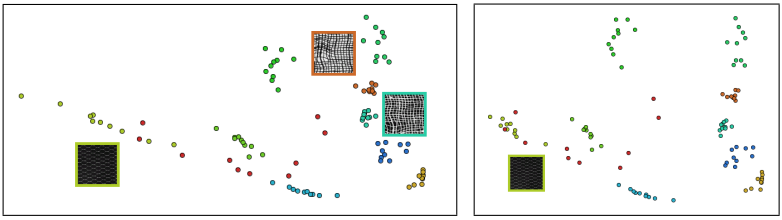
Figura 4. Projeções usando o Filtro de Gabor. (a) Todas as características. (b) Destaque em zoom da classe vermelha em (a). (s) Projeção usando somente as características de Gabor com orientação de 90°.

Para a classe de imagens de arame, as amostras projetadas estão misturadas, como pode ser observado no canto superior esquerdo da Figura 4(a). Para alcançar uma melhor separabilidade, somente característica com orientação 90° foram, portanto, selecionadas (o usuário ignorou as outras três orientações definindo seus pesos iguais a zero) e uma nova projeção é gerada. A visualização resultante é apresentada na Figura 4(c). Note que agora ambas as classes estão separadas, uma vez que o padrão de textura para a orientação 90° é diferente em cada classe de imagem. Entretanto, quando comparado com a projeção anterior, que emprega mais características, as demais classes estão mais espalhadas. Neste caso, o usuário pode concluir que a orientação 90° é eficaz para separar ambas as classes de arames, mas a coesão dos agrupamentos das demais classes é prejudicada. Este resultado não é facilmente percebido sem a utilização da visualização do espaço de características. Em termos do coeficiente de silhueta, as características empregadas para gerar a projeção apresentada na Figura 4(c) são melhores do que aquelas empregadas para gerar a projeção apresentada na Figura 4(a). Entretanto, a diferença é pequena, $0,429$ para a primeira e $0,474$ para a segunda, o que condiz com a inspeção visual.

Quaisquer que sejam as orientações escolhidas neste experimento é possível notar que as características dos filtros de Gabor falham em separar as classes de imagens de linho. As amostras destas classes estão coloridas em marrom e ciano (canto superior direito da Figura 3). Elas também estão presentes nas Figuras 4(a) e 4(c).

De fato, estas duas classes de imagens apresentam um padrão de textura muito similar e uma inspeção mais detalhada revela somente uma leve variação nas intensidades dos pixels. Procurando uma melhor separabilidade, adicionou-se as características de matrizes de co-ocorrência para quatro orientações (0°, 45°, 90° e 135°) e cinco distâncias (1 a 5 pixels). Os vetores de características são calculados sobre as matrizes utilizando cinco diferentes medidas estatísticas: energia, entropia, inércia, momento de diferença inversa e correlação, resultando em um espaço de características com 100 dimensões. A projeção resultante é apresentada na Figura 5(a), com amostras de cada classe de linho ilustradas à direita. Note que as amostras representadas pelas cores marrom e ciano estão agora bem separadas. A inspeção visual indica que a técnica de matrizes de co-ocorrência é uma melhor escolha para discriminar estas imagens, quando comparada com a técnica de filtros de Gabor. O coeficiente de silhueta reforça esta percepção, pois a projeção do espaço de características das matrizes de co-ocorrência resultou em um coeficiente com valor $0,583$, contra $0,474$ da projeção gerada a partir do espaço de características de filtros de Gabor (Figura 4(c)).

Embora o espaço de características das matrizes de co-ocorrência promova uma melhor separabilidade para as duas classes de imagem de linho, a classe de imagens de arame (lado esquerdo da projeção apresentada na



(a) (b)

Figura 5. (a) Projeção das texturas de Brodatz usando as características de co-ocorrência; (b) Projeção das texturas de Brodatz usando as características de co-ocorrência sem a medida de energia.

Figura 5(a)) revelou coesão inferior se comparada com aquela fornecida pelas características de filtros de Gabor. Com o intuito de identificar qual a medida estatística das matrizes de co-ocorrência com maior influência no espalhamento desta classe de imagem, foram geradas projeções considerando cada uma das medidas separadamente. O coeficiente de silhueta calculado destas projeções revela que a medida de energia é a que mais contribui para degradar o resultado. Consequentemente, uma nova projeção com todas as características, exceto aquelas calculadas com a energia, é criada. Esta projeção é apresentada na Figura 5(b). A classe de imagens de arame é mais bem agrupada, enquanto que a coesão das outras classes de imagem é preservada. Pode-se inferir que a medida de energia é um atributo que prejudica a representação deste conjunto de dados. O coeficiente de silhueta para esta projeção tem o valor $0,607$, isto é, um valor mais alto do que aqueles apresentados nas projeções anteriores. Isto indica que o espaço de características usado aqui é apenas quantitativamente o melhor; novamente, mas esta verificação condiz com a inspeção visual.

Os exemplos a seguir procuram apresentar como a abordagem proposta apoia tarefas de classificação baseadas em similaridade. Neste experimento, é utilizado um conjunto de dados de textura[1] com 280 imagens e sete diferentes classes, cada uma contendo 40 amostras distintas. A Tabela 1 apresenta os resultados de aplicação de um classificador K-NN com validação cruzada de dez dobras para as características de filtros de Gabor com diferentes orientações. A característica utilizada para formar o espaço é indicada pelo valor 1, que representa o peso dado à característica. $Test_1$, $Test_2$, $Test_3$ e $Test_4$ correspondem às orientações $0°$, $45°$, $90°$ e $135°$, respectivamente. $Test_5$ é a combinação das orientações $0°$ e $90°$, enquanto que $Test_{All}$ consiste em um espaço de características formado pela combinação de todas as orientações. O coeficiente de silhueta e a taxa de classificação

[1] *http://www-cvr.ai.uiuc.edu/ponce_grp/data/*

são condizentes, sustentando a ideia de que se um usuário utilizar o espaço de características que resultou na melhor projeção, também alcançará os melhores resultados em tarefas de classificação baseada em similaridade.

Tabela 1. Resultados de Silhueta com combinações das características de Gabor.

θ	**0°**	**45°**	**90°**	**135°**	**Silhueta**	**Classificação** (%)
$Test_1$	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

Em um segundo exemplo, extraiu-se características de matrizes de co-ocorrência a partir do mesmo conjunto de imagens e as combinou-se com as características de filtros de Gabor com orientações 0° e 90°. Para as matrizes de co-ocorrência foram escolhidas as medidas estatísticas de Entropia (H), Correlação (COR), Momento de Diferença Inversa (MDI) e Inércia (I). A Tabela 2 apresenta os resultados do coeficiente de silhueta e as taxas de classificação, bem como as características que foram selecionadas para compor o espaço. $Test_1$ combina as características COR e MDI, enquanto que $Test_{All}$ combina todas as características. Adicionalmente, $Test_{Best}$ considera as melhores características com base nos valores dos coeficientes de silhueta calculados a partir das projeções individuais. As características empregadas para o espaço $Test_{Best}$ são filtros de Gabor com orientação em 0° e 90°; e as medidas COR e I das matrizes de co-ocorrência. Novamente confirmou-se que a melhor projeção, isto é, aquela com maiores coeficientes de silhueta, resulta na melhor taxa de classificação.

Tabela 2. Resultados de Silhueta com diferentes combinações de características.

	0°	**90°**	**H**	**COR**	**MDI**	**I**	**Silhueta**	**Classificação** (%)
$Test_1$	–	–	–	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

Para melhorar a robustez desta abordagem em outros cenários, diferentes subconjuntos combinando as característica de filtros de Gabor e matrizes de co-ocorrência foram criados. Para cada espaço, uma projeção foi gerada e o coeficiente de silhueta computado. Além disto, diferentes classificadores baseados em distância, além do K-NN, foram aplicados sobre os sub-conjuntos. Os resultados deste experimento são apresentados na

Figura 6. O eixo horizontal indica os diferentes espaços de características e o eixo vertical mostra as medidas do coeficiente de silhueta e das taxas de classificação obtidas dos classificadores *K-means*, *Expectation Maximization* (EM) e *K-nearest neighbor* (K-NN). Para o classificador K-NN foi utilizado o valor de $K = 25$. Para os demais, o número de agrupamentos é definido como sete, que é a quantidade de classes do conjunto de dados. Embora a taxa de classificação observada entre os classificadores apresentados exiba uma certa flutuação, é interessante notar que, em geral, quanto maior o valor do coeficiente de silhueta, maior é a taxa de classificação.

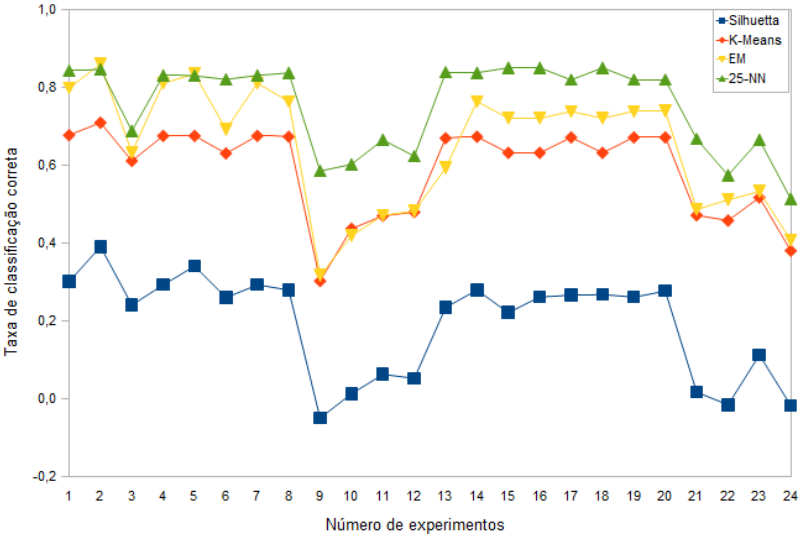


Figura 6. Relação entre silhueta e taxa de classificação para três diferentes classificadores.

Finalmente, o último experimento descreve como aplicar projeções para auxiliar a seleção de características a partir da inspeção de um conjunto reduzido de amostras (daqui em diante referido como conjunto de treino), validado com a tarefa de classificação executada sobre o conjunto de dados completo (conjunto de teste). Foi empregado um subconjunto do conjunto ImageCLEF 2006[2], composto de 3870 imagens e 35 classes (cada classe contendo entre 50 e 200 imagens). Para o conjunto de treino, 15 imagens de cada classe foram empregadas, resultando em um total de 525 imagens. A partir deste conjunto de dados, três diferentes conjuntos de características são formados: a) 16 características de filtros de Gabor; b) 100 características de matrizes de co-ocorrência; e c) 1024 características wavelet de Haar

[2] *http://ir.shef.ac.uk/imageclef/2006/*

com três níveis de decomposição. Projeções são criadas durante as etapas de treino e de teste para cada conjunto de características. A Tabela 3 apresenta os valores do coeficiente de silhueta e das taxas de classificação para o classificador K-NN. Quanto maior o valor do coeficiente de silhueta calculado para a projeção do conjunto de treino, melhor é a taxa de classificação do conjunto de teste. Esta correlação mostra como a qualidade alcançada pelas características selecionadas em um subconjunto é também preservada quando o conjunto de dados maior é empregado.

Tabela 3. Análise das silhuetas e taxas de classificação de três diferentes espaços de características.

Conjunto de Dados	Descritor	% (1-NN)	% (10-NN)	Silhueta
Treinamento	Gabor	18,86	16,95	-0,29
	Co-occorrência	38,67	37,91	-0,22
	Wavelet	77,14	71,62	-0,19
Teste	Gabor	14,11	15,46	-0,34
	Co-occorrência	35,16	36,08	-0,21
	Wavelet	77,10	75,67	-0,18

5.2 Análise visual para seleção de características

O objetivo dos seguintes experimentos é avaliar a abordagem de seleção de características apoiada por projeção, quer seja interativamente, quando o usuário analisa a projeção dos dados transpostos e seleciona manualmente um subconjunto de amostras; ou por meio de algum algoritmo de agrupamento, como por exemplo, o *k-means*. Os subconjuntos de características selecionados são usados para gerar novas projeções. Por fim, o espaço reduzido de características pode ser comparado ao espaço original verificando-se o valor do coeficiente de silhueta da projeção de cada espaço. No experimento, o conjunto transposto é projetado com a técnica *Least Square Projection* (LSP) (Paulovich et al., 2008), enquanto que os espaços de características originais e os construídos a partir da seleção são projetados com a técnica *Classical Scaling* (Cox & Cox, 2000).

5.2.1 Mosaico de texturas

Mais do que demonstrar a qualidade da conjunto selecionado de características, o primeiro experimento visa ilustrar as etapas envolvidas na abordagem de seleção por projeção de dados. Foram empregadas amostras (imagens) provenientes de um mosaico composto por padrões de textura em quatro ângulos de rotações diferentes (0°, 30°, 60° e 90°) do álbum *Rotate* de Brodatz (Brodatz, 1966). As características foram extraídas utilizando os Filtros de Gabor (Bianconi & Fernández, 2007), considerando quatro escalas, 20 orientações e três medidas (média, variância e energia), totalizando 240 características. O conjunto de dados é formado por 784

imagens (196 amostras por classe). Note que ao realizarmos a transposição, teremos 240 amostras e um total de 784 características.

A Figura 7(a) ilustra a projeção do conjunto original de dados com 784 imagens para todas as 240 características. A Figura 7(b) ilustra a projeção do conjunto transposto com a técnica de projeção LSP. Note que esta projeção não revelou agrupamentos perfeitamente delineados, o que impede a seleção interativa. Neste caso, aplicou-se o algoritmo *k-means* (para quatro classes) sobre os dados projetados para realizar o agrupamento das instâncias de dados. Na etapa de seleção automática, para cada um dos quatro agrupamentos, três amostras foram tomadas, totalizando 12 características. O processo aleatório pode ser facilmente substituído por um esquema automático em que as x amostras mais próximas ao centroide de cada grupo sejam tomadas.

Para fins de comparação, aplicou-se o *k-means* diretamente no conjunto transposto, obtendo-se um subconjunto com 12 características. A Figura 7(c) e (d) ilustram as projeções obtidas das 784 imagens para o método proposto e *k-means* diretamente sobre o conjunto transposto, respectivamente. As medidas de silhueta obtidas nas projeções são apresentadas na Tabela 4.

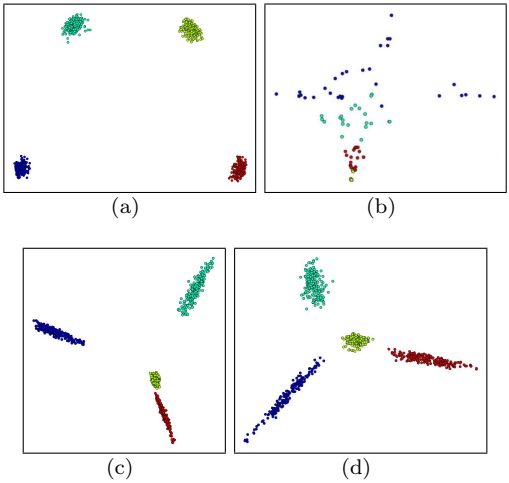


Figura 7. Projeção das 784 amostras. (a) Projeção do conjunto original de dados; (b) Projeção das 240 características, com os quatro agrupamentos computados por *k-means*; (c) Projeção usando o subconjunto de características selecionado pela aplicação do *k-means* a projeção de (b); (d) Projeção usando o subconjunto de características selecionado pela aplicação do *k-means* diretamente sobre o conjunto original.

Tabela 4. Medidas de silhueta obtidas das projeções para o conjunto de 784 imagens do mosaico de textura.

Método de Seleção	Silhueta	Figura 7
Sem seleção (240)	0,9398	(a)
Seleção baseada em projeção (12)	0,8029	(c)
Seleção por meio do *k-means* (12)	0,7541	(d)

Ao observar os resultados, percebe-se que as projeções das imagens apresentam as quatro classes distintas (0°, 30°, 60° e 90°) separadamente, ou seja, a qualidade visual das projeções se mantém após a seleção, com considerável redução do número de características (95% das características foram eliminadas). Também é confirmada a validade da abordagem pela redução pequena no valor da silhueta. É importante também salientar que a composição da amostra usada neste teste (para o dado transposto temos 240 amostras apenas, de dimensão 784) tipifica claramente uma situação suscetível aos efeitos da maldição de dimensionalidade, uma vez que o número de amostras chega a ser inclusive menor que o número de características. Apesar disto, o método consegue selecionar boas características, produzindo projeções com alta segregação entre classes. Isto também é notado pelos valores de silhueta altos, que corroboram a percepção visual.

5.2.2 Cenas naturais

Um segundo experimento foi realizado usando o banco de imagens de cenas naturais *Scene*, composto por 200 imagens, sendo 100 imagens de construções e 100 de costas oceânicas. As características das imagens foram extraídas por meio dos Filtros de Gabor, para quatro escalas, 12 orientações e apenas a função média, totalizando 48 características.

A Figura 8(a) ilustra a projeção do conjunto original de dados com 200 imagens para todas as características, onde pontos azuis representam imagens de construções e pontos vermelhos representam imagens de costas oceânicas. Em seguida, foi realizado o processo de seleção: o conjunto de dados foi transposto e projetado, revelando sete agrupamentos, conforme ilustra a Figura 8(b). O usuário, então, selecionou uma amostra de cada agrupamento, compondo um subconjunto com sete características. Para fins de comparação, aplicou-se o algoritmo *k-means* sobre o conjunto transposto não projetado, para sete agrupamentos, tomando uma única amostra de cada agrupamento (a mais próxima do centroide). As Figuras 8(c) e (d) ilustram, respectivamente, as projeções das 200 imagens para o subconjunto gerado pela seleção visual e para aquele obtido pelo método de seleção por *k-means*. As medidas de silhueta obtidas das projeções são apresentadas na Tabela 5.

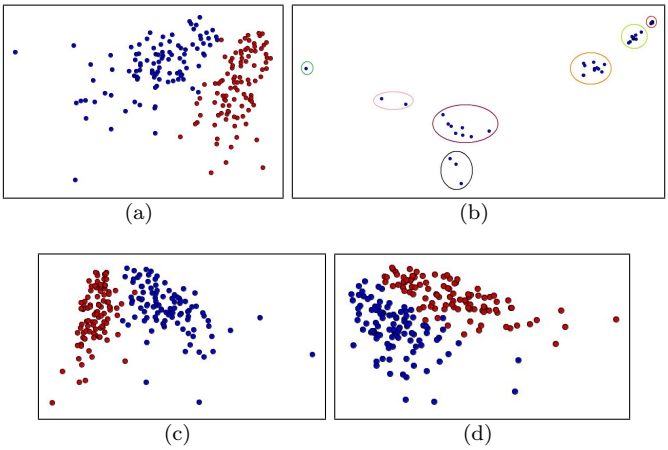


Figura 8. Projeção das 200 amostras. (a) Projeção do conjunto original de dados; (b) Projeção do conjunto transposto; (c) Subconjunto de características selecionado visualmente e (d) Subconjunto de características selecionado pela aplicação do *k-means* diretamente no conjunto transposto.

Tabela 5. Medidas de silhueta obtidas das projeções para o conjunto de 200 imagens de cenas naturais.

Método de Seleção	**Silhueta**	**Figura 8**
Sem seleção (48)	0,4599	(a)
Seleção baseada em projeção (7)	0,4946	(c)
Seleção por meio do *k-means* (7)	0,3651	(d)

Observando os valores de silhueta obtidos nas projeções percebe-se que ocorre uma pequena variação, ressaltando que houve um aumento no valor da silhueta obtido na projeção que utiliza apenas cerca de 15% das características selecionadas por meio visual. Além disto, a qualidade visual das projeções obtidas se mantém, pois as imagens de classes diferentes permanecem separadas.

6. Conclusão

Este trabalho apresentou duas abordagens que combinam exploração de espaços de características por meio de representações visuais com análises quantitativas por meio do cálculo do coeficiente de silhueta, com o intuito de melhorar a qualidade dos espaços de características que representam conjuntos de imagens. Muitos experimentos foram realizados, mostrando

que conforme o valor do coeficiente de silhueta aumenta, por meio de mudanças realizadas nos algoritmos extratores ou em seus parâmetros, a qualidade do espaço de características, em termos de separabilidade e coesão das classes de imagens, também aumenta. Além disto, classificadores baseados em distância apresentaram melhores taxas de classificação para espaços de característica que produzem os maiores valores de coeficiente de silhueta e quem apresentam visualizações de melhor qualidade.

Também comparou-se a utilização de uma abordagem baseada em mapeamentos visuais do conjunto de características com uma abordagem de seleção de características tradicional, que usa o algoritmo de agrupamento *k-means* para a obtenção de um subconjunto de características. Os experimentos realizados mostraram bons resultados, visto que as projeções obtidas, após o processo de seleção de características, mantiveram características de separabilidade semelhante às das projeções que usam o conjunto total de características. Isto é importante pois a redução do número de características é sempre muito significativa.

Ressalta-se que nem sempre é possível identificar grupos bem definidos na projeção de características. Neste casos, a abordagem consiste em aplicar um algoritmo de agrupamento, como o *k-means*, sobre as características projetadas ou diretamente no conjunto de dados transposto. Ao aplicar o algoritmo de agrupamento sobre o conjunto de características, pode-se considerar diferentes valores de k, o que permite selecionar diferentes subconjuntos de características.

Embora existam métodos automáticos para seleção de características em espaços de alta dimensão, a interação do usuário e sua percepção exercem um papel importante neste processo. Muitas vezes, diferentes classes de imagens têm, de fato, amostras muito similares. Em outros casos, amostras de uma dada classe podem ter um baixo grau de uniformidade. Consequentemente, o impacto que tais amostras produzem na classificação ou em algoritmos de seleção de características depende da natureza dos descritores. A capacidade de analisar amostras visualmente e a coesão de classes de imagens é uma abordagem poderosa que supera os limites impostos em esquemas tradicionais de seleção de características e, analogamente,de representar diferentes tipos de imagens por meio de características.

Como trabalhos futuros, é necessário buscar bases teóricas para adaptar a representação visual para que o usuário possa utilizar outros classificadores que não sejam puramente baseados em distância, tal como um classificador *Support Vector Machines* (SVM). Também pretendemos avaliar o método de seleção apoiado por projeção frente à um método de seleção quantitativo baseado em Redes Neurais Artificiais com Saliência (Santos, 2007). Além disto, os grupos de características obtidos nas projeções podem ser usados usados para determinar a quantidade de neurônios da camada oculta da rede neural, com a quantidade de agrupamentos sendo igual a quantidade de neurônios. Também é possível utilizar a discriminação

das projeções e para influenciar no cálculo da saliência de cada característica, o qual vai depender apenas das características pertencentes ao mesmo grupo e não mais do conjunto completo de características. Outro tema importante a ser tratado é a utilização de um método de agrupamento sem número de classes pré-definido, que possa sugerir o melhor subconjunto de características.

Agradecimentos

Os autores agradecem ao CNPq e à FAPESP pelo apoio financeiro.

Referências

Aggarwal, C.C., Towards effective and interpretable data mining by visual interaction. *ACM SIGKDD Explorations Newsletter*, 3(2):11–22, 2002.

Bianconi, F. & Fernández, A., Evaluation of the effects of Gabor filter parameters on texture classification. *Pattern Recognition*, 40(12):3325–3335, 2007.

Botelho, G.M. & Batista Neto, J.E.S., Seleção de características apoiada por mineração visual de dados. *Learning and Nonlinear Models*, 9(1):66–75, 2011.

Brodatz, P., *Textures: A Photographic Album for Artists and Designers*. Mineola, USA: Dover Publications, 1966.

Cox, T.F. & Cox, M.A.A., *Multidimensional Scaling*. 2a edição. Boca Raton, USA: Chapman & Hall / CRC, 2000.

Eler, D.; Nakazaki, M.; Paulovich, F.; Santos, D.; Andery, G.; Oliveira, M.; Batista, J. & Minghim, R., Visual analysis of image collections. *The Visual Computer*, 25(10):923–937, 2009.

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

Jain, A. & Zongker, D., Feature selection: Evaluation, application and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997.

Kudo, M. & Sklansky, J., Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1):25–41, 2000.

Liu, H. & Yu, L., Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):491–502, 2005.

Mardia, K.V.; Kent, J.T. & Bibby, J.M., *Multivariate Analysis (Probability and Mathematical Statistics)*. London, UK: Academic Press, 1979.

Pampalk, E.; Goebl, W. & Widmer, G., Visualizing changes in the structure of data for exploratory feature selection. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, p. 157–166, 2003.

Paulovich, F.V.; Nonato, L.G.; Minghim, R. & Levkowitz, H., 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, 2008.

Rodrigues Jr., J.F.; Castañón, C.A.B.; Traina, A.J.M. & Traina Jr., C., Using efficient visual exploration techniques to evaluate features for content-based image retrieval. In: *Proceedings of the Sixteenth Brazilian Symposium on Computer Graphics and Image Processing*. Los Alamitos, USA: IEEE Computer Society, p. 183–190, 2003.

Rodrigues Jr., J.F.; Traina, A.J.M. & Traina Jr., C., Enhanced visual evaluation of feature extractors for image mining. In: *Proceedings of the ACS/IEEE 2005 International Conference on Computer Systems and Applications*. Los Alamitos, USA: IEEE Computer Society, p. 45–52, 2005.

Santos D. P.; Batista Neto, J.E.S., Feature selection with equalized salience measures and its application to segmentation. In: Proceedings of the Twentieth Brazilian Symposium on Computer Graphics and Image Processing. p. 253–262, 2007.

Tan, P.N.; Steinbach, M. & Kumar, V., *Introduction to Data Mining*. 1a edição. Boston, MA: Addison-Wesley Longman, 2005.

Theodoridis, S. & Koutroumbas, K., *Pattern Recognition*. San Francisco, USA: Academic Press, 2006.

Wong, P.C., Guest editor's introduction: Visual data mining. *IEEE Computer Graphics and Applications*, 19(5):20–21, 1999.

Notas Biográficas

Bruno Brandoli Machado é graduado e mestre em Ciência da Computação (Universidade Católica Dom Bosco, 2007 e Universidade de São Paulo, 2010). Atualmente é doutorando no Instituto de Ciências Matemáticas e de Computação (ICMC-USP) e participa do GCI – Grupo de Computação Interdisciplinar do Instituto de Física de São Carlos. Seus interesses atuais em pesquisa incluem projeção de dados e redução de dimensionalidade, descrição por características locais, sistemas multiagentes, redes complexas e equações diferenciais parciais não lineares.

Danilo Medeiros Eler é bacharel em Ciência da Computação (Universidade do Oeste Paulista, 2001), mestre e doutor em Ciência da Computação e Matemática Computacional (ICMC-USP, 2006 e 2011). Atua como Professor Assistente do Departamento de Matemática Estatística e Computação da Universidade Estadual Paulista (UNESP), câmpus de Presidente Prudente. Seus interesses de pesquisa são em visualização, mineração visual de dados e aplicações de análise visual de dados.

Glenda Michele Botelho é graduada em Ciência da Computação (Universidade Federal de Goiás, 2007) e mestre em Ciência da Computação e Matemática Computacional (ICMC-USP, 2011). Atualmente é doutoranda no mesmo programa de pós-graduação. Possui interesse nas áreas de processamento de imagens, redes complexas e visualização de dados.

Rosane Minghim é graduada em Ciência da Computação (Universidade de São Paulo, 1985), mestre em Engenharia Biomédica (Universidade Estadual de Campinas, 1990) e doutor em Computer Studies (University of East Anglia, 1995). Realizou pós-doutorado na University of Massachusetts Lowell (2000-2001). Atua como professor associado no ICMC-USP. Seus interesses em pesquisa são em visualização, analítica visual, mineração visual de dados e aplicações de análise visual de dados.

João do Espírito Santo Batista Neto é graduado em Ciência da Computação (Universidade Federal de São Carlos, 1988), mestre em Ciência da Computação e Matemática Computacional (ICMC-USP, 1991) e doutor em em Engenharia Biomédica (University of London, 1996). Atua como professor no ICMC-USP e seus interesses de pesquisa são processamento de imagens e reconhecimento de padrões.

Capítulo 20

Mosaicos de Imagens Aéreas Sequenciais Construídos Automaticamente

André de Souza Tarallo*, Francisco Assis da Silva, Alan Kazuo Hiraga, Maria Stela Veludo de Paiva, Lúcio André de Castro Jorge

Resumo: A geração automática de mosaicos de imagens aéreas agrícolas aumenta a eficiência na análise das áreas agrícolas e nas tomadas de decisão relacionadas a controle de pragas, doenças e desmatamento. Neste capítulo é apresentada uma ferramenta para construção automática de mosaicos de imagens sequenciais. As principais características que interferem no desempenho da ferramenta são: construir mosaicos sem distorção e o custo computacional. Dez mosaicos foram obtidos a partir de 200 imagens agrícolas e comparados com aqueles obtidos com os softwares comerciais, mostrando melhor qualidade e menor distorção. Posteriormente, eles foram visualmente inspecionados por um profissional, que confirmou a qualidade da ferramenta desenvolvida.

Palavras-chave: Custo computacional, Imagens aéreas agrícolas, imagens de alta resolução, Mosaicos automáticos.

Abstract: *The automatic generation of mosaics of aerial agricultural images increases efficiency in the analysis of agricultural areas and in decisions-making related to pest control, diseases and deforestation. This chapter presents a tool for automatic construction of mosaics from sequential images. The main features that interfere with the performance of the tool are: building mosaics without distortion and the computational cost. Ten mosaics were obtained from 200 agricultural images and compared with those obtained with commercials software, showing better quality and less distortion. After, they were visually inspected by a professional, which confirmed the quality of the tool developed.*

Keywords: *Computational cost, Aerial agricultural images, High-resolution images, Automatic mosaics.*

*Autor para contato: andre.tarallo@gmail.com

1. Introdução

Até o presente momento, a construção de mosaicos de imagens na agricultura vem sendo feita de maneira semiautomática, necessitando obter o modelo digital do terreno, fazer a ortorretificação das imagens e colocação manual de bandeirinhas (marcadores), para que um software possa reconstruir esta área e gerar um mosaico deste terreno. Deste modo, a construção de mosaicos é demorada e trabalhosa, podendo demorar um dia todo ou mais dias. Isto também envolve uma grande demanda de pessoal para fazer as marcações na área em questão, além de poder gerar um mosaico com pouca precisão.

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 ao controle de pragas, doenças ou queimadas.

Estes fatos levaram à construção de uma metodologia apresentada neste capítulo, para a construção automática de mosaicos de imagens digitais na agricultura. As imagens são fotos 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.

Para a implementação desta metodologia, foi utilizada a transformada SIFT *(Scale Invariant Feature Transform)* para a extração de características das imagens, o algoritmo BBF *(Best-Bin-First)* para determinar pontos correspondentes entre pares de imagens e o algoritmo RANSAC *(Random Sample Consensus)* para filtrar os falsos pontos correspondentes entre os pares de imagens. Por fim, após a aplicação destas técnicas, os pares de imagens foram unidos.

O uso da SIFT para a implementação de mosaicos pode ser encontrado nos trabalhos de Bei & Haizhen (2009) e Li & Geng (2010). Outra técnica usada para este fim é a PCA-SIFT, encontrada no trabalho de Ke & Sukthankar (2004). SURF é uma técnica mais recente que a SIFT e tem se destacado por ter menor custo computacional (Hong et al., 2009). No entanto, de acordo com o trabalho de Juan & Gwun (2009), que faz uma comparação destas técnicas (SIFT, PCA-SIFT e SURF), a transformada SIFT apresenta melhor estabilidade quanto à invariância à escala e rotação, apesar desta transformada ter um custo computacional mais elevado do que as outras técnicas. De acordo com Lowe (2004), as técnicas BBF e

RANSAC, após aplicação da SIFT, são as mais indicadas para encontrarem os pontos correspondentes entre pares de imagens.

O restante deste capítulo está estruturado da seguinte maneira: na Seção 2 é apresentada a fundamentação teórica, na Seção 3 é detalhada a metodologia utilizada, na Seção 4 são apresentados os resultados, na Seção 5 são apresentadas as discussões e conclusões.

2. Fundamentação Teórica

Nesta seção será apresentado como foi construída a base de dados de imagens aéreas bem como a descrição das principais técnicas utilizadas neste artigo.

2.1 A base de imagens

As imagens aéreas utilizadas neste projeto de doutorado foram fornecidas pela EMBRAPA (Empresa Brasileira de Pesquisas Agropecuárias) Instrumentação, situada em São Carlos - SP. As imagens contêm áreas de pastagens, lavouras e áreas urbanas, obtidas na região de Santa Rita do Sapucaí - MG em Setembro de 2007, com o auxílio de um avião de pequeno porte, contendo uma câmera acoplada a um suporte. A base de imagens é composta por 200 imagens sequenciais. As imagens foram adquiridas com dimensão de 3504 x 2336 *pixels*, com 24 bits por *pixel* com 72 dpi (pontos por polegada) no formato JPG, utilizando 8 *Megapixels* de resolução. A Figura 1 ilustra um exemplo de imagem usada no projeto.

Figura 1. Exemplo de imagem usada no projeto.

As imagens foram obtidas seguindo um padrão de aquisição (para evitar distorções), que inclui os seguintes itens:

- A câmera foi posicionada horizontalmente em relação ao solo;
- O avião percorreu sempre a mesma distância, fazendo movimentos horizontais na área demarcada, para obtenção das imagens (Figura 2);
- A obtenção da sequência de imagens foi feita com um tempo determinado e síncrono, entre a imagem anterior e a posterior, para gerar regiões de sobreposição.

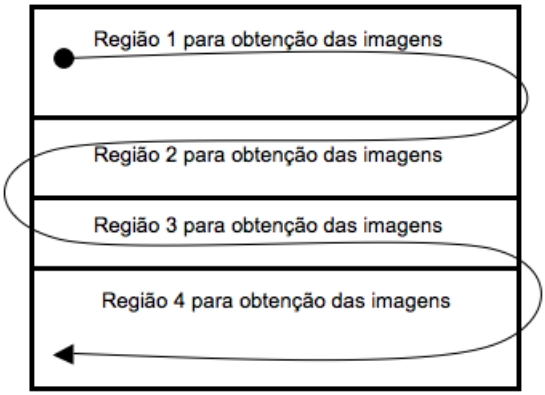


Figura 2. Padrão de rota do avião para obtenção das imagens.

Para uma construção sem muita distorção de um mosaico, é de extrema importância que seja seguido o padrão apresentado para a obtenção das imagens.

2.2 A transformada SIFT

A transformada SIFT (Lowe, 2004) é capaz de transformar uma imagem, em uma coleção de vetores de características locais (descritores de características), e cada um destes vetores são invariantes à escala, rotação e parcialmente invariante à mudanças de iluminação e ponto de vista.

As características fornecidas pela transformada SIFT são bem localizadas em ambos os domínios, o da frequência 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 transformada 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 atribuídos para cada ponto chave localizado, baseado 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 transformada SIFT.

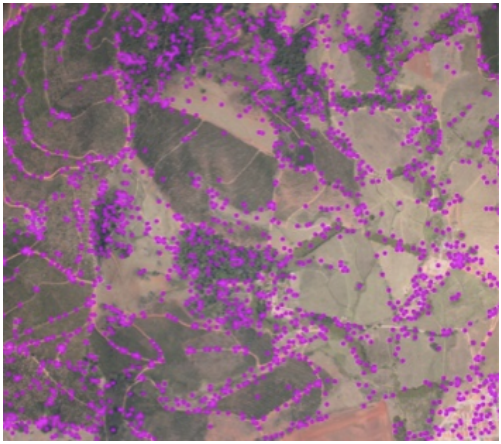


Figura 3. Localização de características com a transformada SIFT.

A transformada SIFT converte 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 (Lowe, 2004).

2.3 BBF

Uma vez aplicada a transformada 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 (Beis & Lowe, 1997) 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 (Beis & Lowe, 1997) 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 têm muitas distâncias similares.

2.4 RANSAC

Após a correspondência dos pontos chave, os mesmos são usados para calcular uma transformada que mapeia as posições dos pontos de uma imagem para as posições dos pontos correspondentes, na outra imagem, de um par de imagens.

Às vezes acontece de pares encontrados corresponderem a falsas correspondências, sendo necessário identificar estas falsas correspondências e de removê-las. A solução para este problema envolve o conceito da geometria epipolar (Oram, 2001) e homografia (Hartley & Zisserman, 2004). Com isso será reduzido o número de falsas correspondências e calculado uma transformação para juntar duas imagens sequenciais.

A correspondência de imagens fornece um conjunto de vetores de deslocamento relativo às características de um par de imagens obtidas, ou seja, cada vetor representa as coordenadas da mesma característica em ambas imagens. Com isto, é possível determinar o movimento entre tais imagens através da homografia.

Como a etapa de correspondência fornece um conjunto de n pontos correlacionados, estes pontos podem ser usados para se achar a matriz H. A matriz homográfica H é determinada, permitindo estimar o movimento entre as imagens.

2.5 Geometria epipolar

Os seres humanos têm a capacidade de distinguir quais objetos estão mais um próximos dos outros quando olham para eles, por possuirmos visão estéreo. Ou seja, cada um dos nossos olhos observa o mundo de pontos de vista diferentes e, a partir disto, o nosso cérebro consegue extrair várias relações geométricas entre as imagens formadas em cada retina. Assim, é capaz de reconstruir o ambiente 3D de forma que possamos perceber as diferenças de profundidade dos objetos que compõem a cena observada. A simulação computacional deste processo de visão que recria o ambiente 3D a partir de duas imagens é baseada na área da geometria, denominada geometria epipolar. Ela depende apenas dos parâmetros da câmera, independente da estrutura da cena (Roberto et al., 2009).

A modelagem da visão estéreo pode ser realizada usando duas câmeras, como mostrado na Figura 4. Pode-se observar que cada câmera possui o seu próprio centro e orientação. Deste modo, cada uma possui também o seu próprio sistema de coordenadas de câmera (Pollefeys, 1999).

Dentre as várias relações possíveis entre pares de imagens, algumas são bastante importantes, pois ocorrem em todos os casos de visão estéreo. A primeira delas é a reta que liga o centro C_1 da primeira câmera com o centro C_2 da segunda, chamada de *baseline*.

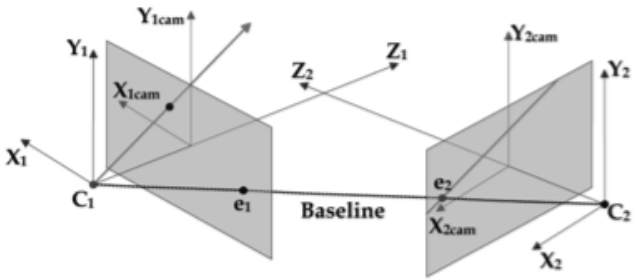


Figura 4. Esquema computacional da visão estéreo (Roberto et al., 2009).

O ponto de intersecção desta reta com o plano de imagem é chamado de epipolo. Para a primeira câmera têm-se o epipolo e_1 e para a segunda câmera tem-se o epipolo e_2.

Se existem dois pontos m_1 e m_2 na primeira e na segunda imagem respectivamente, que são a projeção de um ponto M em coordenadas reais, pode-se dizer que M, C_1, C_2 são coplanares, formando o plano epipolar, como visto na Figura 5. Este plano intersecta com o plano de imagem de cada uma das câmeras formando as linhas epipolares. A Figura 5 ilustra a relação que as linhas epipolares possuem entre uma imagem e outra. Por uma análise, usando o ponto m_1 como referência, é possível definir um raio que parte de C_1 até m_1 . A partir deste raio, pode-se perceber que m_1 na realidade não é apenas a projeção de M, mas sim de todos os pontos que pertencem ao raio. Isto significa que é impossível determinar exatamente a posição espacial de um ponto projetado numa imagem sem que haja uma outra imagem, capturada por uma segunda câmera em uma outra posição. Neste exemplo, m_2 seria este segundo ponto de vista de M. Desta forma, a intersecção dos raios que vão de C_1 à m_1 e de C_2 à m_2 ocorreria no ponto M (Oram, 2001).

Se o primeiro raio for projetado na segunda imagem ele formará uma reta no plano projetivo, que é a linha epipolar correspondente ao ponto m_1 e esta reta contém o ponto m_2. O mesmo acontece se o raio de C_2 à m_2 for projetado na primeira imagem. Desta análise pode ser extraída mais uma importante conclusão: para todos os pontos de uma imagem, seu correspondente na outra figura estará na sua respectiva linha epipolar (Roberto et al., 2009).

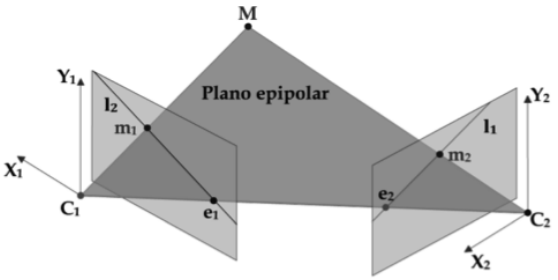


Figura 5. Geometria epipolar e seus principais elementos (Roberto et al., 2009).

Todas as linhas epipolares passam pelo epipolo da imagem e, independente da coordenada espacial do ponto M, todos os planos epipolares passarão pela *baseline*, como mostra a Figura 6. A partir de um ponto m_1 na primeira imagem, a linha epipolar l_1 na segunda imagem, que conterá o ponto m_2, pode ser achada a partir da seguinte relação:

$$l_1 = Fm_1 \tag{1}$$

Sendo F a matriz fundamental (uma representação algébrica da geometria epipolar entre duas imagens). Ela é uma matriz 3×3 que pode ser encontrada a partir da seguinte relação:

$$m_2^T = Fm_1 = 0 \tag{2}$$

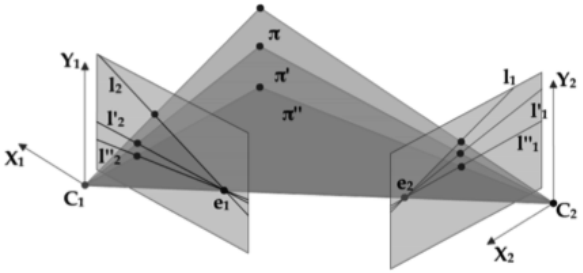


Figura 6. Vários planos epipolares, todos passando pela *baseline*, formando várias linhas epipolares, todas passando pelos epipolos (Roberto et al., 2009).

Para casos onde o objetivo é encontrar a linha epipolar l_2 na primeira imagem, correspondente ao ponto m_2 da segunda, a matriz fundamental também pode ser calculada:

$$l_2 = F^T m_2 \tag{3}$$

Outro papel importante da matriz fundamental é calcular os epipolos da imagem. Eles são definidos como os núcleos da matriz fundamental:

$$Fe_1 = 0 \qquad e \qquad F^T e_2 = 0 \tag{4}$$

Também é possível encontrar linhas epipolares correspondentes. Ou seja, dado que a equação da linha l_1 na segunda imagem é conhecida, encontrada a partir do ponto m_1 na primeira imagem, é possível encontrar a linha epipolar l_2 que contém o ponto m_1, mesmo sem conhecer o ponto m_2 na segunda imagem. Isto é possível porque existe uma matriz homográfica H que mapeia todos os pontos e retas da primeira imagem na segunda, assim como o contrário. Esta matriz é encontrada a partir da matriz fundamental e dos epipolos da imagem pela seguinte equação:

$$H = [e_2]_x F + e_2 a^T = 0 \tag{5}$$

sendo a um vetor qualquer não nulo, usado para garantir que a matriz H tenha uma inversa e $[e_2]_x$ é a matriz anti-simétrica do epipolo da segunda imagem, definida por:

$$[e_2]_x = \begin{bmatrix} 0 & -e_2z & e_2y \\ e_2z & 0 & -e_2x \\ -e_2y & e_2x & 0 \end{bmatrix} \tag{6}$$

Assim, conhecendo a matriz homografica, as linhas epipolares correspondentes podem ser facilmente calculadas usando:

$$I_2 = H^{-T} I_1 \qquad e \qquad I_1 = H^T I_2 \tag{7}$$

Homografias são estimadas entre imagens para detectar características correspondentes nessas imagens. Dentre os algoritmos capazes de estimar a matriz fundamental é possível citar o RANSAC.

Para a estimação da matriz fundamental, o RANSAC calcula as verdadeiras correspondências (*inliers*) para cada matriz H e escolhe a que maximiza esse número. Tendo eliminado os *outliers*, a matriz H é recalculada com o objetivo de melhorar a estimação.

Mesmo com a matriz homográfica calculada, para fazer a junção de um par de imagens é necessário retificar as imagens no sentido de minimizar distorções e suavizar as junções. As retificações utilizadas neste artigo foram: Planar e Cilíndrica. O cálculo da homografia em si com retificação, juntamente com a junção das imagens é feito pelo algoritmo da Transformação Linear Direta – DLT (Hartley & Zisserman, 2004).

2.5.1 Estimando a matriz fundamental

RANSAC é um procedimento de estimação robusto que usa um conjunto mínimo de correspondências amostradas, para estimar os parâmetros de transformação da imagem e achar a solução que tem o melhor consenso com os dados. Os métodos clássicos procuram utilizar o maior número de pontos para obter uma solução inicial e, então, eliminar os pontos inválidos. O RANSAC, ao contrário destes métodos, utiliza apenas o número mínimo e suficiente de pontos necessários para uma primeira estimativa, aumentando o conjunto com novos pontos consistentes sempre que possível (Fischler & Bolles, 1981).

Uma vantagem do RANSAC é a sua habilidade de realizar a estimativa de parâmetros de um modelo de forma robusta, ou seja, ele pode estimar parâmetros com um alto grau de acerto mesmo quando um número significativo de *outliers* esteja presente nos dados analisados. Uma desvantagem do algoritmo é que ele tem de possuir uma quantidade pré-estabelecida de iterações e com isso a solução obtida pode não ser a melhor existente.

Para o problema específico de remoção de *outliers* na correspondência de imagens, a Matriz Fundamental (H) pode ser determinada da seguinte maneira:

- Selecionar randomicamente um subconjunto de oito pontos correlacionados, retirados do conjunto total de pontos correlacionados;
- Para cada subconjunto, indexado por j, calcular a matriz fundamental F_j através do algoritmo de oito pontos;
- Para cada matriz F_j computada, determinar o número de pontos com distância até a linha epipolar, ou residual, menor que um limiar;
- Selecionar a matriz F que apresenta o maior número de pontos com residual inferior ao máximo definido;
- Recalcular a matriz F considerando todos os pontos *inliers*.

Uma visão mais detalhada da relação do RANSAC com a geometria epipolar é apresentada a seguir. Pela geometria epipolar, é possível calcular a matriz fundamental entre dois pares de imagens. Considere $m = [x, y, 1]$ um ponto sobre o plano da imagem L e $n = [x^{'}, y^{'}, 1]$ um ponto sobre o plano da imagem $L^{'}$. Assim, a equação 8 define a matriz fundamental.

$$m^T F n = 0 \tag{8}$$

Diversos métodos para estimação da matriz fundamental são encontrados na literatura, contudo o método mais conhecido é o algoritmo de 8 pontos. Tal método, dado um conjunto com $n \geq 8$ correspondências, estima a matriz fundamental de forma linear, solucionando a Equação 9.

$$\Sigma_{i=1}^{n} \| m_i^T F n_i \|^2 \tag{9}$$

A estimação robusta da matriz fundamental é feita pesando o residual para cada ponto. O resíduo é mostrado pela equação 10, sendo r o resíduo e i o número do par de pontos na lista de pontos correlacionados.

$$r_i = m_i^T F n_i \tag{10}$$

Para o cálculo das homografias neste projeto é selecionado um conjunto mínimo de $S = 4$ correspondências de características e o processo é repetido N vezes ($N = 200$) para um limiar t (distância máxima do modelo que um dado pode estar para ser considerado um *inlier*) de 4 *pixels*.

2.6 Retificação de imagens

Duas maneiras de retificação são conhecidas na literatura e ambas determinam que o par de imagens a ser retificadas deva ser reorganizado a partir de uma reprojeção. Os algoritmos diferem basicamente na forma como as imagens serão reprojetadas (Roberto et al., 2009).

O método tradicional de retificar um par de imagens consiste em reprojetar as imagens num plano em comum paralelo à *baseline*. Desta forma, quando a imagem for mapeada numa região em comum deste plano, têm-se a garantia que linhas epipolares correspondentes estarão na mesma altura (Fusiello et al., 2000). Conhecida como retificação planar, esta abordagem é relativamente simples de ser implementada. Porém, ela falha com alguns movimentos de câmera. Isto se deve ao fato de que, quanto mais próximo da imagem o epipolo estiver, maior será o tamanho da imagem retificada, culminando no caso extremo, onde o epipolo está localizado dentro da imagem, que resultaria numa imagem de tamanho infinito.

A segunda maneira, chamada de retificação cilíndrica, consegue tratar esses casos. Ela se diferencia da retificação planar principalmente por, ao invés de usar um plano em comum, usar um cilindro em comum para reprojetar o par de imagens. O método consiste em determinar um cilindro de raio unitário que tem a *baseline* com eixo de revolução e, em seguida, mapear cada *pixel* da imagem numa coordenada (z, θ) de um sistema de coordenadas cilíndricas, que pode ser usado normalmente, como um ponto (x, y) na imagem (Roy et al., 1997). Apesar de mais geral, esta técnica é bem mais complexa de ser implementada e possui um alto custo computacional, pois todos os cálculos realizados para cada *pixel* da imagem são feitos num espaço tridimensional.

A retificação cilíndrica, entretanto, pode ser simplificada se as informações das geometrias projetiva e epipolar forem usadas. Desta forma, é possível realizar todos os cálculos no plano de imagem, evitando assim operações tridimensionais. A ideia da retificação cilíndrica simplificada é muito semelhante à anterior, ou seja, reparametrizar a imagem num sistema de coordenadas cilíndrico. Entretanto, ela difere no cilindro escolhido. Enquanto na implementação convencional o cilindro é centrado na *baseline*, na forma simplificada as transformações ocorrem ao redor dos epipolos e, como estes estão no mesmo plano da imagem, nenhuma operação ocorrerá no espaço tridimensional (Pollefeys et al., 1999).

Como mostrado na Figura 7, cada linha epipolar possui um ângulo θ em relação ao epipolo, assim como cada *pixel* dela está a uma distância r deste mesmo ponto. Desta forma, as linhas epipolares são reescritas horizontalmente na nova imagem. No final, o par estará retificado porque linhas epipolares correspondentes possuem o mesmo ângulo em relação ao epipolo, já que elas estão no mesmo plano epipolar.

3. A Metodologia Empregada

Para se ter um padrão e reduzir o custo computacional foram selecionadas 20 imagens por vez para a construção dos mosaicos, resultando em 10 grupos de 20 imagens. A construção do mosaico se inicia pela primeira imagem (a esquerda do mosaico) em direção à última imagem (a direita do

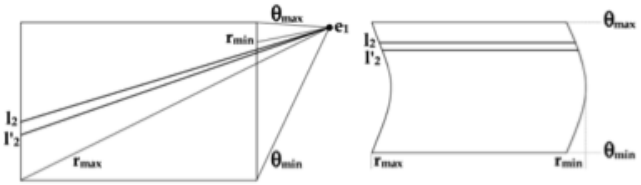


Figura 7. Retificação cilíndrica simplificada (Roberto et al., 2009).

mosaico), do respectivo grupo. As junções são feitas por pares de imagens, ou seja, imagem1 com imagem2, que na sequencia é juntada à imagem3 e assim por diante.

Como as imagens foram obtidas de maneira sequencial, a primeira etapa consiste em extrair as características do primeiro par de imagens pela SIFT, criar os descritores deste primeiro par, fazer as correspondências destas características (pelo método do vizinho mais próximo, com auxílio da BBF para acelerar este processo). Na sequencia é aplicado o RANSAC que estima a matriz homográfica, que é responsável por corresponder partes comuns da primeira imagem com a segunda do par de imagens, eliminando falsos pontos correspondentes (Figura 8).

Com isto é possível realizar a retificação (planar ou cilíndrica) no par de imagens para corrigir possíveis distorções de ângulo ou movimentação entre as imagens e fazer a interpolação nas imagens para que as mesmas possam ser unidas. Enfim, as imagens são unidas pelos pontos correspondentes restantes, após a aplicação do RANSAC e estimação da matriz homográfica. Um momento antes de cada junção ser efetivada é aplicado o algoritmo *Blend Feathering* para suavizar a região de junção. A Figura 9 mostra um exemplo de mosaico sem aplicação do *Blend Feathering*. Todo este processo é repetido para cada par de imagens, até formar um mosaico completo,

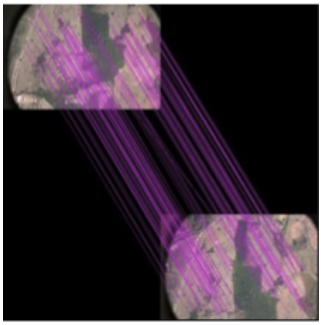


Figura 8. Exemplo de pontos correspondentes após aplicação do RANSAC.

Figura 9. Exemplo de mosaico sem aplicação do *Blend Feathering* nas junções.

incluindo as 20 imagens de cada grupo. Todo este processo é repetido para cada par de imagens, até formar um mosaico completo, incluindo as 20 imagens de cada grupo.

A Figura 10 apresenta um diagrama com a metodologia empregada neste projeto.

4. Resultados

Para gerar os resultados, primeiro foi feita uma análise nas retificações (planar e cilíndrica) para determinar qual é a mais indicada para este projeto. Depois foi realizada a montagem dos mosaicos utilizando a Metodologia Proposta (Projeto), a metodologia comercial livre desenvolvida por Brown & Lowe (2007) (*Autostitch*), e a metodologia comercial livre PTGui. Por fim, os resultados finais foram comparados para verificar a qualidade das junções e o tempo de processamento.

A Figura 11 apresenta os resultados gerados para um mesmo grupo de imagens (contendo 20 imagens), utilizando a retificação cilíndrica e planar.

Observando as retificações realizadas por um profissional da área, foi possível verificar que em 90% dos mosaicos gerados, as retificações planar e cilíndricas geraram resultados similares.

A Figura 12 apresenta um mosaico completo, composto por 20 imagens, que foi construído pelas metodologias citadas anteriormente. Os mosaicos gerados pelo Autostitch e pela Metodologia Proposta são visualmente semelhantes, como pode ser observado na Figura 12, já o mosaico gerado pelo PTGui mostrou-se diferente.

A Tabela 1 apresenta os tempos médios em segundos de processamento para cada metodologia gerar um mosaico contendo 20 imagens.

5. Discussão e Conclusões

Os testes de retificações realizados, conforme exemplo da Figura 11, tiveram uma alta porcentagem de resultados similares devido ao fato de que

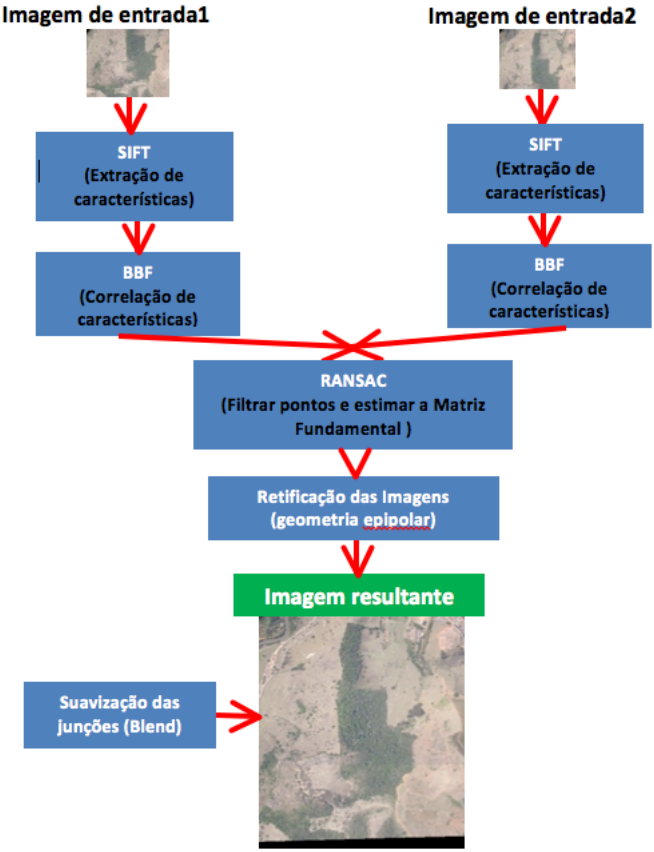


Figura 10. Diagrama da metodologia empregada.

Tabela 1. Tempo de processamento das metodologias usadas.

Metodologia	Tempo (s)
AutoStitch	48
PTGui	39
Metodologia Proposta	60

as transformações se comportaram de maneira estável, pelo motivo de que as imagens obtidas sequencialmente possuem pouca distorção angular entre uma e outra imagem; isto foi comprovado por inspeção visual, por um profissional da área. Para evitar o possível surgimento de uma imagem de

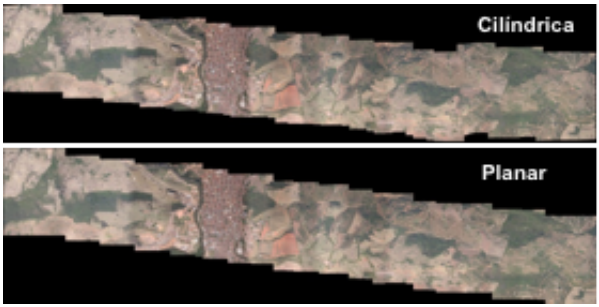


Figura 11. Exemplo da retificação cilíndrica e planar.

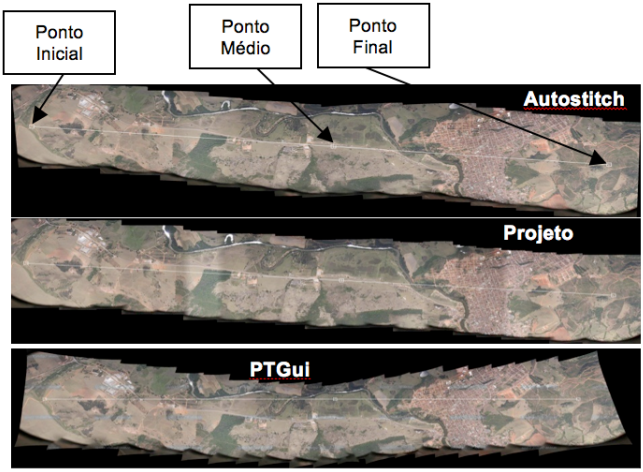


Figura 12. Exemplo de mosaicos gerados.

tamanho infinito (Seção 2.6) que reduz a precisão e qualidade do mosaico resultante, neste projeto foi utilizada a retificação cilíndrica.

Os mosaicos gerados pelas três metodologias (Autostitch, PTGui e Metodologia Proposta) apresentaram resultados visualmente similares, mas se for considerado questões de distorção para gerar os mosaicos, é possível verificar que os resultados das três metodologias diferem muito, como pode ser observado na Figura 13.

Para comprovar as distorções de cada metodologia, foram marcados 3 pontos comuns de junção (ponto inicial, ponto médio e ponto final), conforme pode ser observado pela linha branca na Figura 12 dada como exemplo. O ponto inicial e ponto final sempre foram marcados na mesma localização.

Na Figura 13 dada como exemplo, usando o *Google Earth*, foram feitas as mesmas marcações, nas mesmas posições que foram feitas na Figura 12 e mais uma marcação pelas coordenadas ideais do *Google Earth*. Pelas coordenadas de GPS, foi possível ter uma referência ideal dos pontos marcados do *Google Earth* em relação aos pontos marcados pelas três metodologias utilizadas para comparação. Com isto, foi possível comparar qual metodologia se aproxima mais da marcação do *Google Earth* (ideal) e, verificar qual metodologia gerou mais distorção. De acordo com a Figura 13, a linha vermelha corresponde às coordenadas do *Google Earth*, a linha azul a Metodologia Proposta (Projeto), a linha amarela ao AutoStitch e a linha verde ao PTGui. Na Tabela 2 são apresentadas as coordenadas GPS dos pontos marcados na Figura 13.

Pelas informações da Tabela 2, é possível verificar que os pontos médios pertencentes a cada metodologia variam e isto pode ser comprovado na Figura 13. As metodologias AutoStitch e PTGui distorcem as imagens para que as junções ocorram, não se preocupando muito com a precisão do mosaico, mas visualmente aparentam ser mosaicos de boa qualidade.

Tabela 2. Coordenadas GPS dos pontos marcados.

Coordenada GPS		**Ponto**	**Metodologia**
-22,269857	-45,771812	Inicial	
-22,252313	-45,729144	Médio	Autostich
-22,252913	-45,729219	Médio	Projeto
-22,247868	-45,730388	Médio	PTGui
-22,239154	-45,691300	Final	

Foi feita uma inspeção visual por um profissional da área nos 10 mosaicos gerados por cada metodologia empregada neste artigo. Então verificou-se qual metodologia se aproxima mais da coordenada ideal do *Google Earth* (Figura 9). Com todas estas verificações e inspeções, comprovou-se que em 85% dos casos é a Metodologia Proposta que se mais aproxima da referência ideal do *Google Earth*.

Com isto conclui-se que as metodologias comerciais dão ênfase a um baixo tempo de processamento ao invés da precisão do mosaico a ser gerado. De acordo com os resultados da Tabela 1 e pela análise apresentada anteriormente, é possível concluir que a Metodologia Proposta apresenta mosaicos com melhor qualidade e menor distorção em relação às metodologias comerciais, mas com tempo de processamento mais elevado.

Como trabalhos futuros, pode-se citar a implementação de processamento paralelo na etapa de extração de características para ganho de desempenho, uma vez que esta etapa é a mais custosa computacionalmente, a implementação de processamento paralelo para unir vários pares de imagens simultaneamente e a construção de mosaicos de imagens georreferenciados.

Figura 13. Comparação das metodologias utilizadas.

Referências

Bei, L. & Haizhen, Z., An algorithm of fabric image mosaic based on SIFT feature matching. In: *Proceedings of IEEE International Conference on Artificial Intelligence and Computational Intelligence*. Piscataway, USA: IEEE Press, v. 3, p. 435–438, 2009.

Beis, J.S. & Lowe, D.G., Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: *Proceedings of Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE Computer Society, p. 1000–1006, 1997.

Brown, M. & Lowe, D.G., Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.

Fischler, M.A. & Bolles, R.C., Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

Fusiello, A.; Trucco, E. & Verri, A., A compact algorithm for rectification of stereo pairs. *Journal Machine Vision and Applications*, 12(1):16–22, 2000.

Hartley, R. & Zisserman, A., *Multiple View Geomerty in Computer Vision*. Cambridge, UK: Cambridge University Press, 2004.

Hong, J.; Lin, W.; Zhang, H. & Li, L., Image mosaic based on SURF feature matching. In: *Proceedings of 1st International Conference on Information Science and Engineering*. Piscataway, USA: IEEE Press, p. 1287–1290, 2009.

Juan, L. & Gwun, O., A comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing*, 3(4):143–152, 2009.

Ke, Y. & Sukthankar, R., PCA-SIFT: A more distinctive representation for local image descriptors. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Los Alamitos, USA: IEEE Computer Society, v. 2, p. 506–513, 2004.

Li, L. & Geng, N., Algorithm for sequence image automatic mosaic based on SIFT feature. In: *Proceedings of WASE International Conference on Information Engineering*. Los Alamitos, USA: IEEE Computer Society, v. 1, p. 203–206, 2010.

Lowe, D.G., Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

Oram, D., Rectification for any epipolar geometry. In: *Proceedings of British Machine Vision Conference*. Manchester, UK: BMVA, p. 653–662, 2001.

Pollefeys, M., *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. Phd. thesis, Departement Elektrotechniek, Katholieke Universiteit Leuven, Leuven, Belgium, 1999.

Pollefeys, M.; Koch, R. & van Gool, L., A simple and efficient rectification method for general motion. In: *Proceedings of Seventh IEEE International Conference on Computer Vision*. Piscataway, USA: IEEE Press, v. 1, p. 496–501, 1999.

Roberto, R.A.; Teichrieb, V. & Kelner, J., Retificação cilíndrica: um método eficiente para retificar um par de imagens. In: Conci, A.; Silva, L. & Lewiner, T. (Eds.), *Proceedings of XXII SIBGRAPI Workshops – Undergraduate Work*. Rio de Janeiro, RJ: SBC, 2009.

Roy, S.; Meunier, J. & Cox, I., Cylindrical rectification to minimize epipolar distortion. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE Computer Society, p. 393–399, 1997.

Notas Biográficas

André de Souza Tarallo é graduado em Engenharia de Computação (UNIFEV, 2004) e mestre em Engenharia Elétrica na área de processamento digital de imagens (USP São Carlos, 2007). Atualmente é doutorando na mesma área do mestrado na USP São Carlos e Embrapa Instrumentação.

Francisco Assis da Silva é graduado em Ciência da Computação (UNOESTE, 1998) e mestre em Computação na área de Processamento Digital de Imagens (UFRGS, 2002). Atualmente é doutorando na mesma área na USP São Carlos.

Alan Kazuo Hiraga é graduado em Ciência da Computação (UNOESTE, 2011) e atualmente é mestrando em Computação na UFSCar.

Maria Stela Veludo de Paiva é graduada em Engenharia Elétrica (USP, 1979), mestre e doutor em Física Aplicada (USP São Carlos, 1984 e 1990, respectivamente) e tem pós-doutorado (University of Southampton, 1992). Atualmente é Professor Associado do Departamento de Engenharia Elétrica da USP Carlos.

Lúcio André de Castro Jorge é graduado em Engenharia Elétrica (Faculdade de Engenharia de Barretos, 1987), mestre em Ciência da Computação (USP São Carlos, 2001) e doutor em Engenharia Elétrica na área de processamento digital de imagens (USP São Carlos, 2011). Atualmente é pesquisador da EMBRAPA Instrumentação em São Carlos.

Este livro é composto de uma série de capítulos abordando assuntos relacionados à visão computacional, com foco em análise de imagens médicas, agronomia, biometria, processamento de vídeo, reconhecimento de caracteres, segmentação e visualização.

omnipax editora

ISBN 978-85-64619-09-8

9 788564 619098