Universidade Federal Fluminense

LINCOLN FARIA DA SILVA

Distinção Automática de Texto Impresso e Manuscrito em uma Imagem de Documento

NITERÓI
2009

LINCOLN FARIA DA SILVA

Distinção Automática de Texto Impresso e Manuscrito em uma Imagem de Documento

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do título de Mestre. Área de concentração: Computação Visual e Interfaces.

Orientadora:
Aura Conci

Universidade Federal Fluminense

NITERÓI
2009

Ficha Catalográfica elaborada pela Biblioteca da Escola de Engenharia e Instituto de Computação da UFF

S586 Silva, Lincoln Faria da.
Distinção automática de texto impresso e manuscrito em uma imagem de documento. / Lincoln Faria da Silva. – Niterói, RJ : [s.n.], 2009.
100 f.

Orientador: Aura Conci.
Dissertação (Mestrado em Computação) - Universidade Federal Fluminense, 2009.

1. Mineração de dados (Computação). 2. Análise de documento. 3. Visão computacional. 4. Computação visual. I. Título.

CDD 005.74

Distinção Automática de Texto Impresso e Manuscrito em uma Imagem de Documento

Lincoln Faria da Silva

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do título de Mestre. Área de concentração: Computação Visual e Interfaces.

Aprovada por:

_____

Profª. Aura Conci – IC/UFF (Presidenta)

_____

Profª. Débora Christina Muchaluat Saade – TET/UFF

_____

Prof. Creto Augusto Vidal – LIA/UFC

Niterói, 20 de março de 2009

Ao Deus soberano, criador de todo o universo. Esteve sempre ao meu, deu-me capacidade e sabedoria necessária para o desenvolvimento deste trabalho.

Tudo o que tenho e tudo o que sou devo a Ele.

Agradecimentos

A Aura Conci, que depositou em mim sua confiança e compartilhou comigo sua experiência e conhecimento, profissional e acadêmico, na tarefa de orientar-me. Tenho aprendido muito com ela.

A minha noiva, Bianca dos Santos Maciel, pela compreensão, paciência e apoio durante todo o tempo, e pela ajuda na revisão do texto.

Quero agradecer também a minha mãe, Maria das Graças Faria da Silva, e minha avó, Vitória Ghiotti Faria (em memória), que me educaram, superando situações adversas.

E, por fim, quero agradecer a todos os familiares, amigos e professores que, direta ou indiretamente, me ajudaram e me apoiaram, em especial a minha tia Glacy e ao meu tio Iltom.

Resumo

As metodologias de reconhecimento de texto manuscrito e texto escrito por máquina são totalmente diferentes. Por isso, é importante separar esses dois tipos de texto, em imagens de documentos nas quais eles aparecem juntos, antes de enviá-los aos seus respectivos sistemas de reconhecimento. Documentos que apresentam texto manuscrito e texto impresso, concomitantemente, não são poucos, sendo alguns deles: formulários, cartas, requerimentos, memorandos, envelopes postais e cheques bancários.

O trabalho aqui desenvolvido executa essa separação por meio de regras de classificação mineradas, na fase de treinamento, de um conjunto de dados, os quais representam as características de cada tipo de texto. Primeiramente, a imagem é pré-processada por várias técnicas com a finalidade de: eliminar ruídos, separar o texto do fundo, retirar linhas horizontais e suavizar os contornos verticais das letras das palavras. Em seguida, é realizada a extração de componentes conectados e cada um desses é cercado pelo menor retângulo capaz de contê-lo. Retângulos próximos ou sobrepostos são unidos de modo a formarem palavras. Dos retângulos já unidos são calculadas as características pré-definidas. Essas características têm a função de representar a palavra dentro de cada retângulo e apresentam valores diferentes para as impressas e as manuscritas. Os valores são usados nas regras de classificação, as quais decidem se um retângulo contém uma palavra impressa ou manuscrita. O sistema desenvolvido é testado em duas bases de imagens: na AIM off-line Database 3.0 e na base de imagens de formulários cadastrais criada durante este trabalho e disponível na Internet. Na primeira, a acurácia e a precisão do sistema foi de 100% em 45% das imagens, com acurácia média de 97,55% e precisão média de 96,70% em relação às palavras impressas e com acurácia média de 98,09% e precisão média de 98,10% em relação às manuscritas. Na base criada como parte do trabalho, a acurácia e a precisão do sistema foi

de 100% em 33,33% das imagens, com acurácia média de 97,17% e precisão média de 98,85% em relação às palavras impressas e com acurácia média de 99,46% e precisão média de 98,75% em relação às manuscritas. O trabalho desenvolvido apresenta vantagens quando comparado com outro trabalho, que também utilizou para testes a base de imagens AIM off-line Database 3.0. O sistema foi implementado em C++ e compilado usando o GCC. Ele foi executado em uma máquina equipada com o processador AMD Athlon™ MP 900Mhz consumindo 74 segundos, em média, para realizar 184,04 bilhões de instruções no processamento de cada imagem.

Palavras-chaves: Mineração de Dados, análise de documento, identificação de texto, reconhecimento óptico de caractere, Visão de Máquina.

Abstract

The printed text and handwriting recognition methods are totally different. That is why, it is important to separate those two text types, which appear together in a document image, before sending them to their respective recognition systems. The number of documents which present printed text and handwriting, simultaneously, is significant, for example: Forms, letters, requirements, memorandums, envelopes you post and bank checks.

The separation process proposed in this work uses classification rule mining, on the training phase, of a data set, which represent the characteristics of each type of text. Initially, the image is preprocessed by applying different techniques aimed at: Eliminating noises, separating the text from the background, removing horizontal lines, and smoothing the vertical contours of the words' characters. Then, the extraction of connected components is performed and, for each connected component identified, a bounding rectangle is defined. Neighboring or overlapping bounding rectangles are united in order to form words. Predefined characteristics are computed from the already united rectangles. Those characteristics have the function of representing the word within each rectangle and they present values which are different for printed and handwriting words. Those values are used in the classification rules which decide if a given rectangle contains a printed or a handwritted word. The developed system is applied to two image databases: AIM off-line Database 3.0 and cadastral forms image database constructed simultaneously with this work and available on the Internet. On first, the system's accuracy and precision was of 100% in 45% of the images, with average accuracy of 97.55% and average precision of 96.70% in relation to printed words and with average accuracy of 98.09% and average precision of 98.10% in relation to handwritten. On database constructed as part of this work, the system's accuracy

and precision was of 100% in 33.33% of the images, with average accuracy of 97.17% and average precision of 98.85% in relation to printed words and with average accuracy of 99.46% and average precision of 98.75% in relation to handwritten. This work presents advantages when compared with another work, which also used the AIM off-line Database 3.0 as its test database. The system was implemented using C++ and compiled with GCC. It was carried out in a machine equipped with the AMD Athlon™ MP 900Mhz processor consuming 74 seconds, on the average, in order to perform 184.04 billions of instructions in the process of each image.

Keywords: Data Mining, document analysis, text identification, optical characters recognition, Machine Vision.

Glossário

AIM – Institute of Computer Science and Applied Mathematics

API – Application Programming Interface

ARFF – Attribute-Relation File Format

DCBD – Descoberta de Conhecimento em Bancos de Dados

GCC – GNU Compiler Collection

IDE – Integrated Development Environment

OCR – Optic Character Recognition

VC – Visão Computacional

WEKA – Waikato Environment for Knowledge Analysis

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Algoritmos

1. INTRODUÇÃO

1.1 Motivação

Nos dias atuais, informações são produzidas e armazenadas utilizando tecnologias eletrônicas. Assim, executando ferramentas de busca, tais informações podem ser acessadas facilmente por qualquer pessoa, em qualquer lugar do mundo, bastando que elas estejam disponíveis em uma rede como, por exemplo, a Internet. Porém, uma grande quantidade de informações ainda se encontra e continua sendo armazenada em papel, como: formulários, memorandos, cartas, requerimentos, cheques bancários etc; e o processo de conversão de informações armazenadas em papel para a forma eletrônica é caro e pouco produtivo quando realizado manualmente. A solução para esse problema é dotar os computadores da capacidade de "ler" documentos, tal "leitura" é denominada de reconhecimento óptico de caractere (Optic Character Recognition).

Metodologias de OCR para caracteres impressos por máquina e caracteres manuscritos são totalmente diferentes [Govindan & Shivaprasad, 1990], [Impedovo et al., 1992], [Sampaio et al., 1999], [Koerich, 2004], [L´Homer, 2000], [Kapp, 2004], [Aires, 2005]. Por isso, em documentos que apresentam os dois tipos, existe a necessidade de um estágio anterior à tarefa de OCR. Esse estágio consiste na distinção de quais são caracteres impressos por máquina e de quais são manuscritos. A distinção implica em taxas maiores de reconhecimento de cada tipo, pois, quando enviados aos seus respectivos sistemas de OCR, as chances de um reconhecimento correto aumentam bastante.

O reconhecimento óptico de caracteres é importante não só por converter informações armazenadas em papel para a forma eletrônica, mas também por ser o estágio

inicial de outras tarefas, tais como autenticação de assinaturas, leitura de cheques bancários e determinação de autoria de escrita.

1.2 Metodologia proposta

A metodologia proposta no presente trabalho executa a distinção de texto impresso e texto manuscrito em uma mesma imagem de documento. Para esse fim, são utilizadas regras de classificação baseadas em características extraídas de retângulos envoltórios contendo os dois tipos de textos. As características extraídas são: Desvio da Largura, Desvio da Altura, Desvio da Área, Densidade, Variância da Projeção Vertical, Maior Diferença Encontrada na Projeção Horizontal, Distribuição de Pixels, Divisão da Linha Inferior de Pixels pela Largura, Soma das Divisões de Pixels de Cada Linha pela Largura, Divisão do Maior Contorno Vertical pela Altura e Divisão da Soma dos Comprimentos dos Contornos Verticais pela Área. Elas apresentam valores diferentes para cada tipo de texto nos retângulos envoltórios. Isso é o que torna possível a distinção realizada pelo sistema.

Primeiramente, uma filtragem espacial é executada com o objetivo de reduzir os ruídos na imagem, oriundos de seu processo de captura e digitalização. Em seguida, o texto é separado do fundo com a aplicação de uma técnica de limiarização automática. Então, a imagem é submetida a operações morfológicas com a finalidade de eliminar ruídos remanescentes da fase anterior e suavizar contornos verticais das letras no texto. Esses contornos são importantes em algumas características extraídas. Depois, a extração de componentes conectados é realizada e cada um é cercado por um retângulo envoltório. Feito isso, retângulos envoltórios próximos ou sobrepostos, por alguma porção de área, são unidos formando palavras. Logo após, as características citadas são calculadas para cada retângulo envoltório e as regras de classificação, mineradas na fase de treinamento, aplicadas para

decidir se um retângulo envoltório contém uma palavra impressa ou manuscrita, de acordo como os valores calculados.

Para a obtenção das regras de classificação, é utilizada a ferramenta WEKA [WEKA, 1999]. Essa ferramenta é uma coleção de algoritmos de aprendizagem de máquina, popular e gratuita, utilizada nas tarefas de mineração de dados e desenvolvida pela University of Waikato, New Zealand, na linguagem Java.

Utilizando o ambiente de desenvolvimento integrado wx-DevC++ [DevC++], versão 6.10.2, a metodologia foi implementada na linguagem de programação C++ e o código compilado com o GCC [GCC, 1988]. O sistema contém rotinas OpenGL [OpenGL], uma interface de software para hardwares gráfico, para a manipulação e exibição das imagens de documentos. Para o treinamento e teste do sistema, foram utilizadas duas bases de imagens, a AIM off-line Database 3.0, descrita por Marti e Bunke [1999, 2002] e outra criada como parte do trabalho. Ambas são constituídas por formulários contendo palavras impressas e manuscritas. Imagens desses de cada uma das bases são exibidas no Capítulo 4. Na fase de teste, foi aplicado o método de validação cruzada (*k-Fold Cross Validation*) obtendo-se resultados satisfatórios. Os avaliadores da classificação utilizados foram: a acurácia, a precisão, verdadeiro positivo, falso positivo, verdadeiro negativo, falso negativo, a sensibilidade, a especificidade e o desvio padrão. No Capítulo 5 são expostos os resultados dos testes. O trabalho desenvolvido apresenta vantagens quando comparado com outro de mesmo objetivo (distinção de palavras impressas e palavras manuscritas em uma imagem de documento), que também utilizou para testes a base de imagens AIM off-line Database 3.0. Detalhes dessa comparação estão no Capítulo 6. O sistema foi executado em um computador equipado com o processador AMD Athlon™ MP 900Mhz, onde, para cada imagem, o tempo médio de processamento foi de 74 segundos na realização de 184,04 bilhões de instruções. Resultados completos dos testes em cada base de imagens estão no Apêndice A deste trabalho.

1.3 Organização da dissertação

O Capítulo 2 apresenta uma breve fundamentação teórica relacionada à metodologia proposta, a qual é detalhadamente descrita no Capítulo 4. Trabalhos anteriores de distinção de palavras impressas e palavras manuscritas em uma imagem de documento são revisados no Capítulo 3. O Capítulo 5 descreve os bancos de imagens usados, relata o treinamento, testes e resultados da metodologia. Finalmente no Capítulo 6 são apresentadas as conclusões e idéias de trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são abordados os tópicos teóricos aplicados no presente trabalho. A primeira seção traz um panorama geral da Visão Computacional descrevendo seus vários estágios. Isso é importante porque o sistema aqui desenvolvido realiza uma tarefa dessa área. As demais seções abordam técnicas da Visão Computacional exclusivamente aplicadas na realização da metodologia proposta. Por fim, é discutida a descoberta de conhecimento em bancos de dados, a mineração dos dados, a tarefa de classificação, utilizando regras, e seus avaliadores aqui aplicados.

2.1 Visão Computacional

O ser humano obtém grande parte das informações do mundo que o cerca através de imagens, sejam elas apenas em sua vida pessoal ou realizando tarefas profissionais e científicas. E isso é um fator importante de crescimento das aplicações das imagens digitais [Conci et al., 2008].

Essas aplicações são encontradas, por exemplo, na medicina, transformando sinais em informações visuais para interpretação humana. As imagens biomédicas sofrem processos computacionais como a melhoria de contraste e a codificação dos níveis de intensidades em cores, a fim de facilitar a interpretação por um profissional. Imagens aéreas ou de satélites são processadas computacionalmente para que geógrafos possam estudar padrões de poluição. Imagens borradas de figuras fotografadas, que eram registros únicos de artefatos raros, perdidos ou danificados, têm sido restauradas por técnicas de processamento de imagens [Gonzalez & Woods, 1992]. Outras aplicações são: o reconhecimento automático de assinaturas, de pessoas e de objetos, a inspeção de produtos no controle de qualidade de uma

indústria [Conci et al., 2008], identificação automática de veículos, reconhecimento de impressões digitais, análise de imagens de raio X, monitoração de plantio e reconhecimento automático de caracteres de documentos [Gonzalez & Woods, 1992].

A Visão Computacional é responsável por identificar e classificar os objetos presentes nas imagens, utilizando informações extraídas, após análise das mesmas. Nesse caso, ela deve fornecer aos computadores a capacidade de "compreenderem" o conteúdo de uma imagem [Conci et al., 2008], ou seja, dotar os computadores de capacidade visual cognitiva, como a dos humanos, para realizarem tarefas muitas vezes exaustivas e/ou em lugares perigosos onde a vida de um profissional estaria em risco.

2.1.1 Principais etapas

As principais etapas que compõem um sistema de Visão Computacional estão representadas no esquema da Figura 2.1.

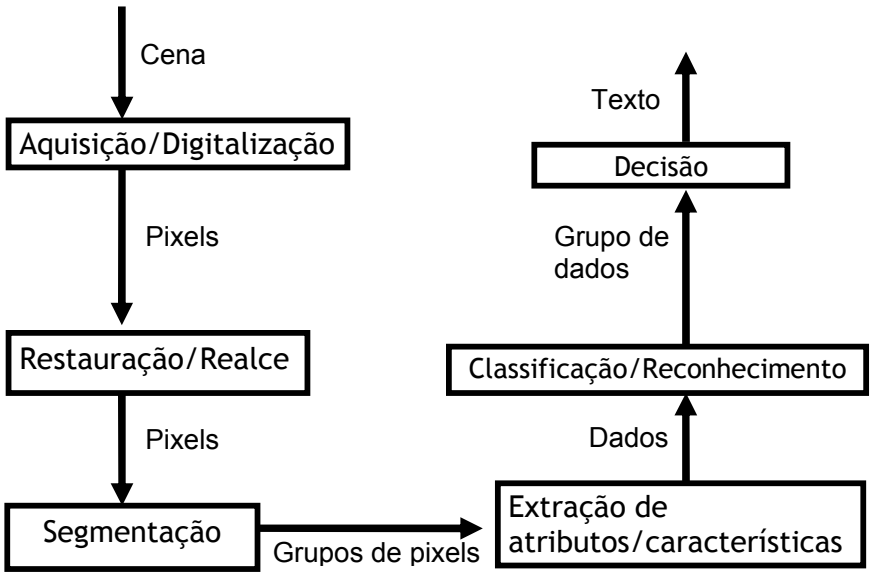


Figura 2.1. Etapas de um sistema de VC genérico [Conci et al., 2008].

2.1.1.1 Aquisição da imagem

Uma imagem pode ser considerada como uma distribuição de energia luminosa no espaço, representando visualmente objetos ou cenas, os quais podem ser adquiridos ou gerados. Ela pode apresentar-se como unidimensional, bidimensional ou tridimensional quanto à forma de representação e armazenamento. Em relação à representação da informação de cada um dos seus pontos, chamados de *pixels*, ela pode ser binária, monocromática, multibanda ou colorida [Conci et al., 2008].

Para aquisição de imagens, é necessário algum dispositivo de captura que, em geral, transforma uma cena em sinais elétricos e esses sinais em dados, os quais são armazenados em algum tipo de mídia utilizável pelo computador. Os dispositivos de captura mais comuns são as câmeras, onde a mais usada atualmente é a CCD (Charge Couple Device), e os escaneadores. Entretanto, existem muitos outros direcionados às aplicações nas áreas médica e industrial como, por exemplo, para a captura de imagens de exames médicos e imagens usadas no setor de inspeção de qualidade.

Em suma, o processo de digitalização de uma imagem consiste em representar uma cena real, a qual possui um número infinito de cores e pontos de coordenadas contínuas, em uma imagem onde os pontos possuem coordenadas discretas e as possibilidades de informações para cada um desses pontos são finitas. Desse processo, resultam dois conceitos importantes: a amostragem e a quantização.

A amostragem determina quantos pontos (pixels) serão usados para representar a imagem. Logo, quanto mais pontos para uma mesma área de imagem mais detalhes serão percebidos por quem a vê. A quantização determina quantos bits (cores ou tons de cinza) estarão disponíveis para cada um dos pontos.

Um modelo matemático utilizado para representar uma imagem bidimensional $f(x, y)$ é descrito por:

$$f(x,y) = i(x,y) \cdot r(x,y) \tag{2.1}$$

onde $i(x,y)$ é função do nível de iluminação do ambiente no qual se localiza o objeto, $r(x,y)$ é uma função que indica o quanto dessa iluminação o objeto reflete e (x,y) são as coordenadas de cada ponto na imagem.

Uma imagem digital bidimensional pode ser pensada como uma matriz *MXN*. Cada elemento da matriz é identificado com um único ponto (pixel) da imagem. O valor desse elemento relaciona-se ao nível de iluminação ou cinza (se tons de cinza são utilizados para representar a imagem) naquele ponto. Nesse caso, o valor é $z = f(x,y)$, onde x representa uma linha e y uma coluna da matriz. Então, uma imagem pode ser representada na forma:

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,M-1) \\ f(1,0) & f(1,1) & \cdots & f(1,M-1) \\ \vdots & \ddots & & \vdots \\ f(N-1,0) & f(N-1,1) & \cdots & f(N-1,M-1) \end{bmatrix} \tag{2.2}$$

2.1.1.2 Pré-processamento

O objetivo desta etapa é “melhorar” a imagem, modificando-a de tal forma que aumente as chances de sucesso dos processos subsequentes [Gonzalez & Woods, 1992]. Essa etapa envolve rearranjo dos pixels, alterando intensidades, ou transferindo valores entre eles, ela tem pixels como entrada e saída do processo [Conci et al., 2008] (Figura 2.1).

Técnicas de realce, de restauração e de remoção dos ruídos da imagem estão presentes no pré-processamento. Os ruídos podem ser introduzidos na imagem no processo de sua aquisição e podem comprometer o resultado final de determinada tarefa a ser realizada. Eles também podem ser causados pelos processos de armazenamento, transmissão, quantização e amostragem.

As técnicas de realce podem ser divididas em duas grandes categorias: técnicas no domínio espacial e técnicas no domínio da frequência. O primeiro refere-se à representação na forma da própria imagem e o segundo refere-se à representação da imagem após o uso das transformadas de Fourier.

As técnicas no domínio espacial, por atuarem diretamente nos pixels da imagem processada, podem ser expressas como funções da forma:

$$g(x,y) = T\left[f(x,y)\right] \tag{2.3}$$

onde $f(x,y)$ é a imagem original, $g(x,y)$ é a imagem processada e T é um operador que age sobre a imagem. Essas técnicas subdividem-se em técnicas de realce ponto a ponto ou por áreas. No realce ponto a ponto, apenas a intensidade do pixel isoladamente é considerada na operação, enquanto, na outra forma, consideram-se as intensidades dos pixels vizinhos ao pixel "visitado", pelo operador, através do uso de filtros (máscaras ou janelas). Usam-se subimagens, geralmente quadradas, para essas áreas (Figura 2.2-a) em que seu centro percorre toda a imagem (Figura 2.2-b), pixel por pixel, realizando uma transformação pré-determinada. Negativos de imagens, aumento de contraste e processamento de histogramas são exemplos de realce ponto a ponto. Filtragem espacial passa-alta, passa-baixa e por mediana são exemplos do segundo tipo mencionado.

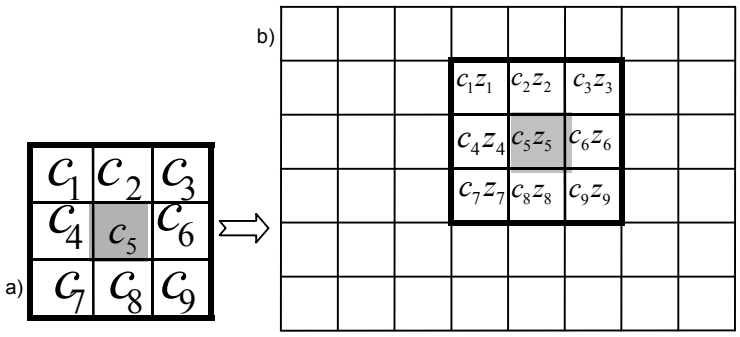


Figura 2.2. Filtragem espacial de uma imagem digital.

No filtro genérico mostrado na Figura 2.2-a, os valores $c_1, c_2, \ldots, c_9$ são coeficientes, definidos de acordo com a tarefa que o filtro exercerá sobre a imagem a ser processada. O centro do filtro é posicionado em um determinado pixel da imagem (Figura 2.2-b) e a tonalidade desse pixel é substituída pelo valor *R* da equação:

$$R = c_1 z_1 + c_2 z_2 + \cdots + c_9 z_9 \tag{2.4}$$

onde $z_1, z_2, \ldots, z_9$ são as tonalidades dos seus oito pixels vizinhos mais a sua, z_5. Esse processo é repetido para todos os pixels da imagem.

Qualquer filtro espacial baseado em uma equação, como a (2.4), é um filtro linear. Caso contrário, é considerado não-linear. Um exemplo de filtro não-linear é o de mediana (abordado com detalhes na Seção 2.2).

O filtro da Figura 2.2-a tem a forma de um quadrado com lado de medida 3 pixels. Porém a forma e o tamanho de um filtro são definidos de acordo com os objetivos do processamento a ser realizado.

Técnicas no domínio da frequência são aplicadas na imagem após a transformação de sua representação do domínio espacial para o domínio da frequência pela transformada de Fourier. Porém, a visualização do resultado só é possível após a aplicação da transformada inversa de Fourier, que retorna a representação para o domínio espacial, onde esse resultado é compreendido pelos olhos humanos. A Figura 2.3 ilustra as etapas citadas. Maiores detalhes em [Conci et al., 2008] ou em [Gonzalez & Woods, 1992].

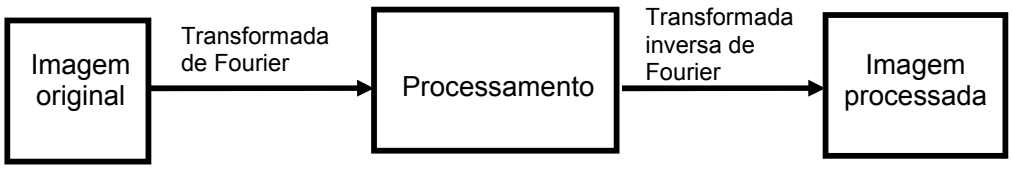


Figura 2.3. Etapas no processamento de imagens no domínio da frequência.

2.1.1.3 Segmentação

Um outro passo de processamento na Visão Computacional é a segmentação da imagem, a qual divide a imagem em suas partes ou objetos constituintes. Ela isola do restante da imagem as partes ou os objetos de interesse da aplicação. Por isso, esse passo do processamento é direcionado pelo objetivo da aplicação, pois seu sucesso depende de uma segmentação adequada. Uma segmentação adequada e ao mesmo tempo autônoma é uma das tarefas mais difíceis [Gonzalez & Woods, 1992].

As diversas técnicas existentes para segmentação de uma imagem, geralmente, baseiam-se na descontinuidade ou na similaridade das propriedades básicas de valores de níveis de cinza, de cores e de texturas. A descontinuidade considera as mudanças bruscas desses valores na imagem para reparti-la, enquanto a similaridade baseia-se no quanto eles podem ser comuns aos diversos pixels da imagem [Conci et al., 2008].

Dentre as técnicas baseadas na descontinuidade estão aquelas com interesse em detecção de pontos isolados, linhas e bordas na imagem. E, na similaridade, técnicas de (1) limiarização, (2) crescimento de região e (3) divisão e fusão de regiões [Gonzalez & Woods, 1992].

A Figura 2.4 ilustra uma segmentação baseada em descontinuidade. A passagem de uma região clara, o fundo, para uma região escura, o objeto, é marcada por uma mudança brusca de tonalidade. Na Figura 2.4-a, a imagem original e na Figura 2.4-b, o resultado da segmentação com os contornos dos objetos em destaque.

Na Figura 2.5, o texto é separado do fundo devido à tonalidade de seus pixels. A imagem da Figura 2.5-a é a original e a da Figura 2.5-b, o resultado da segmentação por similaridade.

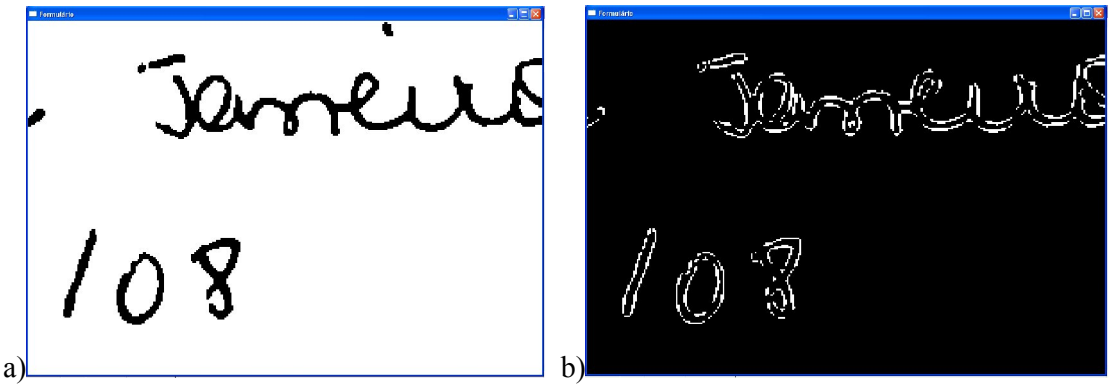


Figura 2.4. Segmentação baseada em descontinuidade.

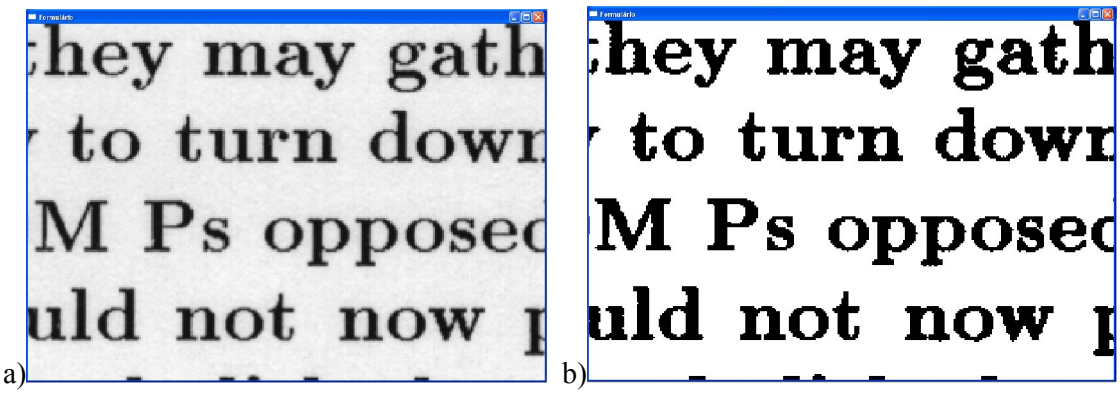


Figura 2.5. Segmentação baseada em similaridade.

A segmentação é obtida mediante as características comuns encontradas nestas partes ou objetos (texturas semelhantes, formas predefinidas etc.). Essas áreas de interesse, onde, posteriormente, atributos são extraídos e cálculos de parâmetros descritivos são realizados, têm seus pixels agrupados e destacados dos demais [Conci et al., 2008]. Na tarefa de distinção de palavras impressas e manuscritas em documentos, por exemplo, os pixels que constituem o texto devem ser identificados e posteriormente destacados.

Muitos trabalhos consideram a segmentação do texto em documentos, como os de Mahadevan e Nagabushnam [1995], Manmatha e Rothfeder [2005], Mital e Leng [1996], Nicolas et al. [2004], Gllavata et al.[2004] e Vellasques [2006].

2.1.1.4 Extração de características

Depois de a imagem ser segmentada em objetos, torna-se possível representar e descrever essas regiões de pixels de forma adequada à próxima etapa. A representação pode ser feita em relação às características do contorno do objeto, quando a forma detiver as informações mais significativas, ou em relação às características internas, quando a área interior do objeto for o foco da atenção do sistema [Gonzalez & Woods, 1992]. Medidas, geometria, propriedade de luminosidade e textura são características dos objetos usadas com frequência. O comprimento, as larguras mínima e máxima, a área, o perímetro e o centro de gravidade podem ser considerados medidas características, as quais descrevem o objeto ou a região. Circularidade, retilineidade, concavidade e eixos principais são exemplos de características geométricas. Cores e níveis de intensidade média de cada banda da região são características luminosas [Conci et al., 2008]. Para descrição de regiões de texturas são aplicadas, principalmente, abordagens estatísticas, estruturais e espectrais. As estatísticas caracterizam essas regiões como suave, áspera, granular etc. As estruturais baseiam-se em elementos geométricos como a quantidade média de linhas paralelas com espaçamentos regulares. As espectrais baseiam-se em propriedades do espectro de Fourier [Gonzalez & Woods, 1992].

Sejam quais forem as características escolhidas para a descrição, é desejável que sejam descorrelacionadas entre si e invariantes quanto à translação, mudança de escala e rotação [Conci et al., 2008].

2.1.1.5 Classificação e reconhecimento

Com os objetos (ou regiões) da imagem segmentados e descritos pelas características escolhidas e extraídas, o sistema de Visão Computacional passa para a etapa de

classificação e reconhecimento. Nela, os objetos são classificados ou reconhecidos de acordo com suas características, pois os de mesma classe devem possuir características semelhantes.

Antes que um novo objeto seja apresentado ao sistema de Visão Computacional para seu reconhecimento, esse sistema passa por uma fase de treinamento com objetos de classe previamente conhecida, ou seja, objetos anteriormente classificados por um especialista no domínio da área. Assim, um objeto fora do conjunto de treinamento e, portanto, desconhecido quanto a sua classe, é apresentado ao sistema e classificado de acordo com suas características.

Para a tarefa de classificação há várias técnicas de decisão disponíveis. São abordagens que vão desde simples critérios, como funções discriminantes e distância mínima, até o uso de inteligência artificial (IA) com a aplicação de algoritmos genéticos, clusterização, lógica nebulosa (fuzzy) ou redes neurais artificiais. Uma seleção prévia e adequada dos atributos torna a fase de classificação menos complexa [Conci et al., 2008].

2.1.1.6 Decisão

Um processo automático de análise só faz sentido se um grande volume de informações é processado. Esse volume, quando organizado, é geralmente denominado de banco de dados.

Após a fase de classificação ou reconhecimento, uma grande quantidade de informações extraídas de imagens passa a estar disponível. Dentro dessa grande quantidade, há conhecimento importante para a tomada de decisões relativas à execução de uma determinada tarefa ou área científica. Porém, esse conhecimento não está explícito no conjunto de informações e deve ser adquirido ou extraído.

A área da computação responsável pela descoberta desse conhecimento no banco de dados é a DCBD (Descoberta de Conhecimento em Bancos de Dados), detalhada na Seção

2.6. Contudo, a complexidade da mineração de conhecimento em banco de imagens é considerável, já os bancos de dados "convencionais" são menos complexos. Com a descoberta realizada o sistema de Visão Computacional pode tomar decisões e, com isso, concluir seu objetivo [Conci et al., 2008].

2.2 Filtros espaciais de mediana e de Prewitt

Como já abordado, a filtragem da imagem por filtros espaciais ou máscaras é uma técnica no domínio espacial, a qual opera diretamente nos pixels da imagem processada, transformando-a em uma de forma mais adequada. A filtragem melhora a imagem por meio da ampliação do seu contraste, acentuação de características, ajuste de foco e eliminação de padrões periódicos ou aleatórios como ruídos adquiridos no momento da sua captura, transmissão ou compressão [Conci et al. 2008].

Filtros espaciais (comentados na Seção 2.1.1.2) são subimagens, pequenas janelas, as quais percorrem toda a imagem a ser processada. A Figura 2.2-a ilustra um filtro de tamanho 3x3 pixels. O centro do filtro é movido pixel a pixel na imagem, Figura 2.2-b, e cada pixel visitado é modificado de acordo com seus pixels vizinhos e com o propósito do filtro.

O filtro de mediana é não-linear (Seção 2.1.1.2) e possui a capacidade de diminuir os ruídos em uma imagem sem modificar as bordas dos objetos nela presentes. O filtro de Prewitt é linear e realça as bordas de objetos presentes na imagem.

No filtro de Mediana, as tonalidades dos pixels vizinhos, e do próprio pixel visitado, são ordenadas de forma crescente ou decrescente e a mediana dessas tonalidades substitui a do pixel posicionado no centro do filtro, ou seja, o pixel visitado. Logo, esse filtro torna-se muito eficiente para eliminar ruídos impulsivos ou do tipo "sal e pimenta". Na Figura 2.6 é possível observar a sua eficácia. Em a) uma imagem de formulário com um ponto de

ruído (no destaque) e em b) o resultado da filtragem por mediana com o desaparecimento do ruído.

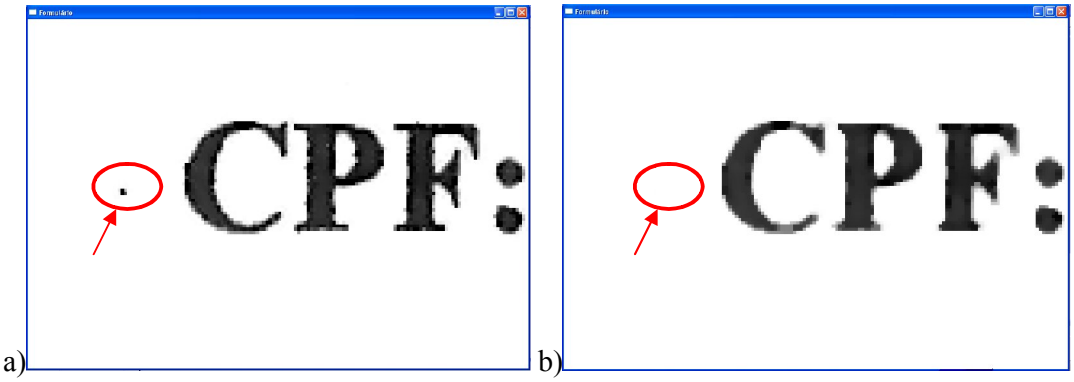


Figura 2.6. Aplicação do filtro de mediana.

O filtro de mediana pouco afeta os contornos dos objetos. Ele pode ser aplicado iterativamente. Quando aplicado na forma de linha e coluna, consecutivamente, ao invés da forma quadrada, apresenta um melhor resultado na preservação dos contornos dos objetos. Como em qualquer outro filtro, seu tamanho também implica no resultado final [Conci et al., 2008].

O filtro de Prewitt é um filtro gradiente. Filtros desse tipo são aplicados para a detecção de contornos de objetos e/ou bordas de regiões em imagens, pois detectam mudanças abruptas de tonalidades. O gradiente de uma função bidimensional, da forma $f(x,y)$, em um ponto (x,y) é definido como o vetor.

$$\nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix} \tag{2.5}$$

de magnitude

$$mag(\nabla f) = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \tag{2.6}$$

O cálculo do gradiente envolve derivadas parciais em relação a *x* e *y*, ou seja, nas direções horizontal e vertical, no caso de sistemas de coordenadas ortogonais. A derivada é interpretada como a taxa de mudança em uma função. Em imagens, essa mudança está relacionada à tonalidade de pixels como, por exemplo, na Figura 2.7, na qual ocorre a mudança de uma região com tonalidade escura para uma outra com tonalidade clara, gerando o perfil de tonalidade exibido na Figura 2.8. Nesse caso, a derivada da imagem na região de fronteira apresenta um valor diferente de zero (Figura 2.9).

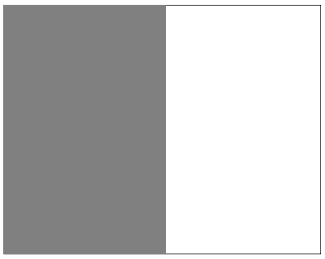


Figura 2.7. Imagem com regiões de tonalidades diferentes.

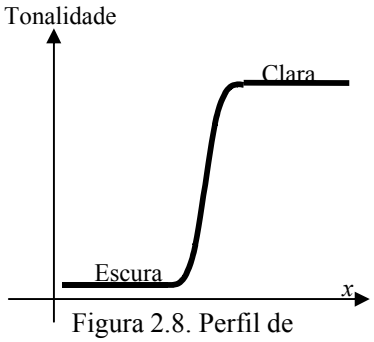


Figura 2.8. Perfil de tonalidade da imagem na Figura 2.7.

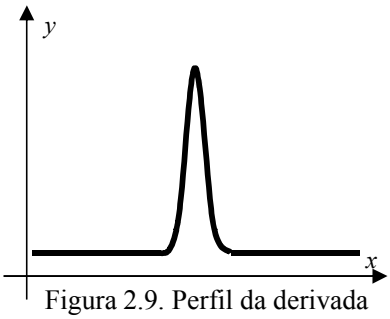


Figura 2.9. Perfil da derivada da imagem na Figura 2.7.

Considerando uma região da imagem de tamanho 3x3, como a ilustrada na Figura 2.2-b, a Equação (2.6) pode ser aproximada para um domínio discreto, dentre outras formas, pela equação:

$$\nabla f \cong \left|(z_7 + z_8 + z_9) - (z_1 + z_2 + z_3)\right| + \left|(z_3 + z_6 + z_9) - (z_1 + z_4 + z_7)\right| \qquad (2.7)$$

em que os coeficientes $c_1, c_2, \ldots, c_9$ (Figura 2.2-a), nesse caso, assumem os valores 1 e -1. É possível implementar a Equação (2.7) pelos filtros de Prewitt, exibidos na Figura 2.10 [Gonzalez & Woods, 1992]. A aplicação deles em uma imagem de formulário é ilustrada na Figura 2.11, onde a) exibe a imagem original e b), os contornos detectados, em que os

verticais e horizontais são detectados com mais eficiência do que os demais devido a natureza dos filtros usados.

–1	0	1
–1	0	1
–1	0	1

–1	–1	–1
0	0	0
1	1	1

Figura 2.10. Filtros de Prewitt vertical e horizontal.

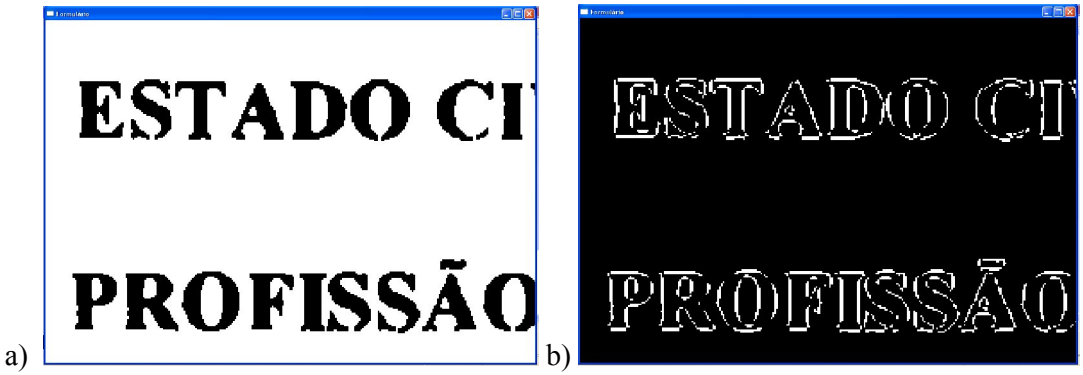


Figura 2.11. Filtragem por filtros de Prewitt.

2.3 Limiarização

A limiarização tem como base as tonalidades de cada ponto da imagem e, por isso, muitas vezes, seu histograma é usado.

O histograma de uma imagem digital é uma ferramenta estatística, a qual fornece o percentual ou a frequência de cada tom presente na imagem. Cada pixel da imagem, em tons de cinza, por exemplo, tem sua tonalidade verificada. E, ao final desse processo, um gráfico cartesiano é construído em que seu eixo horizontal representa os níveis de tons de cinza (0 até 255, por exemplo), e o eixo vertical quantas vezes essas tonalidades apareceram [Conci et al., 2008]. A Figura 2.12 apresenta uma imagem de documento e seu respectivo histograma.

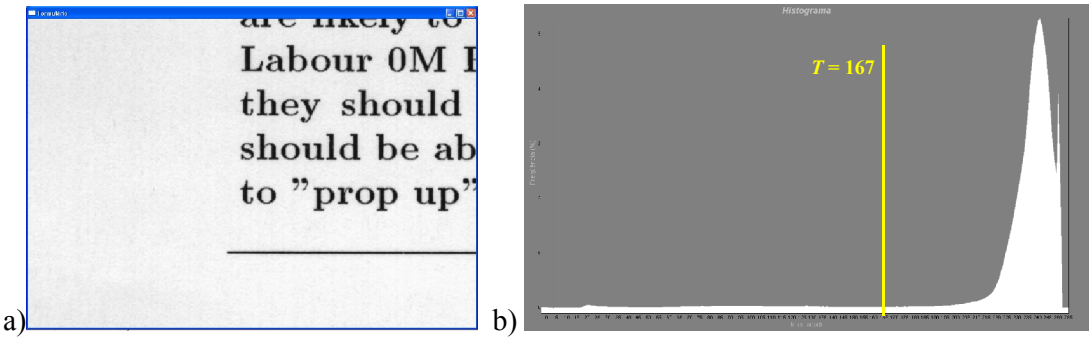


Figura 2.12. Imagem em tons de cinza e seu respectivo histograma.

A segmentação de objetos na imagem pode ser realizada pela análise de seu histograma. Em uma imagem onde objetos de interesse possuem uma tonalidade bem mais escura do que a do restante dela, ou vice versa, obtém-se um histograma composto de dois grupos de tonalidades bem distintas. Eles são separados por um vale entre dois picos, como ilustrado na Figura 2.12-b. Nesse caso, existe um valor *T*, denominado de limiar, entre os picos do histograma que separa de forma eficiente os dois grupos de tonalidades e, consequentemente, o objeto do fundo da imagem. Considerando uma imagem em níveis de cinza descrita por $f(x,y)$, sua limiarização pode ser definida como:

$$g(x,y) = \begin{cases} z_1 & \text{se } f(x,y) > T \\ z_2 & \text{se } f(x,y) \leq T \end{cases} \tag{2.8}$$

onde *g(x,y)* é a imagem limiarizada e z_1 e z_2 são tons de cinza escolhidos. A Figura 2.13 mostra o resultado da limiarização da imagem da Figura 2.12-a. Nessa situação, as tonalidades z_1 e z_2 foram definidas como preta e branca, respectivamente, com T = 167 (Figura 2.12-b).

O histograma mostrado na Figura 2.12 apresenta somente dois picos e, por isso, é chamado de bimodal. Em histogramas desse tipo é necessário apenas um valor de limiar para separar os objetos do fundo da imagem. Mas em outras imagens o histograma obtido é multimodal, ou seja, com vários picos e, dessa forma, são necessários vários limiares para

separar os objetos. Porém, nesse trabalho só será abordado histograma bimodal. Maiores detalhes em [Conci et al., 2008].

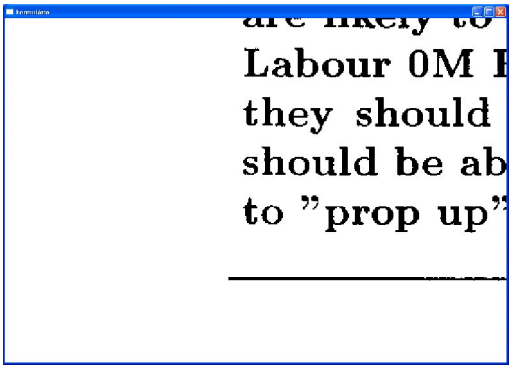


Figura 2.13. Resultado da limiarização.

De maneira mais abrangente, a limiarização pode ser vista como uma função da forma:

$$T = l\big(x, y, p(x, y), f(x, y)\big) \tag{2.9}$$

onde (x, y) são as coordenadas espaciais de um ponto na imagem, $p(x, y)$ representa alguma propriedade local deste ponto e $f(x, y)$ seu nível de cinza. A partir da Equação (2.9), definem-se três tipos de limiar: o global, quando T só depende de $f(x, y)$, o local, quando é função de $f(x, y)$ e (x, y) e o adaptativo, se depender de $p(x, y)$, $f(x, y)$ e (x, y) [Conci et al., 2008]. Para a definição do limiar ótimo global, usando o histograma da imagem, diversos métodos podem ser aplicados, um desses é o de Otsu [1979].

Considerando uma imagem de histograma bimodal, o método de Otsu separa seus pixels em duas classes de tons de cinza, C_0 e C_1. Uma dessas classes representa o objeto (ou os objetos) e a outra o fundo, ou vice-versa. A fronteira entre essas classes é a tonalidade H. Assim, $C_0 = \{0,1,2,3,\ldots,H\}$ e $C_1 = \{H+1, H+2, H+3,\ldots,L\}$, onde L é o número total de tons de cinza. Considerando também ser σ_w^2 a variância dentro de cada classe, σ_b^2 a variância entre as classes e σ_t^2 a variância total, o limiar global T ótimo escolhido é o valor da tonalidade H que maximiza uma das funções:

$$\lambda = \frac{\sigma_b^2}{\sigma_w^2} \tag{2.10}$$

$$\eta = \frac{\sigma_b^2}{\sigma_t^2} \tag{2.11}$$

$$\mathrm{K} = \frac{\sigma_t^2}{\sigma_w^2} \tag{2.12}$$

Como a Função (2.11) é a mais simples entre as três, será detalhada a seguir. Então, na Função (2.11):

$$\sigma_t^2 = \sum_{i=0}^{L-1} (i - ut)^2 P_i \tag{2.13}$$

e

$$\sigma_b^2 = \omega_0 \omega_1 (\mu_0 - \mu_1)^2 \tag{2.14}$$

onde

$$ut = \sum_{i=0}^{L-1} i P_i \tag{2.15}$$

$$\omega_0 = \sum_{i=0}^{H} P_i \tag{2.16}$$

$$\omega_1 = 1 - \omega_0 \tag{2.17}$$

$$\mu_0 = \frac{\mu_t}{\omega_0} \tag{2.18}$$

$$\mu_1 = \frac{\mu t - \mu_t}{1 - \omega_0} \tag{2.19}$$

$$\mu_t = \sum_{i=0}^{H} i P_i \tag{2.20}$$

$$P_i = \frac{n_i}{n} \tag{2.21}$$

Na Equação (2.21) n_i é a quantidade de pixels com o nível de cinza i , n é o total de pixels na imagem e P_i é a probabilidade de ocorrência do nível de cinza i nessa imagem [Monteiro, 2002]. Outro fato é que:

$$n = \sum_{i=0}^{L-1} n_i \tag{2.22}$$

A maximização da Função (2.11), na busca do limiar ótimo é, na verdade, a maximização da variância entre as classes. Mas além desse limiar, o método de Otsu fornece informações de outros aspectos da imagem. Os valores ω_0 e ω_1 são as probabilidades de ocorrência de um elemento das classes C_0 e C_1, respectivamente e μ_0 e μ_1 são níveis médios de cinza em cada classe. A separabilidade de C_0 e C_1, na imagem original, é quantificada pelo valor máximo de η, sendo outra importante informação. Essa medida é invariante para transformações afins da escala de tons de cinza e pertencente ao intervalo $0 \leq \eta \leq 1$ [Monteiro, 2002].

As vantagens do método de Otsu são: (1) o processo é muito simples, (2) somente os momentos cumulativos zero e de primeira ordem do histograma de tons de cinza são utilizados e (3) fornece outras informações de aspectos importantes da imagem. Ele é de natureza não paramétrica e não supervisionada de seleção do limiar [Monteiro, 2002].

2.4 Extração de componentes conectados

A extração de componentes conectados é outra técnica de segmentação da imagem em seus objetos constituintes. Em uma imagem digital, grupos de pixels adjacentes, que possuem alguma similaridade em relação à tonalidade de cinza ou cor, formam objetos ou regiões. Os pixels desses grupos estão ligados [Gonzalez & Woods, 1992] formando componentes conectados. Como exemplo, considere a Figura 2.15-b, onde cada quadrado

representa um pixel. O fundo é formado por pixels brancos e os objetos por pixels de outras cores. Os pixels que formam os objetos são adjacentes e de mesma cor.

Sendo a adjacência uma das exigências à conectividade entre pixels, a definição de vizinhança entre eles é um tópico a ser considerado.

2.4.1 Vizinhanças de um pixel

Devido à amostragem de uma imagem, a vizinhança de um pixel encontra-se geralmente em uma grade regular ou quadrada. As vizinhanças mais utilizadas, pela facilidade de implementação, são: $N_4(p)$, $N_D(p)$ e $N_8(p)$, onde p é um pixel qualquer na imagem. Considerando que as coordenadas de p sejam o par ordenado (x, y), $N_4(p)$ são seus pixels vizinhos horizontais e verticais de coordenadas $(x+1, y)$, $(x-1, y)$, $(x, y+1)$ e $(x, y-1)$, $N_D(p)$ são seus pixels vizinhos diagonais de coordenadas $(x+1, y+1)$, $(x+1, y-1)$, $(x-1, y+1)$ e $(x-1, y-1)$ e $N_8(p)$ é a união de $N_4(p)$ com $N_D(p)$ [Conci et al., 2008]. A Figura 2.14 ilustra cada tipo de vizinhança de um pixel p qualquer, em a), a vizinhança $N_4(p)$, em b), a vizinhança $N_D(p)$ e em c), a vizinhança $N_8(p)$.

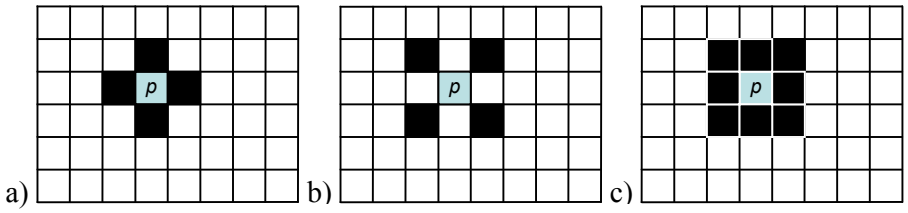


Figura 2.14. Vizinhanças de pixel.

2.4.2 Conectividade entre pixels

A conectividade entre pixels, já comentada, é determinada por dois critérios: a adjacência, relacionada à vizinhança $N_4(p)$ ou $N_8(p)$, por exemplo, e o atributo, como níveis de cinza, cor ou textura.

Nesse contexto, para dois pixels p e q quaisquer, são definidos níveis de conectividade:

- Se $q \in N_4(p)$, e, se p e q possuem valores de atributos que satisfazem a determinados critérios, então eles têm *conectividade – 4*
- Se $q \in N_8(p)$, e, se p e q possuem valores de atributos que satisfazem a determinadas critérios, então eles têm *conectividade – 8.*
- Se $q \in N_4(p)$ ou $q \in N_D(p)$ e $N_4(p) \cap N_4(q) = \varnothing$, e, se p e q possuem valores de atributos que satisfazem a determinado critério, então eles têm *conectividade – m* (mista).

Os critérios aos quais os valores de atributos devem satisfazer, mencionado em cada definição de conectividade acima, podem ser: a igualdade entre eles (mesma tonalidade de cinza, por exemplo) ou atributos com valores dentro de um intervalo tolerável (tonalidade de cinza entre 80 a 120, por exemplo).

2.4.3 Rotulação de componentes conectados

Rotular consiste em atribuir a cada objeto ou região (componentes conectados) da imagem um rótulo único, que, por exemplo, pode ser um número inteiro. Essa rotulação torna-se importante não só para o processo de segmentação, mas também na contagem desses objetos.

Considerando uma imagem binária onde 0 indica o fundo e 1 os objetos, utilizando uma matriz do tamanho da imagem (mesma quantidade de pixels), uma vizinhança $N_4(p)$, onde t e r são vizinhos acima e a esquerda de p, respectivamente, e percorrendo a imagem da esquerda para a direita e de cima para baixo, um algoritmo simples para esta tarefa seria:

se $p = 0$ então vá à próxima posição;

senão:

se $r = 0$ e $t = 0$, então p recebe um novo rótulo;

se $r = 0$ ou $t = 0$, então p recebe o rótulo de *r* ou de *t*;

se $r \neq 0$ e $t \neq 0$ e mesmo rótulo, então p recebe o rótulo de *r* e de *t*;

se $r \neq 0$ e $t \neq 0$ e rótulos diferentes, então p recebe o rótulo de *r* ou de *t* e anota-se que esses rótulos são equivalentes.

Ao final, percorre-se novamente a imagem observando as equivalências anotadas e atribuindo um único rótulo a cada componente conectado [Monteiro, 2002]. A Figura 2.15 ilustra a extração de componentes conectados em uma imagem aumentada em que cada quadrado representa um pixel. Em a), a imagem original e em b), o resultado da extração aplicando o algoritmo descrito, mas usando uma vizinhança $N_8(p)$ para um pixel *p* qualquer. Cada cor representa um rótulo (um objeto).

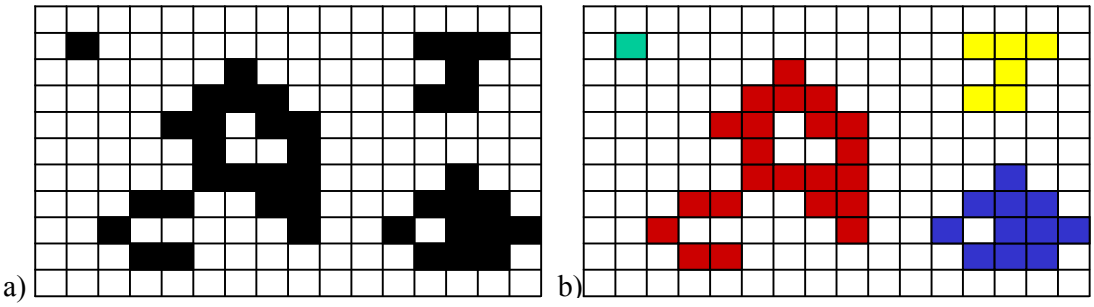


Figura 2.15. Extração de componentes conectados.

2.5 Morfologia matemática em imagens binárias

A morfologia matemática é uma abordagem não linear para as tarefas de processamento de imagens digitais [Calixto, 2005]. Possui muitas finalidades, entre elas estão: o realce, a segmentação, a detecção de bordas, a esqueletização, o afinamento, a análise de formas e a compressão [Facon, 1996]. Na ciência, a palavra morfologia denomina uma área, a qual estuda as formas de animais e plantas. Na gramática da língua portuguesa, morfologia denota a área que estuda a forma das palavras quanto à flexão, à composição, às

variações léxicas e à determinação das categorias gramaticais. Da mesma forma, a palavra morfologia na expressão *morfologia matemática* se justifica, pois ela é uma área que estuda e extrai informações da forma geométrica e da topologia das entidades constituintes de uma imagem, como objetos e regiões.

No escopo desta dissertação, é tratada apenas a morfologia matemática em imagens binárias. Para imagens em tons de cinza há uma morfologia matemática específica e para imagens coloridas uma morfologia matemática ainda se encontra em fase de estudos [Carvalho, 2006].

Como na imagem binária os pixels assumem o valor 1 ou 0 (zero), cada pixel pertence a um objeto da imagem ou ao seu fundo. O valor que representará os objetos está relacionado aos objetivos da tarefa. Sendo D_I o domínio de uma imagem binária, é possível defini-la como um conjunto. Geralmente, D_I é uma região retangular plana em que cada um de seus elementos (pontos) representa um pixel. Considerando uma imagem em um fundo branco define-se:

$$I = \{(x, y) \in D_I : z(x, y) \in \{0,1\}\} \tag{2.23}$$

Naturalmente, em D_I o conjunto de pixels pretos e o conjunto de pixels brancos são excludentes, ou seja, se um pixel não é preto certamente ele é branco e vice versa [Calixto, 2005].

A Figura 2.16 exibe uma imagem binária. Os objetos estão representados pelos pixels pretos e o fundo da imagem pelos pixels brancos. Esses objetos podem ser descritos como conjuntos de pontos no espaço bidimensional Z^2 [Calixto, 2005]. Dessa forma, a linguagem da morfologia matemática passa a ser a teoria dos conjuntos.

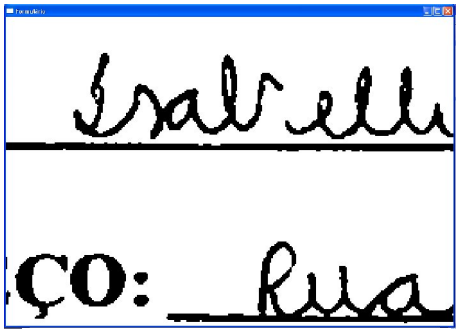


Figura 2.16. Imagem binária.

As informações geométricas e topológicas do conjunto desconhecido (a imagem) são extraídas por meio de transformações entre conjuntos através de outro conjunto, de forma e tamanho definido, chamado de *elemento estruturante* [Facon, 1996]. O elemento estruturante percorre toda a imagem realizando a transformação morfológica por meio das operações descritas a seguir.

2.5.1 Operações morfológicas

A morfologia matemática é caracterizada por operações específicas sobre um conjunto de pixels, a imagem. Neste texto, apenas as operações de erosão, dilatação e abertura serão detalhadas, porém existem outras operações, principalmente resultantes da combinação entre essas.

Como já comentado, a linguagem da morfologia matemática é a Teoria dos Conjuntos, visto que objetos em uma imagem podem ser considerados como conjuntos de pontos no espaço euclidiano Z^2. Assim, algumas definições de conjuntos são relembradas antes das definições das operações morfológicas.

2.5.1.1 Algumas definições da Teoria dos Conjuntos

Considerando $A \subset Z^2$, $x \in Z^2$ e $a \in A$, tal que $x = (x_1, x_2)$ e $a = (a_1, a_2)$ defini-se:

Translação de *A* por *x* como:

$A_x^T = \{y \mid y = a + x, y \in Z^2\}$, onde A_x^T denota a translação de A por *x*;

Reflexão de *A* como:

$A^R = \{y \mid y = -a, y \in Z^2\}$, onde A^R denota a reflexão de A em relação à origem das coordenadas.

2.5.1.2 Erosão

A erosão de certa forma diminui uma imagem. Os efeitos obtidos são a diminuição de partículas, eliminação de grãos pequenos em relação ao elemento estruturante, aumento de buracos e separação de objetos ligados por conexões suficientemente estreitas [Facon, 1996].

Sejam $A \subset Z^2$, $B \subset Z^2$ e $x \in Z^2$, tal que $x = (x_1, x_2)$, a erosão do conjunto A por B é definida como:

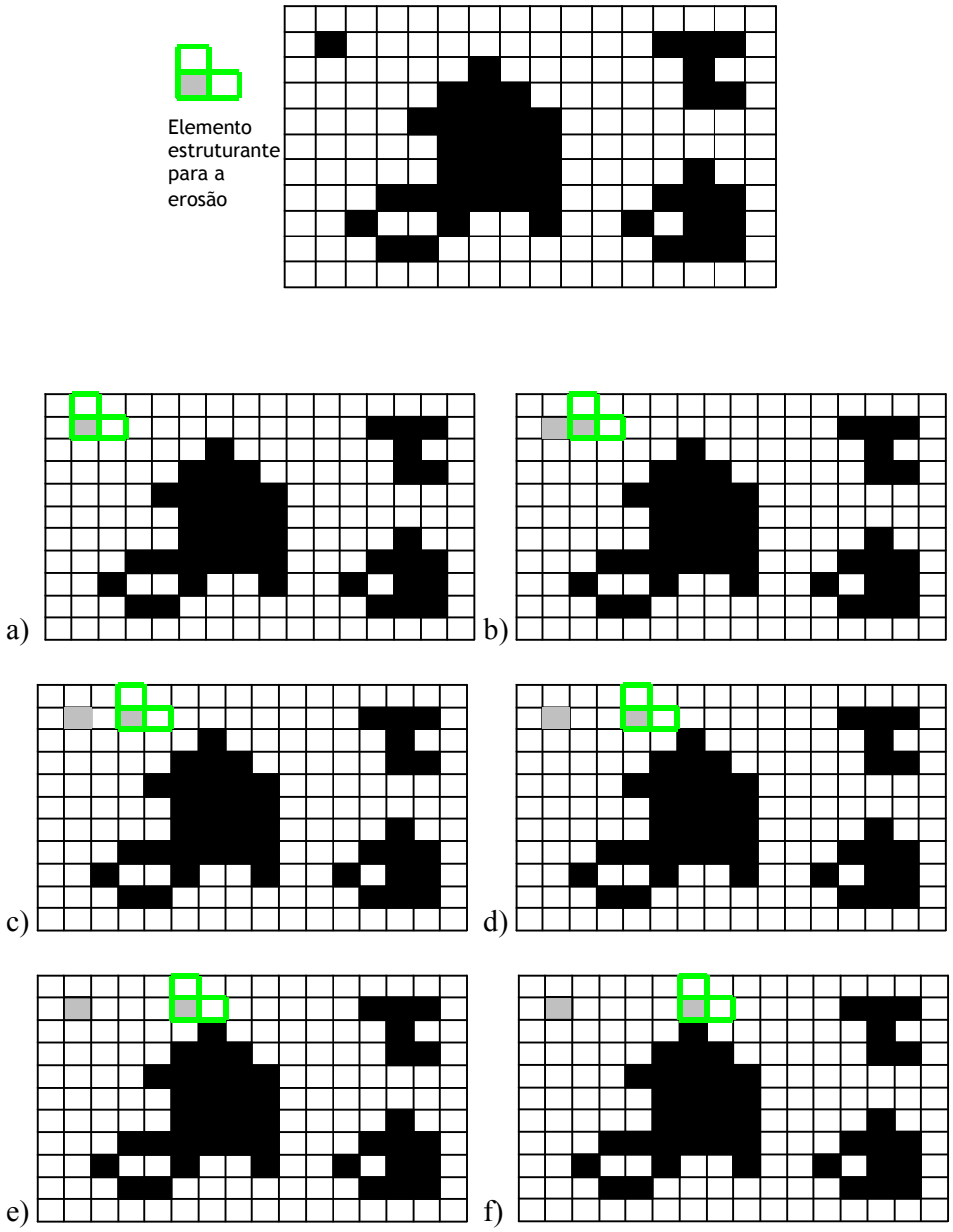
$$A \circ B = \{x \mid B_x^T \subseteq A\} \tag{2.24}$$

onde $A \circ B$ simboliza a erosão de A por B.

O conjunto B, nesse caso, é o elemento estruturante que, na prática, percorre toda a imagem (conjunto A) executando a operação. A Figura 2.17 ilustra a erosão de uma imagem binária aumentada. Cada quadrado na imagem é um pixel, os de cor preta constituem os objetos. O elemento estruturante usado tem o formato de um "L" 2x2 e seu pixel central é o representado em cinza. Esse pixel define a posição onde a operação é efetuada, ou seja, ele se posiciona em um determinado pixel da imagem e a tonalidade (preta ou branca) de cada um de seus vizinhos, na região do "L", é verificada. A operação de erosão é executada, com base na Equação (2.24), e o resultado significa ou o pixel da imagem permanecer preto (ligado) ou ele se tornar branco (desligado). O processo é executado em todos os pixels da imagem, com

exceção dos da borda, pois a definição de erosão não se aplica a eles, pelo fato de a borda de uma imagem ser descontínua.

Na Figura 2.17, primeiramente é exibida a imagem original e o elemento estruturante que irá ser usado na operação. De a) até m) mostra o deslocamento do elemento sobre a imagem, pixel por pixel, n) mostra os pixels "marcados" de cinza por esse, após percorrer toda a imagem, e o) o resultado da erosão. Os pixels "marcados" deixam de fazer parte dos objetos.



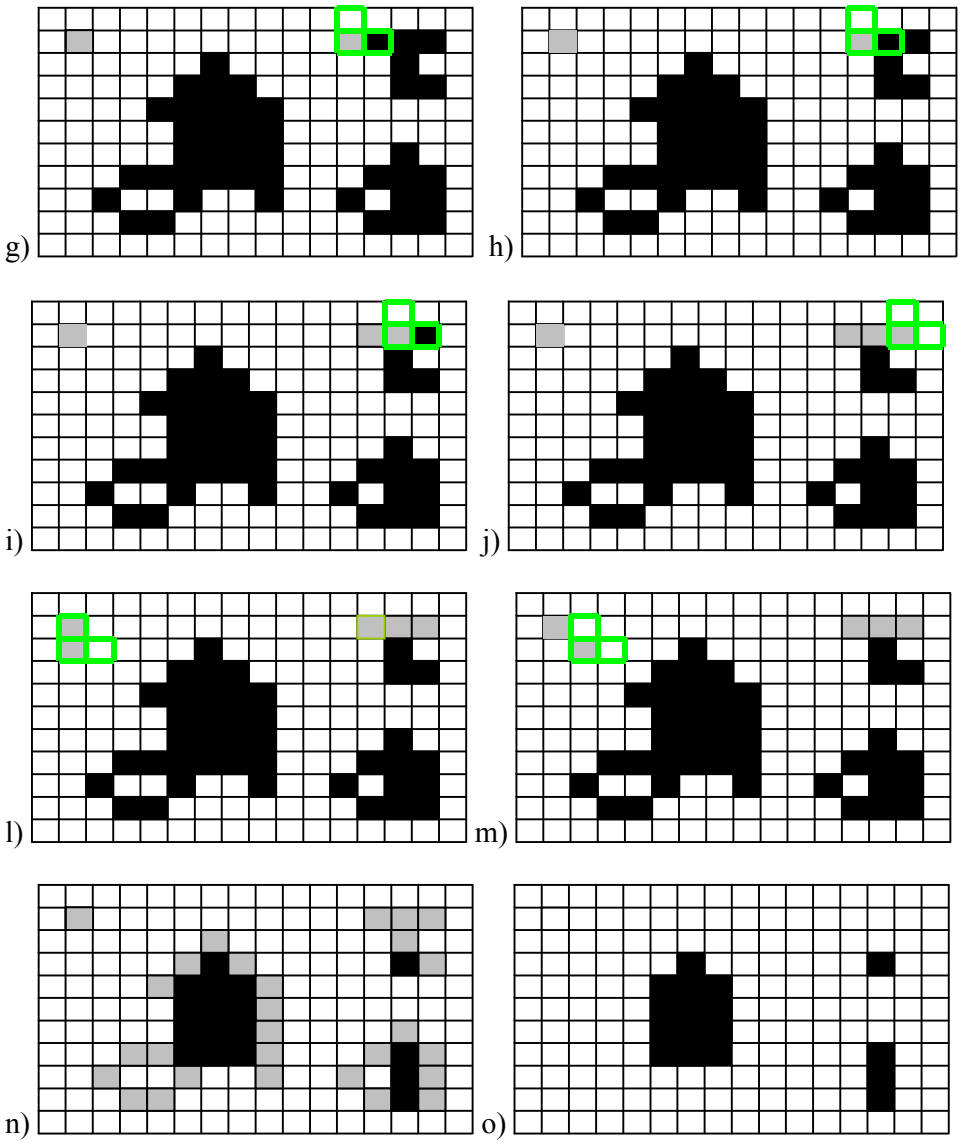


Figura 2.17. Erosão de uma imagem binária.

2.5.1.3 Dilatação

A dilatação, ao contrário da erosão, provoca um aumento na imagem. Os efeitos obtidos são: partículas engordadas, buracos preenchidos e objetos próximos conectados [Facon, 1996].

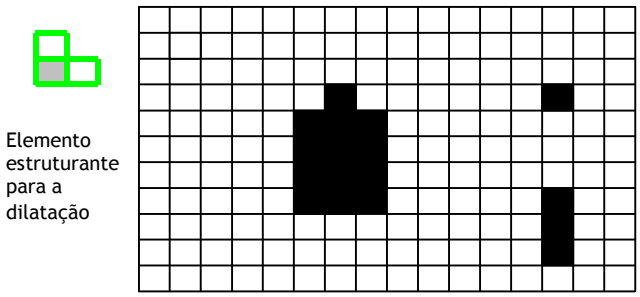
Sejam $A \subset Z^2$, $B \subset Z^2$ e $x \in Z^2$, tal que $x = (x_1, x_2)$, a dilatação do conjunto A por B é definida como:

$$A \oplus B = \left\{ x \mid \left(B^R \right)_x^T \cap A \neq \varnothing \right\} \tag{2.25}$$

onde $A \oplus B$ simboliza a dilatação de A por B. O conjunto B, mais uma vez, é o elemento estruturante que percorre toda a imagem (conjunto A) executando a operação.

A Figura 2.18 ilustra a dilatação da imagem resultante da erosão na Figura 2.17. O elemento estruturante tem o mesmo formato e tamanho, porém agora executa a operação de dilatação com base na Equação (2.25). Seu pixel central se posiciona em um pixel da imagem, o qual tem seus vizinhos, na região do "L", analisados em relação à tonalidade. O resultado, como na erosão, significa: ou o pixel da imagem permanecer preto (ligado) ou se tornar branco (desligado). Todos os pixels são visitados pelo elemento, com exceção da borda, semelhantemente à erosão.

Na Figura 2.18, primeiramente é exibida a imagem resultante da erosão mostrada na Figura 2.17 e o elemento estruturante que será usado na operação. De a) até h) mostra o deslocamento do elemento (já refletido conforme a Equação (2.25)) sobre a imagem, pixel por pixel, i) mostra os pixels "marcados" na cor cinza por esse, após percorrer toda a imagem, e j) o resultado da dilatação. Os pixels "marcados" passam a fazer parte dos objetos.



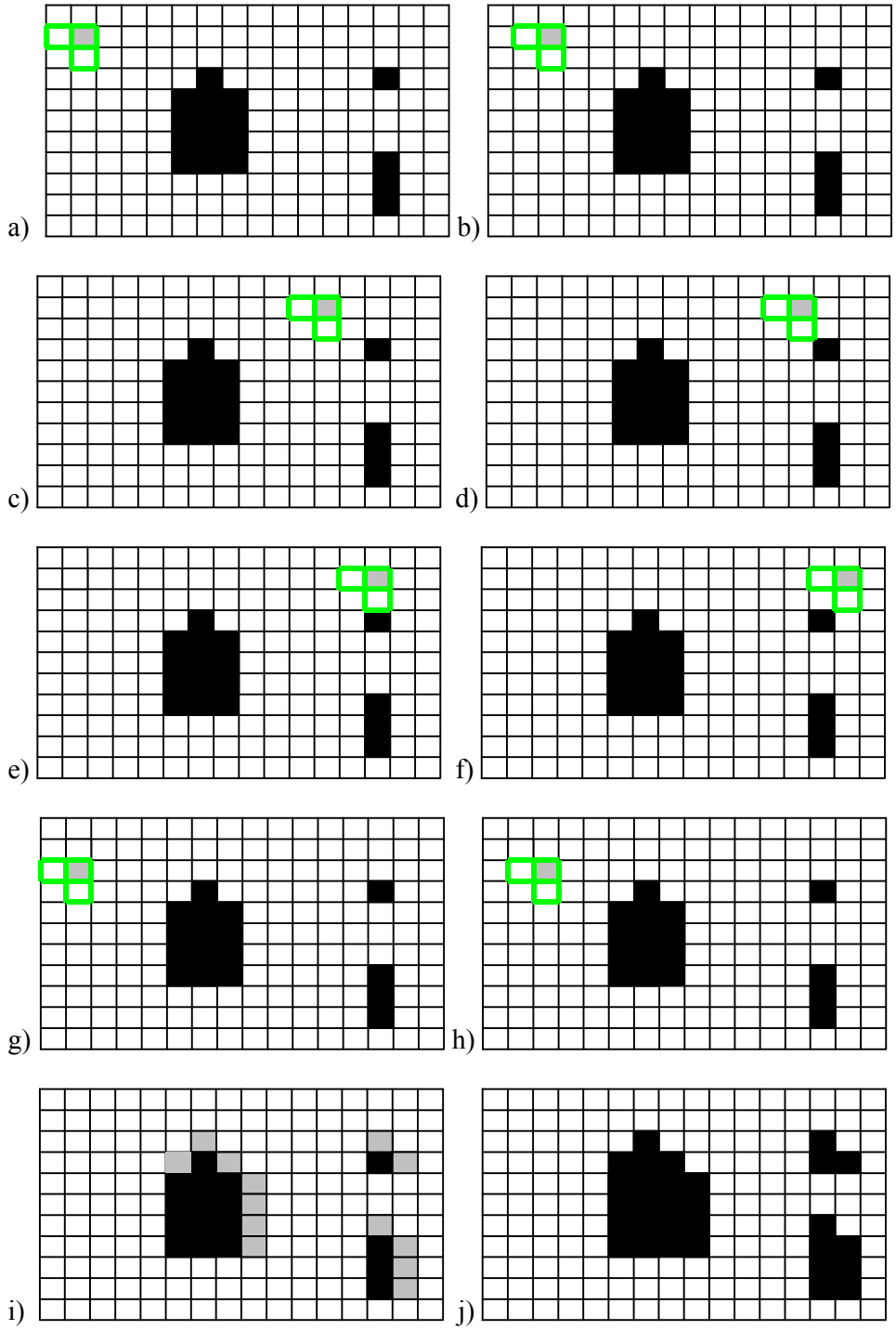


Figura 2.18. Dilatação de uma imagem binária.

2.5.1.4 Abertura

A erosão e a dilatação são operações elementares, as quais dão origem a outras operações da morfologia matemática. Uma delas é a abertura, a qual é uma sequência de

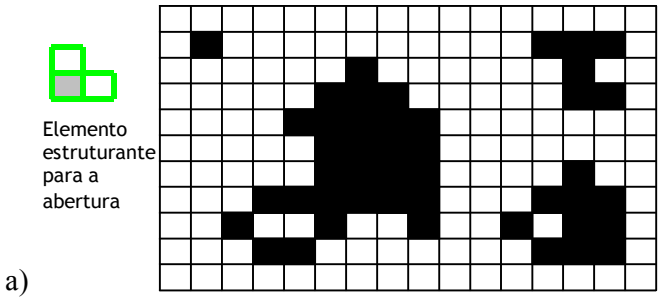
erosão e dilatação da imagem, ou seja, a imagem erodida é em seguida dilatada pelo mesmo elemento estruturante. O efeito, de uma forma em geral, é a suavização dos contornos da imagem [Gonzalez & Woods, 1992]. O tamanho de um objeto não é modificado, porém os grãos indesejáveis são eliminados resultando em uma imagem com menor grau de detalhes. Também, ocorre a separação de objetos de estreita ligação [Facon, 1996].

Sejam os conjuntos $A \subset Z^2$ e $B \subset Z^2$, a abertura do conjunto A por B é definida como:

$$A * B = \left(A \circ B \right) \oplus B \tag{2.26}$$

onde $A * B$ simboliza a abertura de A por B.

Na definição acima, o conjunto B é o elemento estruturante usado na execução da abertura que percorre toda a imagem (conjunto A) realizando a operação. A Figura 2.19 ilustra a abertura com os resultados dos exemplos anteriores de erosão e dilatação. Em a) a imagem original e o elemento estruturante - em forma de "L" e tamanho 2x2 - que será usado. Após o elemento percorrer toda a imagem, pré-definido para operar a erosão, alguns pixels são "marcados" (b) e eliminados (c). A imagem erodida é novamente percorrida pelo elemento, agora definido para operar a dilatação, e então, pixels são "marcados" (d) e passam a fazer parte dos objetos na imagem (e). Observando o resultado da abertura (e) e comparando-o com a imagem original (a) nota-se que pontos isolados são eliminados e os objetos ganham contornos mais suaves.



a)

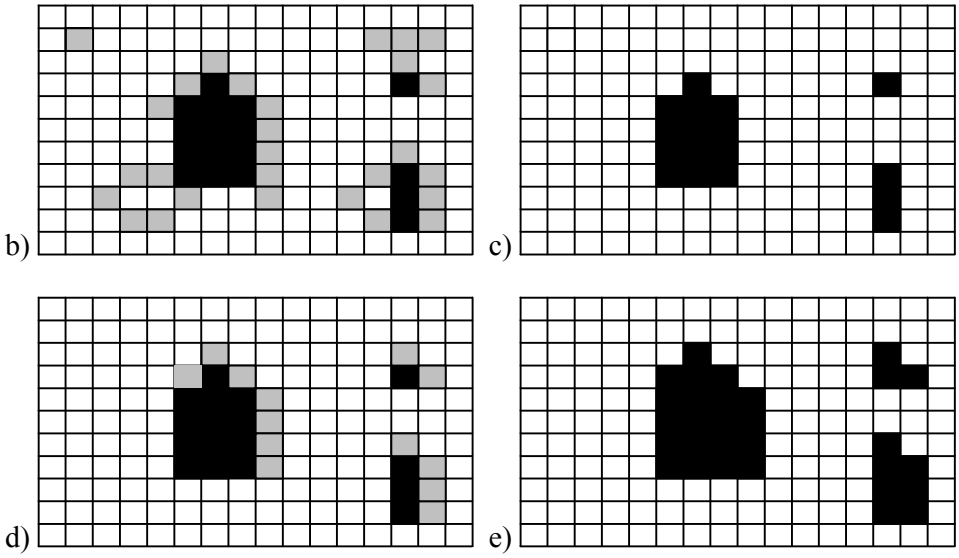


Figura 2.19. Abertura de uma imagem binária.

Diversas outras operações resultantes dessas elementares são muito importantes para as várias aplicações. Uma ótima forma de conhecer, manipular e assimilar essas outras, é a utilização do pequeno programa disponível em: http://www.ic.uff.br/~aconci/Morfologia, o qual é uma ferramenta em *JavaScript*, orientada ao objeto, para o estudo da Morfologia Matemática binária, entre outros tópicos. Essa ferramenta é descrita no trabalho de Nuñez e Conci [2007].

2.6 Descoberta de Conhecimento em Bancos de Dados e Mineração de Dados

Nas últimas décadas, grande quantidade de dados vem sendo adquirida ou produzida por empresas, centros de pesquisas, etc. Isso, cria a necessidade e a oportunidade de extração do conhecimento, precioso e necessário para tomadas de decisões estratégicas, guardado por esses dados. Porém, a extração do conhecimento, baseada na observação dos dados, extrapola a capacidade humana. Do desenvolvimento de técnicas e ferramentas computacionais, capazes de extrair informações úteis dessa grande quantidade de dados,

nasceu a área conhecida como DCBD e a Mineração de Dados [Fayyad et al., 1996 e Piatetsky-Shapiro, 1990].

Pelo fato de o conhecimento estar implícito nos bancos de dados, tem-se a DCBD como um processo não trivial de extração do conhecimento inserido nos dados (dados na forma bruta), usando métodos e técnicas desenvolvidas especialmente para automatizar parcialmente a tarefa de análise dos dados nesses bancos [Silva, 2003].

A DCBD pode ser dividida em três fases fundamentais: preparação de dados, mineração de dados e pós-processamento, onde a mineração de dados é considerada a mais importante, e, por isso, denomina todo o processo em alguns trabalhos da literatura. Porém, a preparação dos dados é muito importante para o sucesso do processo, tal como o pós-processamento. A Figura 2.20 ilustra essas fases e suas etapas.

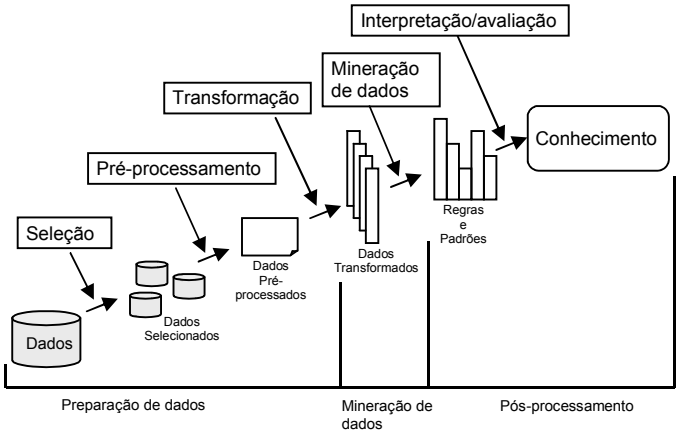


Figura 2.20. Fases e etapas da DCBD.

2.6.1 Preparação de dados

O foco da fase de preparação de dados é a obtenção de melhores resultados na fase subsequente, a mineração de dados. Dessa forma, o sucesso de todo o processo de DCBD tem início em atividades como o aumento da qualidade e disposição dos dados, no intuito de facilitar a extração de conhecimento. As etapas da preparação de dados são: seleção, pré-processamento e transformação dos dados [Silva, 2003].

Nem todos os dados de uma base são importantes para se alcançar os objetivos traçados pelo usuário, então é interessante que sejam retirados. Isso é realizado, quando necessário, na etapa de seleção dos dados. Em outras vezes, ocorre que todos os dados de uma base são importantes para se alcançar o objetivo, porém se apresentam em uma quantidade muito grande, tornando impossível uma análise. Nesse caso, é necessário reduzir esses dados, sem que haja um prejuízo da integridade dos originais. Essa tarefa é realizada também na etapa de seleção dos dados [Silva, 2003].

A etapa de pré-processamento tem o objetivo de identificar erros em dados selecionados. Esses erros podem ser: a duplicação de registro, ausência de valores em atributos obrigatórios e atributos preenchidos com valores incorretos [Silva, 2003]. Após a identificação dos erros, são executadas operações como a correção ou a eliminação dos dados por alguma metodologia. Dentre as metodologias de correção está o preenchimento de um atributo do dado de valor ausente. No caso de atributos categóricos o valor escolhido para esse preenchimento pode ser a moda, e no caso de valores contínuos pode ser a média ou a mediana, dos valores desse atributo em todos os dados da base [Romão, 2002].

A etapa de transformação modifica os dados conforme o objetivo da aplicação e o formato exigido pela técnica de mineração de dados usada na fase de mineração. Por exemplo, a transformação de atributos contínuos em categóricos, no caso em que o algoritmo de mineração não aceite o primeiro formato, ou para que o conhecimento descoberto seja compreendido mais facilmente [Romão, 2002].

2.6.2 Mineração de Dados

Após os dados serem preparados, passa-se à fase da Mineração de Dados (Figura 2.20). Uma técnica é escolhida para executar uma determinada tarefa e um modelo é

construído com os padrões e regras identificadas nos dados da base analisada. As tarefas, as técnicas, os tipos de modelos, e tipos de aprendizagens são conceitos comentados a seguir.

A Tabela 2.1 exibe uma relação de tarefas encontradas na fase de mineração de dados, as técnicas utilizadas na realização dessas tarefas, os modelos gerados e os tipos de aprendizagens.

Tabela 2.1. Tarefas e técnicas da Mineração de Dados

Técnica	Tarefa	Tipo do modelo gerado	Tipo de aprendizagem	Forma de representação do conhecimento
Árvore de decisão	Classificação	Preditivo	Supervisionado	Árvore de decisão
Regras de classificação	Classificação	Preditivo	Supervisionado	Regras de classificação
Identificação de regras associativas	Associação	Descritivos e preditivo	Não-supervisionado	Regras associativas
Rede Neural	Classificação	Preditivo	Supervisionado	Caixa preta
Algoritmos de Clusterização (K-Média e Agrupamento demográfico)	Clusterização (agrupamento)	Descritivo	Não-supervisionado	Centróide

A fase de mineração de dados pode ser dividida em dois estágios: a escolha da técnica e a mineração dos dados (propriamente dita) com a técnica escolhida. A escolha da técnica depende, fundamentalmente, da tarefa a ser realizada e do objetivo final do processo de DCBD. Porém, outros fatores são considerados, como o modelo a ser gerado, o tipo de aprendizagem e o formato da representação do conhecimento a ser descoberto [Silva, 2003].

A aplicação da técnica escolhida sobre os dados da base, constrói um modelo (como já mencionado) o qual pode ser preditivo, quando é construído para prever a classe à que pertence determinado dado (registro, objeto ou instância) novo, e, por isso, desconhecido, ou descritivo, quando descreve os dados da base de modo que uma pessoa possa compreender.

No caso da tarefa de classificação, por exemplo, o modelo funciona como um "mecanismo" onde o dado novo entra e retorna classificado (Figura 2.21).

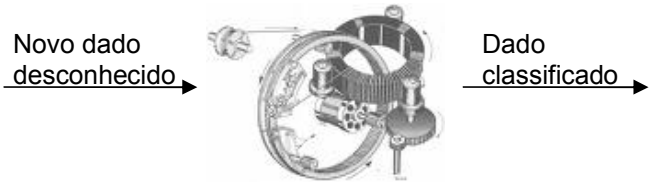


Figura 2.21. Modelo de classificação.

A etapa de construção do modelo é geralmente chamada de aprendizagem. Essa aprendizagem pode ser supervisionada ou não-supervisionada. Quando supervisionada, parte da base de dados é separada para treinamento. Essa parte possui seus dados já classificados por um especialista da área. O algoritmo de mineração recebe os dados classificados e constrói um modelo para representá-los. Comumente, a aprendizagem supervisionada é usada em técnicas preditivas. Quando não-supervisionada, não existe dado com classe ou grupo predefinido, ou seja, não há intervenção humana. Quase sempre, a aprendizagem não-supervisionada é usada em técnicas descritivas como algoritmos de clusterização, que agrupam os elementos de acordo com as suas similaridades. Em muitos casos, o número de classes ou grupos é desconhecido e cabe ao algoritmo descobri-lo [Silva, 2003].

Na Tabela 2.1, são citadas as tarefas de classificação, associação e clusterização, que são algumas dentre várias outras na área de Mineração de Dados. A classificação consiste basicamente em identificar a que classe pertence um determinado dado novo mediante suas características. Essas classes são pré-definidas e discretas. A associação pesquisa padrões de relacionamento entre os dados de uma base que aparecem com determinada frequência. Esses padrões, são descritos por regras de associação do tipo $X \Rightarrow Y$, onde X é chamado de antecedente e Y, de consequente. Um exemplo é a regra $\{\text{salaminho}\} \Rightarrow \{\text{cerveja}\}$ minerada de uma base de dados de transações de compra de um supermercado. Essa regra indica que quem

compra salaminho tende a comprar cerveja. Geralmente, são usadas medidas para avaliar sua qualidade [Gonçalves, 2005]. E, por último, a tarefa de clusterização (agrupamento) consiste em agrupar, em classes, os elementos de uma base de dados a partir da similaridade entre eles, ou seja, elementos de uma mesma classe são mais similares do que os de classes diferentes. Por possuir um treinamento não-supervisionado não há elementos pré-classificados na base de dados e, em várias circunstâncias, o número de classes também não é pré-estabelecido, cabendo ao algoritmo determinar esse número através da análise dos dados, como já comentado [Silva, 2003].

2.6.3 Pós-processamento

A fase de mineração dos dados gera uma quantidade grande de regras ou padrões descobertos, dos quais poucos são de fato interessantes ou úteis ao objetivo pretendido. Desse modo, a fase de pós-processamento do processo de DCBD interpreta e compreende os padrões e regras mineradas, avaliando o quanto são úteis, novos e válidos para dados desconhecidos, com alguma taxa de certeza, e também se são de fácil compreensão por uma pessoa ou se confirmam uma hipótese levantada pelo usuário [Han & Kamber, 2001].

Dentre as tarefas comentadas no texto, a classificação foi a escolhida para ser aplicada neste trabalho e por isso ela será mais detalhada a seguir.

2.6.4 Classificação

Em linhas gerais, a tarefa de classificação, por gerar um modelo preditivo com aprendizagem supervisionada, utiliza dados do passado, previamente classificados, para prever a classe dos dados, dos quais ainda não se conhece a classe, a partir de suas características. Um exemplo desse tipo de tarefa seria antever a falência de uma organização

de negócios mediante a análise de características, tais como sua flexibilidade financeira, sua credibilidade, sua competitividade, seu risco de administração, etc [Kim & Han, 2003] .

O conjunto de dados (o banco de dados) disponível é separado em dois subconjuntos disjuntos, um para treinamento e outro para teste. Através de um algoritmo, executado sobre o subconjunto de treinamento, um modelo de classificação é criado e utilizado para classificar os dados do subconjunto de teste e os demais dados (Figura 2.22).

O modelo de classificação possui as classes já pré-definidas e pode ser formado por regras do tipo:

SE{antecedente}ENTÃO{consequente}

onde o antecedente é formado por uma combinação, geralmente uma conjunção, de condições de valores dos atributos previsores, e o consequente é formado por um valor do atributo meta, dentre os previstos [Romão, 2002]. Os atributos previsores descrevem características e o atributo meta descreve uma classe correspondente às características. Então, a forma generalizada de uma regra de classificação é do tipo:

$$\text{SE } \{(\textit{atributo}_1 \text{ satisfaz } \textit{condição}_1) \text{ e } (\textit{atributo}_2 \text{ satisfaz } \textit{condição}_2) \text{ e ... e } (\textit{atributo}_n \text{ satisfaz } \textit{condição}_n)\} \text{ ENTÃO } \{\textit{classe} = c_k\}$$

onde n é o número de condições e k a quantidade de possíveis classes.

Vale lembrar que o uso de regras de classificação não é a única forma de construção de um modelo de classificação, mas é a utilizada neste trabalho.

Para exemplificar, na tarefa de classificação de uma palavra como impressa ou manuscrita, considerando como características hipotéticas das palavras a largura e a altura, poderiam surgir regras de classificação, mineradas do conjunto de treinamento, do tipo:

- SE {(*Largura da Palavra* > 100 pixels) e (*Altura da Palavra* > 55 pixels) } ENTÃO { *classe* = *manuscrito* }
- SE {(*Largura da Palavra* ≤ 100 pixels) e (50 pixels < *Altura da Palavra* ≤ 55 pixels) } ENTÃO { *classe* = *impresso* }

Se os atributos (características) de um novo dado (registro, instância, objeto etc) satisfazem à conjunção de condições do antecedente de uma regra, então esse novo dado é classificado como pertencente à classe indicada pelo atributo meta da regra. Logo, considerando a ilustração anterior, se a largura de uma palavra for igual a 70 pixels e a altura for igual a 43 pixels, por exemplo, então essas características satisfazem à segunda regra e, assim, a palavra será classificada como impressa.

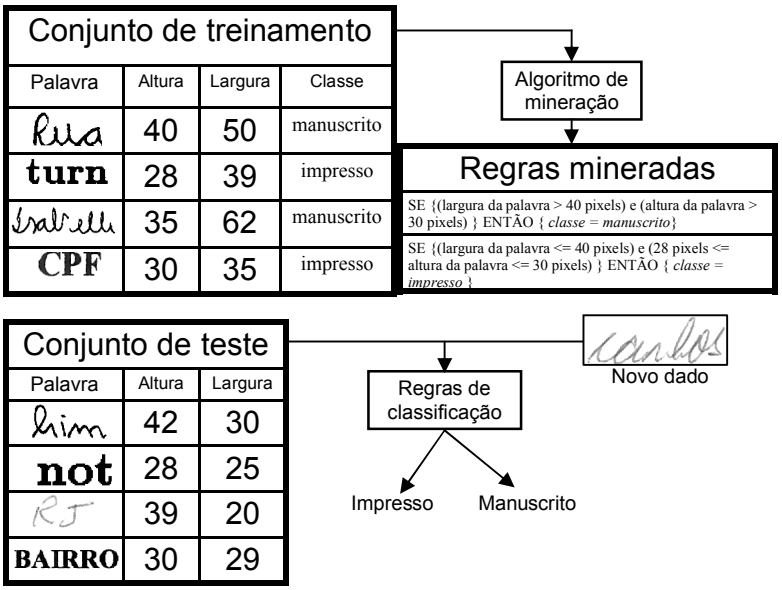


Figura 2.22. Construção e aplicação do modelo de classificação.

2.6.5 Avaliadores da classificação

Para avaliar o resultado da classificação realizada pelo modelo construído na fase de treinamento, existem alguns avaliadores frequentemente usados. Nesta dissertação foram eles: a acurácia, a precisão, verdadeiro positivo, falso positivo, verdadeiro negativo, falso negativo, a sensibilidade, a especificidade e o desvio padrão. Considerando ser c_k uma classe qualquer, esses avaliadores são definidos a seguir.

Acurácia

A *acurácia* é o grau de veracidade de uma quantidade calculada ou medida [Wikipedia]. A equação usada para avaliar a *acurácia* do sistema, desenvolvido neste trabalho, em relação a sua classificação é:

$$\text{Acurácia} = \frac{\text{nº de dados da classe } c_k \text{ classificados corretamente}}{\text{nº de dados da classe } c_k \text{ no banco de dados}} \tag{2.27}$$

[Zheng et al., 2004].

Precisão

A *precisão* é o grau de reprodutibilidade de uma quantidade calculada ou medida [Wikipedia]. A equação usada para avaliar a *precisão* do sistema, desenvolvido neste trabalho, em relação a sua classificação é:

$$\text{Precisão} = \frac{\text{nº de dados da classe } c_k \text{ classificados corretamente}}{\text{nº de dados classificados como da classe } c_k} \tag{2.28}$$

[Zheng et al., 2004].

Verdadeiro positivo - VP

Verdadeiro positivo é o total de dados classificados como sendo da classe c_k, os quais realmente são de c_k [Castanho et al., 2004].

Falso positivo - FP

Falso positivo é o total de dados classificados como sendo da classe c_k, os quais não são de c_k [Castanho et al., 2004].

Verdadeiro negativo - VN

Verdadeiro negativo é o total de dados classificados como não sendo da classe c_k, os quais realmente não são de c_k [Castanho et al., 2004].

Falso negativo - FN

Falso negativo é o total de dados classificados como não sendo da classe c_k, os quais são de c_k [Castanho et al., 2004].

Sensibilidade

A *sensibilidade* é a probabilidade de um dado da classe c_k ser classificado corretamente pelo classificador usado [Castanho et al., 2004], é obtida pela equação:

$$\text{Sensibilidade} = \frac{VP}{VP + FN} \tag{2.29}$$

Especificidade

A *especificidade* é a probabilidade de um dado não pertencente à classe c_k ser classificado corretamente pelo classificador usado [Castanho et al., 2004], é obtida pela equação:

$$\text{Especificidade} = \frac{VN}{VN + FP} \tag{2.30}$$

Desvio padrão

A *variância* é uma medida quantitativa utilizada para representar a dispersão dos dados ao redor da média em um conjunto. O *desvio padrão* é a raiz quadrada da *variância* e é obtido pela equação:

$$\text{dp} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - M)^2}{n}} \tag{2.31}$$

onde n e M são a quantidade e a média, respectivamente, dos elementos x_i do conjunto analisado [Nunes, 2006].

2.6.6 Validação Cruzada

Dentre os métodos de validação do resultado do teste da classificação, a validação cruzada (*Cross Validation*) foi o aplicado no presente trabalho. Esse método é recomendável quando não se possui grandes quantidades de dados no banco.

O método consiste em dividir o conjunto de dados disponível em *k* subconjuntos ou pastas (daí alguns trabalhos denominarem esse método de *k-fold Cross Validation*). Então, $k - 1$ subconjuntos são usados para o treinamento do sistema, enquanto o subconjunto que restou é usado como conjunto de teste. Esse processo é repetido *k-vezes* até que todos os subconjuntos tenham sido usados como conjunto de teste pelo menos uma, e somente uma, vez.

Geralmente, o resultado final nesse método é a média dos resultados dos avaliadores em cada subconjunto testado.

Com esse método de validação dos resultados dos testes, o capítulo é finalizado. Nele, todos os tópicos teóricos aplicados neste trabalho são abordados. No próximo capítulo, a metodologia proposta é detalhada.

3. TRABALHOS ANTERIORES

Nos últimos anos, foram propostas várias metodologias para a tarefa de distinguir textos (caracteres) impressos e textos (caracteres) manuscritos em uma imagem de documento. Porém, a maioria tem em comum as seguintes etapas: pré-processamento, segmentação do texto, extração de características e a classificação em impresso ou manuscrito. A Figura 3.1 ilustra essas etapas.

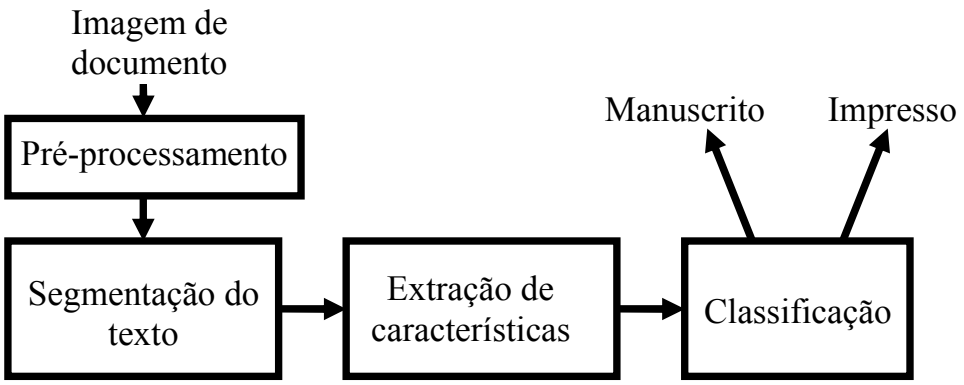


Figura 3.1. Etapas de metodologias de distinção de texto impresso e manuscrito.

A fase de pré-processamento tem por objetivo preparar a imagem do documento para a fase de segmentação do texto. Alguns procedimentos comumente realizados nessa fase são: localização de regiões de texto, detecção de orientação do texto (retrato ou paisagem), correção do ângulo de inclinação da página, binarização da imagem e eliminação de ruídos.

Após o pré-processamento, as regiões de textos são divididas em áreas menores, as quais podem delimitar cada caractere ou cada palavra ou cada linha ou, até mesmo, zonas de texto. A classificação ocorre nessas áreas, isto é, se elas delimitam palavras, cada uma é classificada como contendo uma palavra impressa ou manuscrita.

A etapa de extração de características consiste em extrair, de cada área, aquelas que são significativas na distinção de impressos e manuscritos. Essas características,

geralmente, descrevem a forma e/ou os aspectos da textura da área. Outras, descrevem a forma do conteúdo interior.

A classificação é a fase que decide se uma determinada área contém impresso ou manuscrito de acordo com as características extraídas. Existem várias técnicas para esse propósito, entre elas estão: redes neurais, classificadores estatísticos, discriminantes lineares e árvore de decisão.

Na década de 90, o trabalho de Imade et al. [1993] descreve uma metodologia para segmentar uma imagem de documento em regiões e classificá-las como contendo um dos seguintes elementos: caractere (Kanji ou Kana) impresso, caractere (Kanji ou Kana) manuscrito, fotografia ou imagem pintada. Kanji e Kana são tipos de caracteres da escrita japonesa. Depois da aplicação de uma técnica de limiarização, a imagem binarizada é dividida em blocos de 8x8 pixels. Em seguida, para cada bloco é verificada a existência de algum pixel preto. Se existir pelo menos um, o bloco é substituído por um "elemento preto". Caso contrário, ele é substituído por um "elemento branco". Elementos isolados, considerados ruídos, são eliminados por filtragem espacial. Todos os "elementos pretos" conectados são cercados por retângulos, os quais são considerados as áreas segmentadas. De dentro desses retângulos são selecionados, aleatoriamente, quadrados pequenos de onde as características são calculadas e a classificação executada. As características são: o histograma de direções do vetor gradiente, calculado em cada pixel, e o histograma de níveis de luminância. Com base nelas, a classificação é executada por uma rede neural. Uma área segmentada é classificada em determinada classe de acordo com a classificação da maioria dos quadrados pequenos nela selecionados. Segundo os autores, os melhores resultados de classificação foram alcançados quando o tamanho dos quadrados foi fixado em 32x32 pixels. Nesse caso, obtiveram 84,5% de acertos para caracteres impressos, 77,7% de acertos para caracteres manuscritos, 98,8% de acertos para fotografias e 67,7% de acerto para pinturas.

Franke e Oberlander [1993] desenvolveram uma metodologia para classificar os campos de dados de uma imagem de cheque bancário em impressos ou manuscritos. Nesse caso, o leiaute do cheque é conhecido. Dos campos são extraídos componentes conectados, os quais são cercados por retângulos envoltórios (o menor retângulo possível limitante). Quatro características são extraídas de cada campo: o histograma das larguras (contendo a distribuição das larguras de seus retângulos envoltórios), o histograma das alturas (contendo a distribuição das alturas de seus retângulos envoltórios), o histograma das distâncias (com as distâncias de cada dois retângulos envoltórios consecutivos horizontalmente) e o histograma das distâncias dos centros (com as distâncias dos centros de cada dois retângulos envoltórios consecutivos horizontalmente). Cada característica é enviada a um classificador polinomial linear, com as devidas adaptações. De cada um, é gerado um vetor de decisão bidimensional, onde a primeira coordenada é uma estimativa probabilística do conteúdo do campo de dado ser um manuscrito e a segunda de ser um impresso (a soma das duas é igual a 1). Como a segunda coordenada é linearmente dependente da primeira, somente a primeira de cada vetor gerado é usada para gerar um novo vetor de característica com quatro coordenadas. Esse novo vetor formado é usado como entrada em outro classificador polinomial, o qual finalmente faz a classificação do campo em impresso ou manuscrito. Para o desenvolvimento e teste do sistema foram usados 800 cheques bancários. Cada cheque contendo 13 campos, nem sempre todos preenchidos. Segundo os autores, a taxa de erro na classificação foi de 1,04% no conjunto de treinamento e 1,18% no conjunto de teste. Não foram descritas comparações com outros trabalhos.

Violante et al. [1995] apresentam outro trabalho para essa tarefa. O sistema faz a distinção entre impressos e manuscritos em textos de endereços de cartas. São processadas imagens com resolução de 512x512 e com 256 tons de cinza. Devido às condições de iluminação não adequadas, no momento da captura das imagens das cartas, elas tendem a ser mais brilhantes na parte esquerda. Para corrigir isso, a intensidade de cada pixel é

multiplicada por (0.5 +0.5x/512), onde *x* é a coordenada horizontal do pixel. As regiões de endereços são manualmente delineadas por retângulos envoltórios. Com o objetivo de separar o texto do fundo da imagem, é usada uma técnica de limiarização adaptativa. As características extraídas de cada retângulo envoltório são: Contagem de Componentes Conectados, Retidão da Aresta, Perfil Horizontal, Altura, Largura e Área. O texto é dividido em componentes conectados, onde os com mais de nove pixels são contados. Textos impressos nessas imagens, quando divididos, formam palavras, enquanto os manuscritos formam letras. Esse fato gera um número maior de componentes conectados em imagens com manuscritos do que com impressos. A aresta é definida como pares de pixels em que cada par é formado por um pixel do texto e outro do plano de fundo do retângulo envoltório que o contém. A quantidade de pares de pixels alinhados é determinada e considerada como a característica Retidão da Aresta. A projeção horizontal dos pixels do texto no retângulo envoltório é obtida e a diferença absoluta de linhas adjacentes dessa é calculada e considerada como a característica Perfil Horizontal, na qual as diferenças em impressos são maiores. A classificação é feita por uma rede neural treinada, baseada nas características extraídas. Foram utilizadas 400 imagens, metade delas para treinamento e outra para testes. Segundo os autores, a acurácia de classificação alcançada foi da ordem de 95%. A base de dados usada não foi especificada.

Kuhnke et al. [1995] descrevem um trabalho que classifica uma imagem de caractere em impresso ou manuscrito. Primeiramente, o caractere da imagem é cercado por um retângulo envoltório, pois ele não a preenche totalmente, e seu contorno é detectado. De cada retângulo envoltório são extraídas algumas características. Uma delas é a retidão das linhas verticais e horizontais dos contornos dos caracteres. Inicialmente, essas são detectadas com a aplicação da Transformada de Hough. Para avaliar a retidão de uma linha é calculada a variância das coordenadas de seus pixels. Outras características são: a simetria relativa a pontos diferentes e a simetria dos loops internos, ambas em relação ao centro de gravidade do

caractere. Essas são invariantes em relação ao tamanho e a rotação dos caracteres. A classificação é realizada por uma rede neural com base nas características extraídas. Para treinamento e teste do sistema, foi formada uma base com 4700 imagens de caracteres, 3632 para o treinamento e 1068 para o teste, com distribuição uniforme em relação às classes. Parte dos caracteres manuscritos foi obtida da Nist Special Database 3 e os caracteres impressos foram obtidos de formulários escaneados. De acordo com os autores, os resultados foram de 96,8% de acertos na classificação no conjunto de treinamento e 78,5% no conjunto de teste. Não há comparações com outros trabalhos.

No trabalho de Srihari et al. [1996] é desenvolvido um sistema integrado de tempo real para "ler", em formulários fiscais, nomes e endereços, tanto impressos quanto manuscritos. A entrada do sistema é uma imagem em formato TIFF e a saída é a forma ASCII do nome e endereço escritos no formulário. Primeiramente, é feita a extração de componentes conectados da imagem. Após a análise desses componentes, blocos de endereços são extraídos. Para determinar se um bloco de endereço é de texto impresso ou de texto manuscrito, algumas características são extraídas dos seus componentes conectados, são elas: desvio padrão das alturas e larguras, densidade média, relação entre a altura e a largura e quantidade de alturas e larguras pouco frequentes. Para a classificação é empregado um discriminante linear Fisher. Durante o treinamento, o hiperplano que melhor separa os vetores de características em dois grupos é encontrado. Um texto qualquer é classificado pela determinação de que lado seu vetor de característica se encontra. Em um conjunto de teste de 800 endereços postais a taxa de acertos na classificação foi de 95%, segundo os autores.

Fan et al. [1998] desenvolveram uma metodologia para classificar uma linha de texto em impressa ou manuscrita. Analisando as larguras dos vales dos perfis das projeções horizontal e vertical dos pixels do texto, a orientação do leiaute do documento é decidida: vertical ou horizontal. Através de um algoritmo de corte, o texto é dividido em linhas e depois em blocos de caracteres. Um processo para eliminar os espaços em branco dos blocos é

executado. Esses são agrupados formando palavras ou sentenças dentro das linhas de texto, com base nas distâncias entre eles e em seus tamanhos. De cada bloco de caractere é determinado um ponto. Se a orientação do texto for horizontal esse ponto se encontra no centro do lado inferior do bloco. Se a orientação do texto for vertical esse ponto se encontra no centro do bloco. A variância das coordenadas y ou x (depende da orientação do texto) dos pontos de cada grupo de blocos de caracteres é calculada. Se a variância for maior do que um determinado limiar, o grupo de blocos é de texto manuscrito, caso contrário, é de impresso. Uma desvantagem dessa metodologia é não trabalhar com textos inclinados em relação à página. Para a avaliação da abordagem, foram usadas 25 imagens escaneadas de revistas e jornais como textos impressos e 25 de textos escritos por várias pessoas como manuscritos. Somente textos em inglês e chinês estavam presentes nas imagens. A taxa de acertos alcançada foi de 85%, segundo os autores.

Nos trabalhos de Pal e Chaudhuri [1999, 2001], é apresentada uma metodologia para separar as linhas de texto manuscrito das linhas de texto impresso em Bangla e Devnagari (escritas mais populares na Índia). Para tanto, características estruturais e estatísticas são extraídas das linhas. Uma técnica de limiarização baseada no histograma das tonalidades converte a imagem em tons de cinza para binária. As projeções vertical e horizontal dos pixels do texto na imagem são obtidas. Pela análise do perfil de cada projeção, é determinada a orientação do texto (retrato ou paisagem) e suas linhas são segmentadas. A análise é baseada nos picos e vales dos perfis. A classificação é realizada por meio de uma árvore de decisão de três níveis (nós) em fila, onde em cada nível se encontra uma característica extraída. Característica do primeiro nível: as letras do alfabeto dessas escritas possuem uma linha base superior. Então, em uma palavra de texto impresso, onde os caracteres se tocam, forma-se uma longa linha (horizontal ou vertical, dependendo da orientação), o que nem sempre acontece em palavras manuscritas. Se essa linha possuir comprimento menor do que um determinado limiar em todas as palavras da linha de texto em

análise, então tal linha de texto é de palavras manuscritas, caso contrário, ou é de texto impresso ou manuscrito e a decisão desce ao segundo nível. Característica do segundo nível: em linhas de texto impresso, se for apagada a região da linha base de todos os caracteres, cada caractere ficará isolado, o que nem sempre acontece com linhas de texto manuscrito. Portanto, se após apagar essa região nem todos os caracteres ficarem isolados, a linha é de texto manuscrito. Se isso não ocorrer, ela é de texto impresso ou manuscrito e a decisão desce ao terceiro nível. Característica do terceiro nível: com os caracteres isolados, após apagar a região da linha principal no nível anterior, a coordenada relativa à altura na página do ponto mais inferior (mais baixo) de cada caractere é anotada formando um conjunto de valores. Após algumas adaptações, a regularidade desse conjunto é avaliada pelo cálculo do desvio padrão. Se o valor desse desvio ficar abaixo de um determinado limiar, então esse conjunto é de coordenadas de pontos inferiores de caracteres impressos, senão, é de manuscritos. Foram usadas 600 imagens escaneadas de vários tipos de documentos contendo texto impresso e manuscrito. A acurácia do sistema, segundo os autores, foi de 98,6%. Os erros aconteceram onde a linha de texto era curta ou quando nela continha uma única palavra. O sistema é invariante em relação a estilos, tamanhos ou fonte do texto. Porém, as características extraídas são muito dependentes do formato das letras do tipo de escrita analisada.

Outro trabalho que separa texto manuscrito de impresso é descrito por Guo e Ma [2001]. Nele, a imagem é primeiramente segmentada em componentes conectados. Com base na regularidade das alturas e distâncias de letras vizinhas, componentes conectados são unidos em palavras. Pequenos componentes (ruídos) são eliminados antes do agrupamento. Para cada letra de uma palavra, o perfil da projeção vertical é obtido e quantizado em M níveis (número de símbolos). Na fase de classificação é analisada uma palavra de cada vez. Nesse caso, o perfil da projeção vertical de cada uma de suas letras é determinado. Esses perfis são concatenados formando uma sequência de observação. Baseados na teoria dos Modelos Ocultos de Markov, dois modelos probabilísticos são construídos na fase de treinamento, um

com letras impressas e outro com manuscritas. Esses modelos são usados para calcular a probabilidade da sequência de observação. O modelo que produzir a maior probabilidade é selecionado como a classe reconhecida. Há também uma fase de pós-processamento para correção de algumas classificações erradas. Devido à natureza da teoria usada (Modelos Ocultos de Markov), a abordagem só pode ser aplicada em linguagens de origem latina. Para a avaliação do sistema foram escaneados 25 documentos com textos impressos e manuscritos. A maior taxa de acertos alcançada foi de 92,86%, segundo os autores da pesquisa. Não houve comparações com outros trabalhos.

Santos et al. [2002-a, 2002-b] descrevem uma abordagem que distingue texto impresso e texto manuscrito em uma imagem de cheque de banco. O conteúdo da imagem de cheque é dividido em "objetos" de texto, dos quais são extraídas características relativas ao seu conteúdo e ao seu formato e, por serem características locais, dão ao sistema independência em relação ao leiaute do documento, podendo ser aplicada em diferentes circunstâncias. As características que descrevem o conteúdo fornecem informações referentes ao comportamento dos níveis de cinza dentro do "objeto" analisado. Nesse caso, os momentos estatísticos, média e desvio padrão, são usados para obter informações dos padrões de textura do objeto. Outras características são: a variação de intensidade, a densidade de pixels dentro do objeto, a curtose e a entropia. As extraídas relacionadas à forma dos objetos são: excentricidade, solidez e consistência. A classificação é realizada por uma rede neural treinada. Cheques de diferentes bancos de diferentes países foram utilizados para o treinamento e teste do sistema. Segundo os autores, em relação aos exemplos extraídos das imagens desses cheques, a taxa de erros foi de 12% e outros 12% foram considerados ambíguos. Para diminuir a taxa de ambíguos, um pós-processamento foi aplicado.

Zheng et al. [2002] apresentam uma abordagem que segmenta uma imagem de documento ruidosa e identifica manuscritos, principalmente anotações e correções. É realizada uma extração dos componentes conectados na imagem binária. Após esse passo, os

componentes próximos são agrupados em palavras, baseando-se na distância horizontal entre eles. Essas palavras formadas são cercadas por um retângulo envoltório de onde são extraídas características estruturais e de textura. As estruturais estão divididas em dois grupos, as referentes ao retângulo envoltório e as referentes aos componentes conectados dentro do retângulo envoltório. As do primeiro grupo são: largura, altura, relação de largura e altura e densidade de pixels do retângulo envoltório. E as do segundo grupo são: a média das alturas e larguras, relação de largura e altura, taxa de cobrimento e variância do perfil da projeção vertical dos componentes conectados. Uma das características de textura é obtida pelo histograma de co-ocorrência, o qual registra quantas vezes pares de pixels pretos ocorrem em uma fixada direção e distância. Para tanto, foram usadas quatro direções e quatro distâncias, em pixels, diferentes. Outras características referentes à textura são extraídas com auxílio da aplicação das técnicas: NxM-grams (intuitivamente são pequenos padrões em uma imagem e o princípio é de que se duas imagens possuem as mesmas ocorrências de padrões, então provavelmente pertençam a mesma classe), Pseudo Run Lengths e filtros de Gabor. A classificação é feita por um classificador linear Fisher. Para treinamento e teste do sistema, foram escaneadas 318 imagens de documentos do arquivo de uma indústria de tabaco. Três tipos de segmentação de texto foram testados: ao nível de caracteres (onde a menor porção para análises é um caractere), ao nível de palavras (onde a menor porção para análises é uma palavra) e ao nível de zona de texto. Segundo os autores, melhores resultados foram alcançados quando o texto foi segmentado ao nível de palavras, atingindo uma taxa de acertos de 97,3%.

Em um trabalho posterior, Zheng et al. [2003, 2004] reportam à outra abordagem. Uma das diferenças da abordagem anterior é que o texto da imagem de documento é segmentado somente no nível de palavras. Outra, está na classificação, pois, além de impressos e manuscritos, o sistema classifica ruídos. Esse fato torna o sistema uma ferramenta eficiente de pré-processamento para a eliminação de ruídos. A última diferença é uma

reclassificação baseada em Campos Aleatórios de Markov (MRF - Markov Random Field) para corrigir os casos de classificação incorreta. A acurácia total do sistema, segundo os autores, foi de 98,1%, considerando as três classes.

Kavallieratou e Stamatatos [2004] e Kavallieratou et al. [2004] apresentam um trabalho que tem por objetivo classificar linhas de texto em impresso ou manuscrito. Primeiramente, a inclinação do documento em relação à imagem é estimada. Depois, as áreas de texto são localizadas e têm seus respectivos ângulos de inclinação de igual modo estimados. Das áreas de textos, são extraídos componentes conectados os quais são cercados por retângulos envoltórios. Dependendo das características do retângulo envoltório o componente nele contido é eliminado, pois o mesmo é considerado como uma linha vertical ou linha horizontal ou ruído. Os componentes conectados remanescentes são unidos formando linhas de texto impresso ou manuscrito. As características extraídas de uma linha de texto são baseadas no perfil formado com o primeiro e o último pixel preto de cada coluna da linha. O histograma horizontal (projeção horizontal) desse perfil é obtido e divido, por determinados critérios, em três regiões: superior, principal (a do meio) e inferior. Os valores numéricos - largura da região superior dividida pela largura da região principal, largura da região inferior divida pela largura da região principal e área do histograma horizontal dividido pelo maior valor do próprio histograma - formam o vetor de características de cada linha de texto, o qual será enviado à fase de classificação. Essa, é baseada em uma análise de discriminante linear. Durante o treinamento, o centróide de cada classe (ponto representando a média da classe no conjunto de treinamento) é determinado. Então, um novo exemplo (um vetor de características de uma linha de texto) tem sua distância medida em relação ao centróide de cada classe pela distância de Mahalonobis. A menor distância decide a classe. Para avaliação do sistema, duas bases de imagens foram usadas. Uma delas é a IAM off-line Database 3.0 descrita por Marti e Bunke [1999, 2002] e disponível na Internet. A outra é a GRUHD Database, a qual não existe mais. Segundo os autores, a acurácia alcançada no conjunto de

teste foi de 97,9%. Para validação do resultado foi utilizada a técnica *10-fold Cross Validation*, porém não é especificado como a acurácia foi calculada. Não houve comparações com outros trabalhos. A maior ocorrência de erros foi registrada em linhas de texto manuscrito com uma única palavra ou quando elas eram muito curtas.

O trabalho descrito por Farooq et al. [2006] propõe uma metodologia para distinguir texto impresso e manuscrito na língua árabe. O texto é dividido em palavras que são cercadas por retângulos envoltórios. De cada retângulo envoltório, características de direções dos traços e uniformidade das palavras foram extraídas aplicando o filtro de Gabor direcional em seis direções e duas escalas diferentes, totalizando doze características. Foram coletados 34 formulários de imigração de 18 escritores diferentes. Cinco deles foram usados para o treinamento e o restante, para o teste do sistema. Usando uma rede neural, o sistema obteve, segundo os autores, precisão de 98,68% e acurácia de 85,58% em relação à distinção de manuscritos e precisão de 94,93% e acurácia de 98,25% em relação à distinção de impressos.

Trabalhos recentes de Koyama et al. [2008-a, 2008-b] descrevem uma metodologia para decidir se caracteres dos idiomas japonês e chinês são impressos ou manuscritos sem a necessidade de segmentar o texto em caracteres ou linhas de texto. Nessa metodologia, são consideradas regiões da imagem de tamanho geralmente igual à de um caractere. No primeiro passo, cada região é transformada em um espectro de energia aplicando a transformada rápida de Fourier bidimensional. Dentro desses espectros são calculados os valores das características, os quais representam a energia em orientações pré-determinadas. No total, foram estabelecidas 16 orientações de observação. Segmentos de linhas em caracteres impressos constroem um número limitado de padrões, enquanto segmentos de linhas em caracteres manuscritos não, pois esses últimos possuem segmentos irregulares. Comparados com os valores das características de segmentos de caracteres impressos, os de manuscritos têm grandes variações. Esses valores são enviados a uma rede

neural, a qual realiza a classificação. Segundo os autores, a taxa de distinção correta alcança 99,7% em determinada classe de caractere devido a sua estrutura simples.

É importante lembrar que, na maioria dos trabalhos, não há comparações com outras metodologias que realizam a mesma tarefa e os avaliadores da classificação não são descritos. Também, é relevante registrar que as bases de imagens utilizadas para treinamento e teste dos sistemas desenvolvidos, não estão disponíveis para serem utilizadas, impossibilitando comparações. Foram feitos contatos, por e-mail, com os autores desses trabalhos solicitando informações sobre as bases utilizadas em seus respectivos trabalhos, mas não houve êxito na quase totalidade dos casos. As respostas recebidas foram as seguintes: a política da instituição não permite o acesso por outros grupos de pesquisas ou a base não existe mais. Em outros casos, o endereço eletrônico do autor não é mais válido. No apêndice B desta dissertação, encontram-se as cópias dos e-mails enviados e suas respostas.

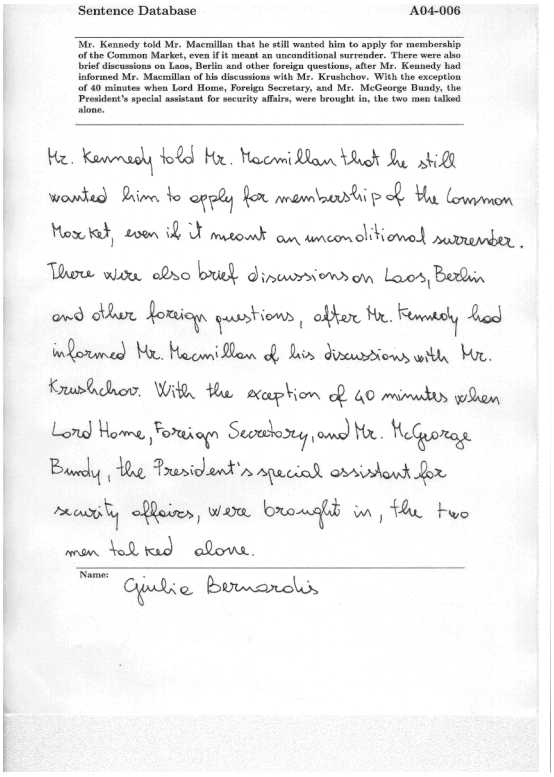
4. METODOLOGIA PROPOSTA

Neste capítulo, a metodologia proposta é detalhada, etapa por etapa: (1) a descrição do tipo de documento processado pelo sistema; (2) as técnicas de processamento de imagens aplicadas; (3) a segmentação do texto em palavras e a extração de suas características e (4) a classificação executada pelo sistema a fim de distinguir texto impresso e texto manuscrito.

4.1 Tipo de documento analisado pelo sistema

O sistema analisa documentos do tipo formulário. São documentos preenchidos por pessoas com finalidades diversas como, por exemplo, cadastros institucionais, questionários de pesquisas, cheques bancários, etc. Desta forma, possuem palavras impressas e manuscritas. Contudo, não possuem figuras, tabelas, gráficos ou outro tipo de elemento.

As figuras 4.1 e 4.2 exibem dois exemplos desses documentos. Na Figura 4.1 observa-se um formulário onde o texto impresso e o manuscrito nunca se apresentam na mesma linha, o que não acontece com o formulário da Figura 4.2, uma espécie de cadastro. Nesse último, a pessoa preenche seus dados sobre linhas horizontais logo após uma palavra, ou palavras, impressas.



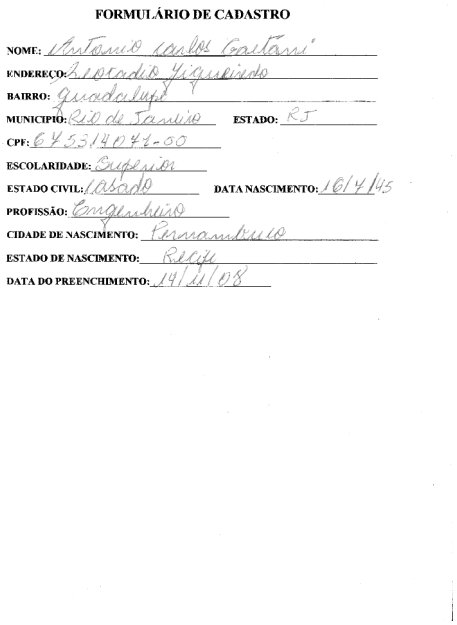
Sentence Database A04-006

Mr. Kennedy told Mr. Macmillan that he still wanted him to apply for membership of the Common Market, even if it meant an unconditional surrender. There were also brief discussions on Laos, Berlin and other foreign questions, after Mr. Kennedy had informed Mr. Macmillan of his discussions with Mr. Krushchov. With the exception of 40 minutes when Lord Home, Foreign Secretary, and Mr. McGeorge Bundy, the President's special assistant for security affairs, were brought in, the two men talked alone.

Mr. Kennedy told Mr. Macmillan that he still wanted him to apply for membership of the Common Market, even if it meant an unconditional surrender. There were also brief discussions on Laos, Berlin and other foreign questions, after Mr. Kennedy had informed Mr. Macmillan of his discussions with Mr. Krushchov. With the exception of 40 minutes when Lord Home, Foreign Secretary, and Mr. McGeorge Bundy, the President's special assistant for security affairs, were brought in, the two men talked alone.

Name: Giulie Bernardis

Figura 4.1. Formulário com manuscrito separado do texto impresso.



FORMULÁRIO DE CADASTRO

NOME: Antonio Carlos Gaetani
ENDEREÇO: Estrada Figueiredo
BAIRRO: Guadalupe
MUNICIPIO: Rio de Janeiro ESTADO: RJ
CPF: 64531407-00
ESCOLARIDADE: Superior
ESTADO CIVIL: Casado DATA NASCIMENTO: 16/4/45
PROFISSÃO: Engenheiro
CIDADE DE NASCIMENTO: Pernambuco
ESTADO DE NASCIMENTO: Recife
DATA DO PREENCHIMENTO: 14/11/08

Figura 4.2. Formulário de cadastro.

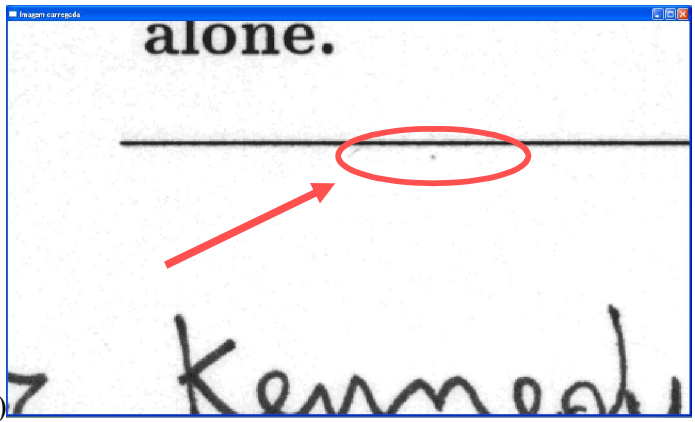
4.2 Pré-processamento da imagem

Com a imagem do formulário capturada e "carregada", o sistema inicia o pré-processamento. Essa etapa tem por finalidade preparar a imagem para sua segmentação em palavras.

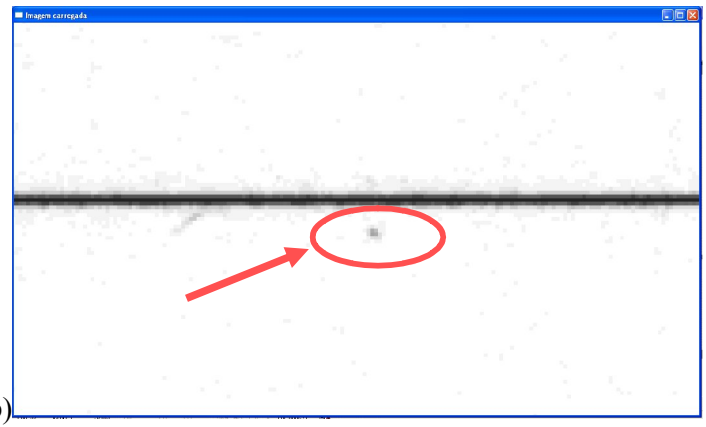
4.2.1 Redução de ruídos por filtragem espacial

Um filtro de mediana percorre a imagem, ainda em tons de cinza, para a redução de ruídos. Na forma usada neste trabalho, a filtragem por mediana torna os pixels de pontos pequenos (ruídos), provenientes do processo de captura da imagem, mais claros. O objetivo é que esses pontos desapareçam na etapa de binarização, quando o texto é separado do fundo.

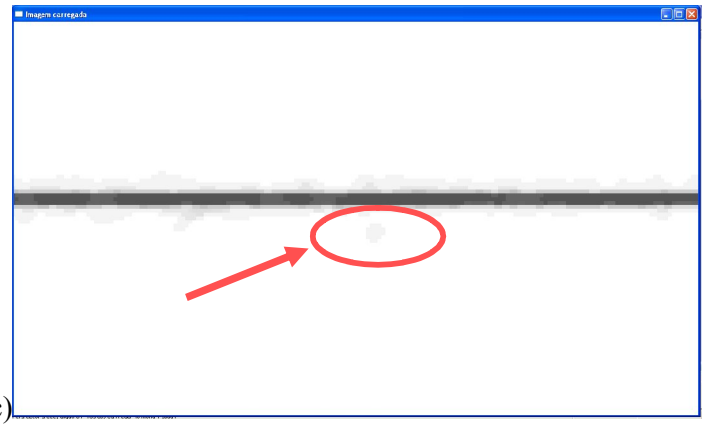
Para a ordenação crescente das tonalidades na aplicação do filtro, foi implementado o algoritmo descrito por Cormen et al. [2002]. A Figura 4.3 mostra a aplicação desse filtro pelo sistema. Em a) uma imagem no tamanho original, em b) sua ampliação e em c) o resultado da filtragem por mediana, destacado pela seta.



a)



b)



c)

Figura 4.3. Filtragem espacial por filtro de mediana.

4.2.2 Binarização

O texto é separado do fundo da imagem aplicando a técnica de limiarização, apresentada na Seção 2.3. Nesse caso, um limiar global ótimo T é automaticamente determinado pelo método de Otsu [1979], também descrito na Seção 2.3 e implementado no sistema aqui desenvolvido (Algoritmo 4.1). A imagem é percorrida pixel por pixel e todo aquele com nível de cinza menor ou igual ao limiar determinado passa a ter tonalidade preta (0), caso contrário, branca (255). O resultado é a binarização da imagem. Esse processo é descrito pela equação:

$$g(x,y) = \begin{cases} 0 & se \quad f(x,y) \leq T \\ 255 & se \quad f(x,y) > T \end{cases} \tag{4.1}$$

onde $g(x,y)$ é a imagem binarizada produzida e $f(x,y)$ é a imagem original.

O método de Otsu foi escolhido por mostrar-se muito adequado a esse tipo de tarefa e às características da imagem processada, a qual possui caracteres pretos sobre um fundo banco. Imagens similares foram objetos de análise no trabalho de Monteiro [2002], onde resultados satisfatórios foram alcançados. Outro motivo da escolha, são os resultados superiores na segmentação de imagens, quando comparados com os de outros métodos, de acordo com o trabalho de Seixas et al. [2008]. Nesse trabalho, vários métodos são testados e seus resultados são confrontados. Eles executam a segmentação de imagens em tons de cinza.

```
1. Limiarizacao( )
2.     Para i = 0 até i = Largura_da_imagem
3.         Para j = 0 até j = Altura_da_imagem
4.             r = Tonalidade_do_pixel (i, j)
5.             Tonalidade[r] = Tonalidade[r] + 1
6.             Número_de_pixels = Número_de_pixels + 1
7.             j = j + 1
8.         i = i + 1
9.     Para i = 0 até i = 255
10.        Probabilidade[i] = Tonalidade[i]/Número_de_pixels
11.        i = i + 1
```

```
12.  Para i = 0 até i = 255
13.        Ut = Ut + i*Probabilidade[i]
14.        i = i + 1
15.  Para i = 0 até i = 255
16.        Variância_total = Variância_total + (i – Ut)*(i – Ut)* Probabilidade[i]
17.        i = i + 1
18.  Função = 0,   Funçãoaux = 0
19.  Para k = 0 até k = 255
20.        w0 = 0,   w1 = 0,   u0 = 0,   u1 = 0,  uk = 0
21.        Para i = 0 até i = k
22.              w0 = w0 + Probabilidade[i],    i = i +1
23.        w1 = 1 – w0
24.        Para i = 0 até i = k
25.              uk = uk + i*Probabilidade[i]
26.        u0 = uk/w0
27.        u1 = (Ut – uk)/(1 – w0)
28.        variância_entre_classe = w0*w1*(u1 – u0)*(u1 – u0)
29.        Função = Variância_entre_classes/Variância_total
30.        Se (Função > Funçãoaux)
31.              Limiar = k
32.              Funçãoaux = Função
33.        k = k + 1
```

Algoritmo 4.1. Método de Otsu.

O texto, por ser a porção mais escura na imagem original, possui pixels de tonalidades menores ou iguais a *T* e, por isso, na imagem binarizada terá a cor preta. Já o fundo é a porção mais clara da imagem original, seus pixels têm tonalidades maiores que *T*, e, por esse motivo, na imagem binarizada terá a cor branca. Na Figura 4.4-a o texto em tons de cinza e na Figura 4.4-b (imagem binarizada) o texto já separado.

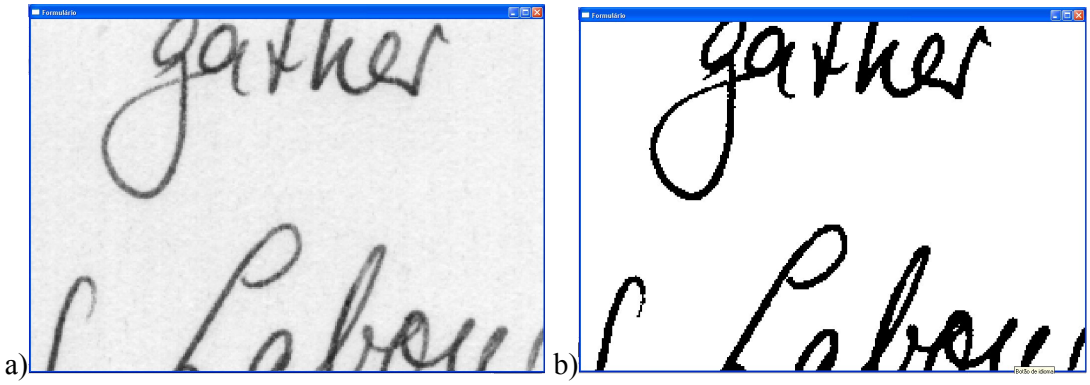


Figura 4.4. Binarização de uma imagem de formulário.

4.2.3 Linhas horizontais no formulário

Mesmo que não contenha figuras, tabelas, gráficos e outros elementos, um formulário, quase que inevitavelmente, possui linhas horizontais. A metodologia proposta possui dois destinos para elas: ou as reconhece e as ignora ou as extrai (as elimina da imagem).

Inicialmente, o sistema apenas extraía as linhas por meio de um simples procedimento implementado especificamente para isso (Algoritmo 4.2). A linha era identificada e extraída. O objetivo era extrair as que serviam de base para o preenchimento do manuscrito, como ocorre no formulário da Figura 4.2, porque elas diminuem o desempenho do sistema. Porém, em formulários, nos quais as linhas horizontais apenas separam regiões, como os da Figura 4.1, ou tenham qualquer outra função que não seja a de servir de base para o preenchimento do manuscrito, o sistema as reconhece, mas não as extrai, e sim as ignora. O reconhecimento é feito com base em suas medidas. Linhas horizontais possuem valores de comprimento muito acima da média e de altura muito abaixo da média dos valores de comprimento e altura, respectivamente, de palavras.

Na Figura 4.5, linhas horizontais são extraídas e na Figura 4.6 uma linha é ignorada e, por isso, não está cercada por um retângulo envoltório como as palavras.

```
1. Eliminando Linhas( )
2.   Para y = 50 até y = altura_da_imagem – 50
3.       k = 0
4.       Para x = 50 até x = largura_da_imagem – 50
5.               r = cor_do_pixel de coordenada (x, y)
6.               Se (r = preta)
7.                   k = k + 1
8.               Se (r = branca)
9.                   k = 0
10.              Se (k = 100)
11.                  Para  t = –300 até t = 500
12.                      Pintar o pixel (x + t, y) de branco
13.                      t = t + 1
14.                  k = 0
15.              x = x + 1
16.      y = y + 1
```

Algoritmo 4.2. Eliminando linhas horizontais.

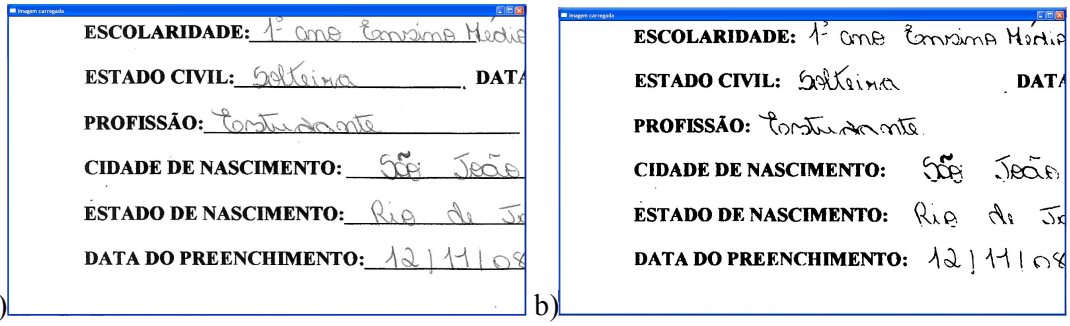


Figura 4.5. Linhas horizontais extraídas.

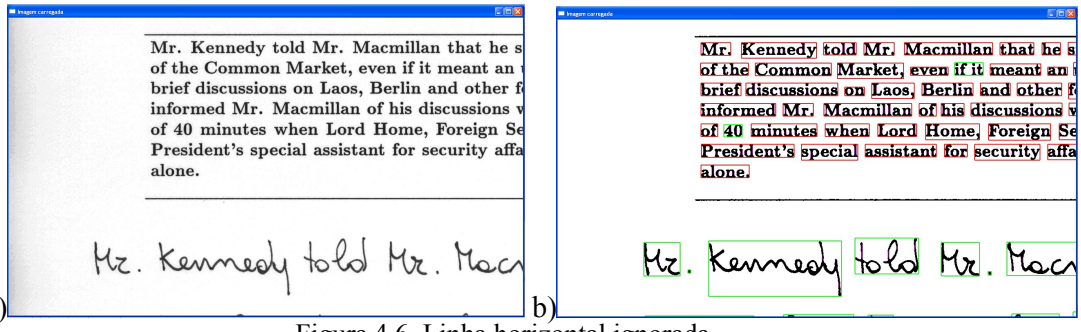


Figura 4.6. Linha horizontal ignorada.

4.2.4 Eliminação de ruídos e suavização de contornos verticais por abertura morfológica

A abertura morfológica (Seção 2.5.1.4) é aplicada para eliminar ruídos, os quais permaneceram na imagem após sua binarização, e segmentos pequenos, os quais surgem na região da borda das linhas, depois que elas são eliminadas. A Figura 4.7-a mostra pequenos pontos e segmentos isolados, eliminados na Figura 4.7-b.

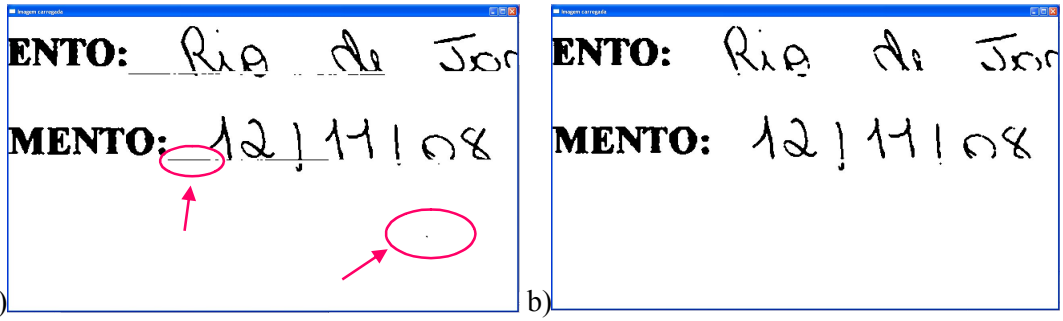


Figura 4.7. Eliminado ruídos por abertura morfológica.

Outro efeito da aplicação da abertura morfológica é a suavização dos contornos verticais das letras no texto. Ele é importante, pois os contornos verticais das letras são posteriormente detectados e utilizados em características extraídas, as quais são reforçadas pela suavização. A Figura 4.8 mostra no destaque (interior da elipse) o que acontece com os contornos e a Figura 4.9 o elemento estruturante usado pelo sistema, em que seu pixel central está representado pelo quadrado preto.

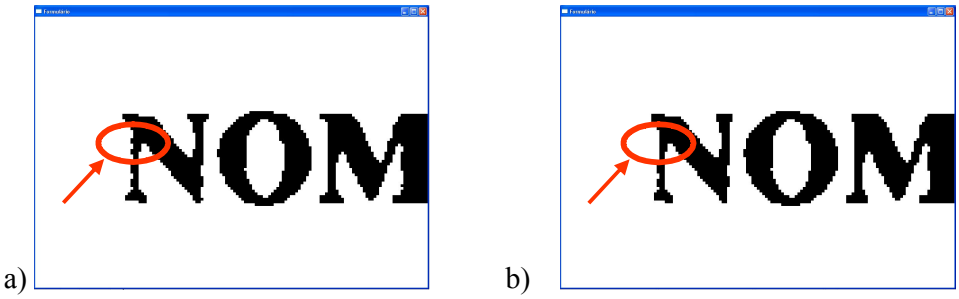


Figura 4.8. Suavização de bordas verticais.

Figura 4.9. Elemento estruturante.

4.3 Extração de componentes conectados

A extração de componentes conectados (Seção 2.4) é realizada utilizando a vizinhança $N_8(p)$ (Seção 2.4.1). Porém o algoritmo de rotulação desses, é diferente do apresentado na Seção 2.4.3, onde a imagem é novamente percorrida observando as equivalências anotadas. O algoritmo aqui implementado (Algoritmo 4.3) é baseado no descrito por Pitas [1995]. Por ele ser recursivo, a imagem não precisa ser percorrida novamente.

```
ComponeteConectadoAux(x, y, c)
    Para m = –1 até m = 1
        Para n = –1 até n = 1
            r = cor_do_pixel_de_coordenada (x + m,y + n)
            Se((r = 0) e (c ≠ Componente[x + m][y + n]))
                Componente[x + m][y + n] = c
                xaux = x + m,  yaux = y + n
                ComponenteConectadoAux(xaux, yaux, c)
        n = n + 1
    m = m + 1
```

```
11. Fim  ComponenteConectadoAux
12. ComponenteConectado( )
13.         Para x = 1 até x = largura_da_imagem – 1
14.                     Para y = 1 até y = altura_da_imagem – 1
15.                                 r = cor_do_pixel_de_coordenada (x, y)
16.                                 Se((r = 0) e (Componente[x][y] = 0))
17.                                             c = c + 1
18.                                             Componente[x][y] = c
19.                                             ComponenteConectadoAux(x, y, c)
20.                     y = y + 1
21          x = x + 1
22. Fim ComponenteConectado
```

Algoritmo 4.3. Extração de componentes conectados.

4.4 União dos componentes conectados em palavras

Após a extração dos componentes conectados, eles são unidos com o objetivo de formarem palavras. Cada componente conectado é cercado por um retângulo envoltório, o qual é o menor retângulo possível capaz de contê-lo. Retângulos envoltórios, que estão em uma mesma linha de texto na imagem, são unidos, dependendo da distância entre eles. O avaliador dessa distância é dado pela equação:

$$Distância = \frac{M_L = \frac{\sum_{i=0}^{k} L_i}{k}}{2} \tag{4.2}$$

onde L_i é a largura de cada retângulo envoltório; k é a quantidade de retângulos envoltórios na imagem e M_L é a média de suas larguras. Geralmente, a largura do retângulo envoltório de um componente conectado é igual a de uma letra impressa e igual a distância entre palavras impressas no texto. A média dessas larguras dividida por dois é um valor que, normalmente, é menor do que a distância entre palavras impressas e é maior do que entre letras impressas. Assim, apenas as letras de uma mesma palavra são unidas, justificando o uso da Equação (4.2). A distância não é baseada nas medidas dos elementos manuscritos, pois eles são muito irregulares, inclusive em uma mesma página de documento.

Neste trabalho, a distância entre dois retângulos envoltórios é a distância entre as retas paralelas que passam pela lateral direita de um e esquerda do outro, Figura 4.10.

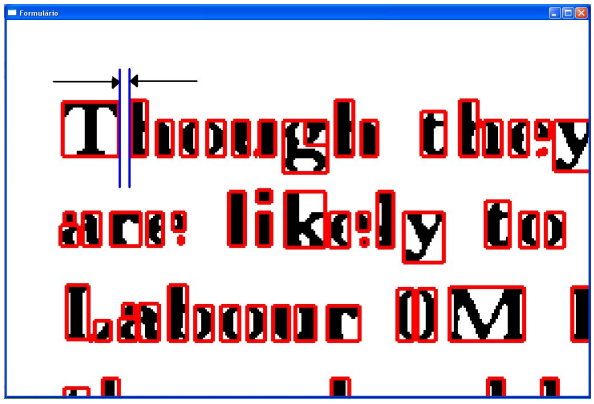


Figura 4.10. Distância entre dois retângulos envoltórios.

Como a imagem está referenciada por um sistema de coordenadas ortogonal, o cálculo da distância entre essas retas é obtido pelo módulo da diferença entre as coordenadas *x* mínimo de um retângulo envoltório e *x* máximo do outro. Esse cálculo é feito para cada dois retângulos envoltórios em uma mesma linha de texto na imagem e se o seu valor for menor ou igual ao valor obtido na Equação (4.2), esses retângulos envoltórios são unidos. A Figura 4.11 ilustra a união dos retângulos envoltórios próximos da Figura 4.10.

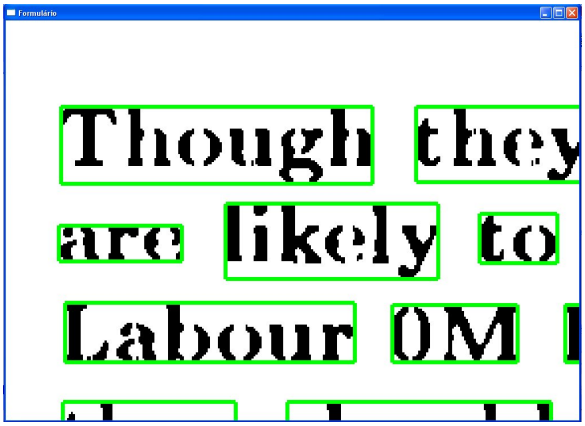


Figura 4.11. União de retângulos envoltórios da Figura 4.10.

Outro procedimento de união de dois retângulos envoltórios é baseado na interseção de áreas. O sistema verifica para cada dois retângulos envoltórios de uma mesma linha se há alguma porção de área em comum e, quando comprova a existência, os une. A

interseção de áreas de dois retângulos envoltórios ocorre principalmente quando um está cercando uma palavra manuscrita e o outro o ponto de uma letra “i”. A Figura 4.12 ilustra como isso ocorre. Na Figura 4.12-a dois retângulos envoltórios com interseção de áreas e na Figura 4.12-b os dois já unidos.

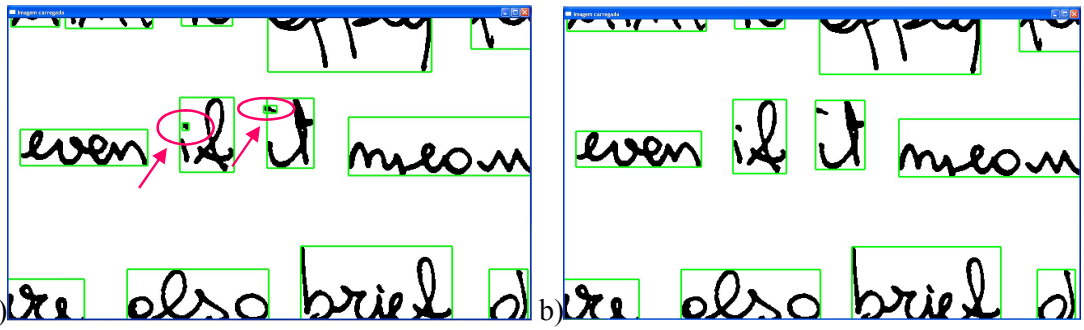


Figura 4.12. União de retângulos envoltórios com interseção de áreas.

A escolha de unir os componentes no nível de palavras é baseada nas conclusões do trabalho de Zheng et al. [2002], onde os melhores resultados de classificação de impressos e manuscritos são alcançados quando o texto é segmentado em palavras.

4.5 Características extraídas

As características usadas para a distinção de palavras impressas e manuscritas são extraídas dos retângulos envoltórios. Isso quer dizer que o sistema classifica um desses como contendo uma palavra impressa ou manuscrita, ou seja, os retângulos envoltórios são regiões da imagem que abarcam um desses dois tipos de palavras. As características extraídas são: Desvio da Largura, Desvio da Altura, Desvio da Área, Densidade, Variância da Projeção Vertical, Maior Diferença Encontrada na Projeção Horizontal, Distribuição de Pixels, Divisão da Linha Inferior de Pixels pela Largura, Soma das Divisões de Pixels de Cada Linha pela Largura, Divisão do Maior Contorno Vertical pela Altura e Divisão da Soma dos Comprimentos dos Contornos Verticais pela Área. Elas têm a função de representar a palavra dentro de cada retângulo envoltório e apresentam valores diferentes para as impressas e as

manuscritas. Esses valores são usados nas regras de classificação, as quais decidem o tipo de palavra que o retângulo envoltório contém. Cada uma dessas características é descrita detalhadamente a seguir.

4.5.1 Desvio da Largura, Desvio da Altura e Desvio da Área

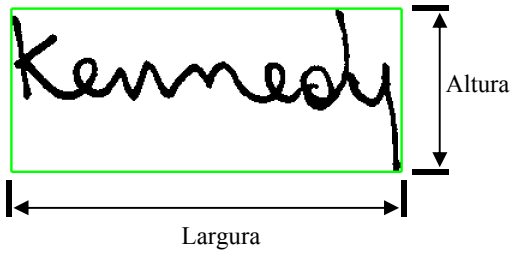
Primeiramente, as médias das larguras, das alturas e das áreas dos retângulos envoltórios, na imagem do documento são calculadas. Para cada retângulo envoltório, o módulo da diferença de sua largura pela média das larguras é registrado como uma característica. O mesmo é feito com sua altura e sua área. Essas características podem ser representadas pelas equações:

$$D_{LG} = \left| LG - M_{LG} \right| \tag{4.3}$$

$$D_{AT} = \left| AT - M_{AT} \right| \tag{4.4}$$

$$D_{AR} = \left| AR - M_{AR} \right| \tag{4.5}$$

onde LG é a largura, AT é a altura e AR é a área e M_{LG}, M_{AT} e M_{AR} médias das larguras, das alturas e das áreas, respectivamente.



$$Área = \text{largura} \times \text{altura}$$

Figura 4.13. Largura, altura e área de um retângulo envoltório.

Para um retângulo envoltório, a largura é a diferença entre as abscissas máxima e mínima e a altura, a diferença entre as ordenadas máxima e mínima (Figura 4.14). A área é o produto dessas diferenças. Outro fato é que a imagem se encontra sempre no primeiro quadrante do sistema de coordenadas e, por isso, as coordenadas x e y são sempre positivas.

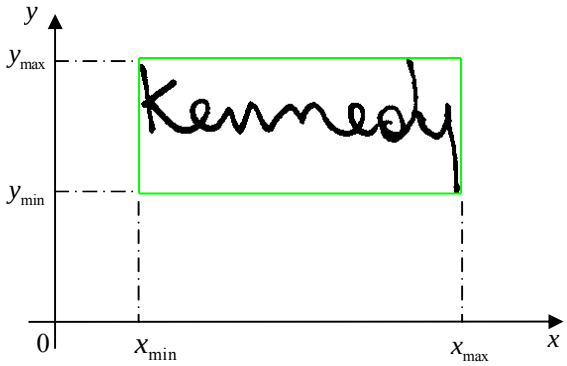


Figura 4.14. Coordenadas de um retângulo envoltório.

A extração dessas características se justifica pelo fato de as palavras impressas possuírem medidas mais regulares do que as manuscritas, principalmente a altura. Essas três características, embora muito simples, não foram aplicadas em trabalhos anteriores.

4.5.2 Densidade

Como as palavras impressas por máquinas são mais regulares e compactas, geram retângulos envoltórios com menos espaços em branco do que os gerados pelas palavras manuscritas (Figura 4.15). Dessa forma, a densidade de pixels pretos dentro de um retângulo envoltório, é maior para o primeiro tipo. Essa característica também é usada nos trabalhos de Zheng et al. [2002, 2003, 2004]. A densidade é dada pela equação:

$$Densidade = \frac{\text{Quantidade de pixels pretos dentro do retângulo envoltório}}{\text{Área do retângulo envoltório}} \tag{4.6}$$

Figura 4.15. Retângulos envoltórios envolvendo uma palavra impressa e outra manuscrita.

4.5.3 Variância da Projeção Vertical

Antes da exposição dessa característica, e da próxima, é necessário detalhar as projeções horizontal e vertical dos pixels em uma imagem binarizada.

Essas projeções são frequentemente usadas em características extraídas para a tarefa de distinção de textos impressos e manuscritos. A projeção horizontal/vertical dos pixels é obtida acumulando os pixels pretos existentes em uma mesma linha/coluna. Isso é feito para cada linha/coluna da imagem e, ao final desse processo, os resultados são resumidos em um histograma [Fan et al., 1998], o qual é geralmente chamado de histograma horizontal/vertical da imagem [Kavallieratou & Stamatatos, 2004].

Nessa característica, a projeção vertical é feita em relação aos pixels da palavra (os pixels pretos) dentro do retângulo envoltório. As figuras 4.16 e 4.17 mostram essa projeção para uma palavra impressa e outra manuscrita, respectivamente.

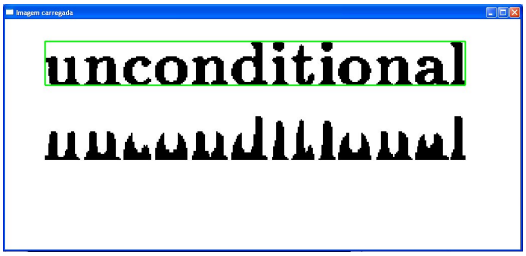


Figura 4.16. Projeção vertical de uma palavra impressa.

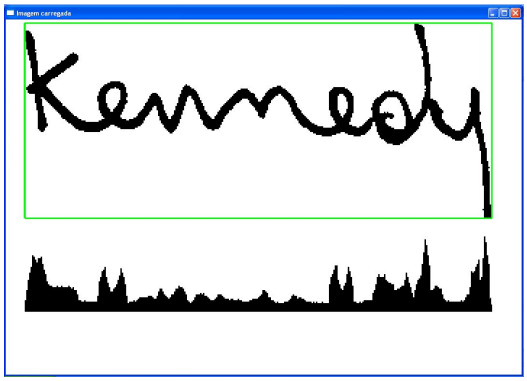


Figura 4.17. Projeção vertical de uma palavra manuscrita.

A variância da projeção vertical mede o quanto o seu perfil é suave ou cheio de picos altos e vales profundos. Dado um conjunto de valores, a variância (Seção 2.6.5) verifica a sua homogeneidade, ou seja, se há grandes diferenças de seus valores em relação a sua média. No problema em questão, esse conjunto é formado pelas coordenadas verticais de cada ponto do perfil da projeção. A Figura 4.18 é o perfil da projeção da Figura 4.17.

Figura 4.18. Perfil da projeção vertical de uma palavra manuscrita.

Como ilustrado nas figuras 4.16 e 4.17, o perfil da projeção vertical de uma palavra manuscrita é mais suave, ou seja, o conjunto de valores das coordenadas verticais de cada um de seus pontos é mais homogêneo do que o de uma palavra impressa, conduzindo a uma variância menor. Como a anterior, essa característica também foi usada nos trabalhos de Zheng et al. [2002, 2003, 2004].

4.5.4 Maior Diferença Encontrada na Projeção Horizontal

A projeção horizontal nessa característica, assim como a projeção vertical, é determinada em relação aos pixels da palavra dentro do retângulo envoltório. As figuras 4.19 e 4.20 mostram-na para uma palavra impressa e outra manuscrita, nessa ordem.

Figura 4.19. Projeção horizontal de uma palavra impressa.

Figura 4.20. Projeção horizontal de uma palavra manuscrita.

Com a projeção horizontal disponível, seu perfil é percorrido e o módulo da maior diferença das abscissas de pontos adjacentes desse perfil é armazenado como uma característica. Analisando os perfis das projeções nas figuras 4.19 e 4.20 nota-se que, pela regularidade, o perfil de uma palavra impressa possui, em determinadas regiões, diferenças grandes de um ponto a outro adjacente em relação à abscissa, enquanto uma palavra manuscrita gera um perfil de projeção de curvas mais suaves e então as diferenças de pontos adjacentes, em relação à abscissa, são geralmente menores. Essa característica é usada no trabalho de Violante et al. [1995] para fazer a distinção de impressos e manuscritos em textos de endereços de cartas.

4.5.5 Distribuição de Pixels

Para analisar sua distribuição de pixels, o retângulo envoltório é dividido ao meio por uma linha horizontal. Em seguida, é calculada a densidade (Equação 4.6) na porção superior e na porção inferior, e o módulo da diferença entre essas densidades é armazenado como uma característica, a qual é representada pela equação:

$$Distribuição\ de\ Pixels = |DS - DI| \tag{4.7}$$

onde *DS* é a densidade da porção superior e *DI* a densidade da porção inferior. A distribuição dos pixels do texto em um retângulo envoltório é mais equilibrada quando ele contém uma palavra impressa em vez de uma manuscrita, devido à falta de regularidade da última. As figuras 4.21 e 4.22 ilustram a divisão de retângulos envoltórios.

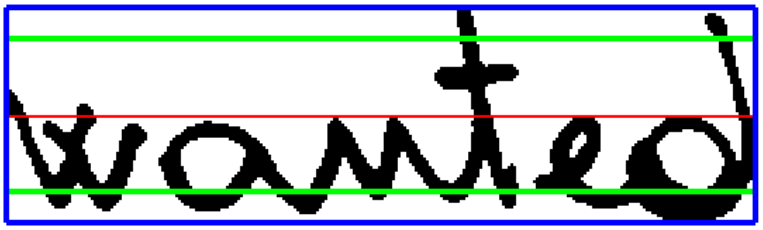


Figura 4.21. Divisão de retângulo envoltório com palavra manuscrita.

Figura 4.22. Divisão de retângulo envoltório com palavra impressa.

Nas figuras anteriores, a linha vermelha indica o local da divisão. A densidade da porção superior é calculada entre a linha vermelha e a linha verde superior e a densidade da porção inferior é calculada entre a linha vermelha e a linha verde inferior.

Inicialmente, o procedimento para o cálculo dessa característica era diferente. Antes, o retângulo envoltório também era dividido na direção vertical. O módulo da diferença entre as densidades das porções esquerda e direita era somado ao módulo da diferença entre as densidades das porções superior e inferior para constituir o valor numérico da característica. Além do mais, a altura considerada não era reduzida como indicada pelas linhas verdes nas figuras 4.21 e 4.22. No entanto, na fase de análise de características mais significativas, ela nunca era apontada. Então, alterações foram experimentadas chegando à atual versão. Apenas a divisão horizontal foi considerada por apresentar melhores resultados, e a decisão de redução da altura no retângulo envoltório, indicada pelas linhas verdes, foi motivada pela observação de que algumas palavras impressas contendo letras como “p” e “l” (Figura 4.22) aumentavam o espaço em branco dentro do retângulo envoltório de forma não-uniforme, devido o formato dessas letras.

As mudanças realizadas fizeram com que essa característica apresentasse valores bem diferentes para retângulos envoltórios contendo palavras impressas e manuscritas. A grande diferença nos valores se deve, como já comentado, à falta de regularidade das palavras manuscritas, claramente notada na Figura 4.21, e a presença dela nas impressas, percebida na Figura 4.22. Nessa forma de avaliação essa característica não foi encontrada em nenhum trabalho pesquisado.

4.5.6 Divisão da Linha Inferior de Pixels pela Largura

Para cada retângulo envoltório é determinado quantos pixels (pontos) da palavra, nele contida, possuem o valor da coordenada vertical igual ao valor do seu *y* mínimo, ou seja, quantos pixels da palavra "encostam" na sua linha inferior. O valor do número de pixels nessa condição é então dividido por sua largura e considerado como uma característica. As figuras 4.23 e 4.24 ilustram-na.

Na Figura 4.23, a seta destaca a região onde a palavra manuscrita "encosta" na linha inferior do retângulo envoltório. Na Figura 4.24, são várias as regiões em que a palavra "encosta" na linha inferior. Essa característica apresenta um valor geralmente maior em palavras impressas do que nas palavras manuscritas, pois, nas palavras impressas, uma quantidade maior de pontos "encostam" na linha inferior do retângulo envoltório.

Figura 4.23. Linha inferior de pixels de uma palavra manuscrita.

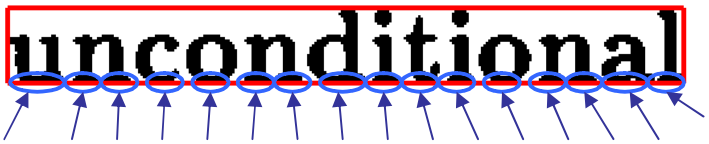


Figura 4.24. Linha inferior de pixels de uma palavra impressa.

A motivação para a extração dessa característica baseia-se, mais uma vez, na compacidade e na regularidade das palavras impressas, as quais apresentam uma linha horizontal imaginária usada no posicionamento vertical das letras, enquanto palavras manuscritas não, pois seres humanos não possuem coordenação motora suficiente para gerarem essa linha enquanto escrevem. Essa forma de avaliação também não foi encontrada em nenhum trabalho pesquisado.

4.5.7 Soma das Divisões de Pixels de Cada Linha pela Largura

Para cada retângulo envoltório, o número de pixels da palavra nele contida (pixels pretos) em cada linha é dividido por sua largura. A soma dos resultados dessas divisões é armazenada como uma característica. A razão para a extração dessa característica é a maior compacidade das palavras impressas em relação às manuscritas. Essa característica, nessa forma, não foi encontrada em nenhum trabalho pesquisado.

4.5.8 Divisão do Maior Contorno Vertical pela Altura

No texto, as letras impressas e manuscritas possuem alguns contornos verticais, os quais, nas figuras 4.25 e 4.26, estão destacados na cor magenta. Esses são detectados pelos filtros espaciais mostrados na Figura 4.27, os quais são baseados no filtro de Prewitt direcional abordado na Seção 2.2.

Figura 4.25. Contornos verticais em palavras impressas.

Figura 4.26. Contornos verticais em palavras manuscritas.

- 1	0	1
- 1	0	1
- 1	0	1
- 1	0	1
- 1	0	1
- 1	0	1
- 1	0	1

1	0	- 1
1	0	- 1
1	0	- 1
1	0	- 1
1	0	- 1
1	0	- 1
1	0	- 1

Figura 4.27. Filtros espaciais para detecção dos contornos verticais.

O comprimento dos contornos verticais em algumas letras de uma palavra impressa é quase igual à altura do retângulo envoltório que a cerca. Porém, em letras de palavras manuscritas, os contornos verticais (quando existem) são de comprimentos bem menores do que a altura do retângulo envoltório. O fator fundamental para essa diferença é a forma mais regular das letras impressas do que a das manuscritas. Então, a razão entre o maior contorno vertical encontrado dentro do retângulo envoltório e sua altura é mais uma característica usada na implementação desta dissertação. Consequentemente, em retângulos envoltórios contendo palavras impressas, essa característica possui valores maiores do que naqueles contendo palavras manuscritas. As figuras 4.28 e 4.29 destacam o maior contorno vertical de uma palavra impressa e de uma manuscrita, nessa ordem. Essa característica não foi encontrada em nenhum trabalho pesquisado.

Figura 4.28. Maior contorno vertical em uma palavra impressa.

Figura 4.29. Maior contorno vertical em uma palavra manuscrita.

4.5.9 Divisão da Soma dos Comprimentos dos Contornos Verticais pela Área

Essa característica utiliza os contornos verticais detectados para o cálculo da característica anterior. Como ilustrado pelas figuras 4.25 e 4.26, retângulos envoltórios contendo palavras impressas possuem número maior, e de maiores comprimentos, de contornos verticais do que os retângulos envoltórios com palavras manuscritas. Por outro lado, os do primeiro tipo possuem áreas geralmente menores do que os do segundo tipo, devido à regularidade e à compacidade das palavras impressas. Logo, a soma dos comprimentos de todos os contornos verticais detectados nas letras em um retângulo envoltório dividida por sua área é considerada uma característica interessante a ser avaliada. Naturalmente, nos retângulos envoltórios contendo palavras impressas ela apresenta valores bem maiores na maioria das vezes. Essa característica também não foi usada em nenhum trabalho pesquisado.

A Tabela 4.1 é um resumo das características extraídas, nela encontram-se uma descrição breve de cada uma e seu respectivo autor. A ausência de autor na terceira coluna indica que a característica correspondente foi desenvolvida neste trabalho.

Tabela 4.1. Resumo das características extraídas.

Característica	Descrição	Autor
Desvio da Largura	É o módulo da diferença da largura do retângulo envoltório pela média das larguras dos retângulos envoltórios na imagem do documento.	
Desvio da Altura	É o módulo da diferença da altura do retângulo envoltório pela média das alturas dos retângulos envoltórios na imagem do documento.	
Desvio da Área	É o módulo da diferença da área do retângulo envoltório pela média das áreas dos retângulos envoltórios na imagem do documento.	
Densidade	É a densidade de pixels pretos dentro do retângulo envoltório	Zheng et al. [2002, 2003, 2004]
Variância da Projeção Vertical	É o cálculo da variância das coordenadas verticais dos pontos do perfil da projeção horizontal dos pixels pretos do retângulo envoltório.	Zheng et al. [2002, 2003, 2004]

Maior Diferença Encontrada na Projeção Horizontal	É a maior diferença encontrada entre as coordenadas horizontais de pontos adjacentes do perfil da projeção horizontal dos pixels pretos dentro do retângulo envoltório.	Violante et al. [1995]
Distribuição de Pixels	É a análise da homogeneidade da distribuição de pixels pretos dentro do retângulo envoltório.	
Divisão da Linha Inferior de Pixels pela Largura	É o valor da divisão do número de pixels pretos dentro do retângulo envoltório, que encostam no seu lado inferior, por sua largura.	
Soma das Divisões de Pixels de Cada Linha pela Largura	É a soma das divisões do número de pixels pretos, encontrados em cada linha do retângulo envoltório, por sua largura.	
Divisão do Maior Contorno Vertical pela Altura	É o valor da divisão do maior contorno vertical, encontrado dentro do retângulo envoltório, por sua altura.	
Divisão da Soma dos Comprimentos dos Contornos Verticais pela Área	É a divisão da soma dos comprimentos de todos os contornos verticais, encontrados dentro do retângulo envoltório, por sua área.	

4.6 Classificação do sistema

A etapa de classificação é, sem dúvida, a mais importante, pois executa a tarefa esperada pelo sistema implementado, ou seja, determina se uma palavra no texto é impressa ou manuscrita. Na realidade, essa classificação é executada aqui sobre os retângulos envoltórios. Como na imagem, cada palavra é limitada por um retângulo envoltório, esse é classificado como contendo uma palavra impressa ou manuscrita.

A classificação de um retângulo envoltório qualquer é efetuada por meio de regras de classificação (Seção 2.6.4) mineradas na fase de mineração dos dados (Seção 2.6.2) e baseadas nas características extraídas da imagem (Seção 4.5). Se as características de um retângulo envoltório satisfazem à conjunção de condições do antecedente de uma regra, em que sua conclusão é um manuscrito, então ele é classificado como contendo um manuscrito.

Como exemplo, considere a regra de classificação: SE {(*Divisão do Maior Contorno Vertical pela Altura* <= 0.422) e (*Variância da Projeção Vertical* <= 99)} ENTÃO {*classe = manuscrito*}. O retângulo envoltório cujas características satisfazem ao antecedente

dessa regra, e que não tenha sido classificado anteriormente por outra regra, será classificado como contendo uma palavra manuscrita. Então, se estabelece que nessa região da imagem existe uma palavra manuscrita. Por outro lado, se as características do retângulo envoltório satisfazem ao antecedente da regra, por exemplo: SE {(*Divisão do Maior Contorno Vertical pela Altura* > 0.422) e (*Maior Diferença Encontrada na Projeção Horizontal* > 27) e (*Distribuição de Pixels* <= 594)} ENTÃO {*classe = impresso*}, ele será classificado como contendo uma palavra impressa, caso ainda não tenha sido classificada por outra regra.

O conjunto de regras é formado por regras que cobrem todos os casos possíveis referidos às características e, por isso, no final não existirá retângulo envoltório sem classificação no formulário.

5. TREINAMENTO, TESTES E RESULTADOS

Este capítulo, aborda inicialmente o modo de executar o treinamento do sistema, a fase de aprendizagem, onde as regras de classificação são obtidas. Em seguida, descreve as características das imagens dos bancos de dados utilizados para treinamentos e testes e por último, discute os resultados obtidos nos testes.

5.1 Treinamento do sistema

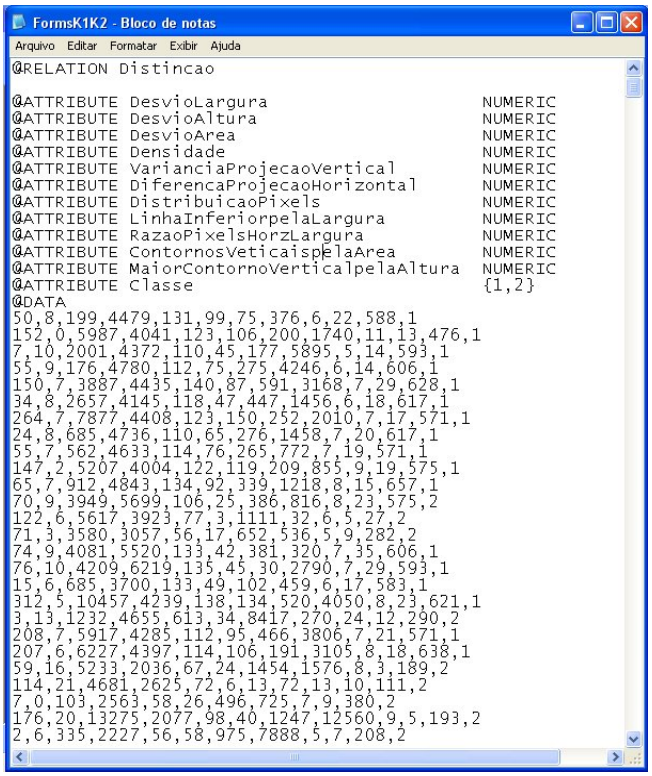
A abordagem proposta realiza uma tarefa de classificação (Seção 2.6.4). Na fase de mineração dos dados, gera um modelo de classificação preditivo de aprendizagem supervisionada. Isso quer dizer que, sendo supervisionada, necessita de treinamento com dados pré-classificados por um especialista da área (Seção 2.6.2).

Esses dados previamente classificados são retângulos envoltórios de formulários separados para o treinamento do sistema e, por isso, já estão classificados como contendo uma palavra impressa ou manuscrita.

Cada imagem desse conjunto de treinamento é carregada pelo sistema e passa pelas etapas descritas no Capítulo 3: pré-processamento, segmentação do texto em palavras e extração das características, porém não passa pela etapa de classificação. Essa, é feita manualmente pelo especialista.

As características e a classificação dos retângulos envoltórios, classificados manualmente, são armazenadas em um arquivo de extensão ARFF [ARFF], que é um arquivo de entrada para a ferramenta WEKA (Seção 1.3) [WEKA, 1999]. Ele é composto de duas partes distintas. A primeira parte, o cabeçalho, contém o nome da relação, uma lista de atributos e seus tipos. A segunda parte contém os dados que serão minerados logo após a declaração “@DATA”, Figura 5.1.

Na Figura 5.1, depois da declaração “@DATA”, a primeira linha contém os valores: “50,8,199,4479,131,99,75,376,6,22,588,1”, referente ao primeiro retângulo envoltório encontrado na imagem de um dos formulários do conjunto de treinamento. Nesse caso, o retângulo envoltório contém uma palavra impressa, pois termina com o valor 1. Os números: 50,8,199,...,588 são os valores das características (ou atributos) desse retângulo envoltório e que aparecem no cabeçalho do arquivo ARFF nessa ordem, ou seja, 50 é o desvio da largura, 8 é desvio da altura, 199 é o desvio da área e assim sucessivamente. Entretanto, a décima segunda linha, “70,9,3949,5699,106,25,386,816,8,23,575,2”, refere-se a um retângulo envoltório contendo uma palavra manuscrita, pois termina com o valor 2. Lembrando que, como esses retângulos envoltórios estão sendo usados para o treinamento do sistema, cada linha de valores termina com o valor 1 ou 2, preenchidos no momento da classificação manual.


FormsK1K2 - Bloco de notas

Arquivo Editar Formatar Exibir Ajuda

```
@RELATION Distincao

@ATTRIBUTE DesvioLargura                    NUMERIC
@ATTRIBUTE DesvioAltura                     NUMERIC
@ATTRIBUTE DesvioArea                       NUMERIC
@ATTRIBUTE Densidade                        NUMERIC
@ATTRIBUTE VarianciaProjecaoVertical        NUMERIC
@ATTRIBUTE DiferencaProjecaoHorizontal      NUMERIC
@ATTRIBUTE DistribuicaoPixels               NUMERIC
@ATTRIBUTE LinhaInferiorpelaLargura         NUMERIC
@ATTRIBUTE RazaoPixelsHorzLargura           NUMERIC
@ATTRIBUTE ContornosVeticaispelaArea        NUMERIC
@ATTRIBUTE MaiorContornoVerticalpelaAltura  NUMERIC
@ATTRIBUTE Classe                           {1,2}
@DATA
50,8,199,4479,131,99,75,376,6,22,588,1
152,0,5987,4041,123,106,200,1740,11,13,476,1
7,10,2001,4372,110,45,177,5895,5,14,593,1
55,9,176,4780,112,75,275,4246,6,14,606,1
150,7,3887,4435,140,87,591,3168,7,29,628,1
34,8,2657,4145,118,47,447,1456,6,18,617,1
264,7,7877,4408,123,150,252,2010,7,17,571,1
24,8,685,4736,110,65,276,1458,7,20,617,1
55,7,562,4633,114,76,265,772,7,19,571,1
147,2,5207,4004,122,119,209,855,9,19,575,1
65,7,912,4843,134,92,339,1218,8,15,657,1
70,9,3949,5699,106,25,386,816,8,23,575,2
122,6,5617,3923,77,3,1111,32,6,5,27,2
71,3,3580,3057,56,17,652,536,5,9,282,2
74,9,4081,5520,133,42,381,320,7,35,606,1
76,10,4209,6219,135,45,30,2790,7,29,593,1
15,6,685,3700,133,49,102,459,6,17,583,1
312,5,10457,4239,138,134,520,4050,8,23,621,1
3,13,1232,4655,613,34,8417,270,24,12,290,2
208,7,5917,4285,112,95,466,3806,7,21,571,1
207,6,6227,4397,114,106,191,3105,8,18,638,1
59,16,5233,2036,67,24,1454,1576,8,3,189,2
114,21,4681,2625,72,6,13,72,13,10,111,2
7,0,103,2563,58,26,496,725,7,9,380,2
176,20,13275,2077,98,40,1247,12560,9,5,193,2
2,6,335,2227,56,58,975,7888,5,7,208,2
```

Figura 5.1. Exemplo de arquivo ARFF.

No final, quando todas as imagens do conjunto de treinamento já tiverem sido usadas pelo sistema, se obtém um arquivo ARFF possuindo os valores das características e da classe de um número grande de retângulos envoltórios contendo manuscritos e impressos. Esse arquivo é utilizado pela ferramenta WEKA que irá minerar, dessas características calculadas (dados), as regras de classificação. Na metodologia, foi usada a versão 3.4 da ferramenta WEKA, executando-se o algoritmo PART [Cao & Wu, 2004] para obtenção das regras.

As regras de classificação mineradas são incorporadas ao sistema e utilizadas para a classificação de novos dados (retângulos envoltórios) fora do conjunto de treinamento. A Tabela 5.1 contém as regras mineradas de um determinado conjunto de treinamento.

Tabela 5.1. Regras mineradas de um determinado conjunto de treinamento

Antecedente (conjunção de condições)	Consequente (Classe)
SE{(*Divisão do Maior Contorno Vertical pela Altura* <= 0.422) e (*Variância da Projeção Vertical* <= 99)}	ENTÃO {*manuscrito*}
SE {(*Divisão do Maior Contorno Vertical pela Altura* > 0.422) e (*Maior Diferença Encontrada na Projeção Horizontal* > 27) e (*Distribuição de Pixels* <= 594)}	ENTÃO {*impresso*}
SE{(*Maior Diferença Encontrada na Projeção Horizontal* <= 20)}	ENTÃO {*manuscrito*}
SE{(*Divisão do Maior Contorno Vertical pela Altura* > 0.237) e (*Distribuição de Pixels* <= 934) e (*Densidade*> 3125) e (*Densidade* <= 5669)}	ENTÃO {*impresso*}
SE{(*Divisão do Maior Contorno Vertical pela Altura* <= 0.558) e (*Divisões de Pixels de Cada Linha pela Largura* > 8)}	ENTÃO {*manuscrito*}
SE{(*Densidade* <= 5703)}	ENTÃO {*manuscrito*}

O conjunto de características mais significativas para a tarefa de classificação sofre mudanças de acordo com a base usada para treinamento. Ao executar o avaliador de características ***CfsSubsetEval*** com o método de pesquisa ***GeneticSearch*** na ferramenta

WEKA, foi possível constatar esse fato. A Tabela 5.2 mostra as características mais significativas na base de imagem de formulários cadastrais, criada neste trabalho.

Tabela 5.2. Características mais significativas na classificação

Características
• Densidade • Variância da Projeção Vertical • Maior Diferença Encontrada na Projeção Horizontal • Distribuição de Pixels • Divisão da Soma dos Comprimentos dos Contornos Verticais pela Área • Divisão do Maior Contorno Vertical pela Altura

5.2 Bases de imagens utilizadas para testes

Alguns testes foram realizados utilizando bases contendo imagens de formulários. Uma dessas bases, a IAM-Database versão 3.0, é descrita no trabalho de Marti e Bunke [2002]. Uma versão mais antiga dessa base de dados e métodos de segmentação e reconhecimento de manuscrito são apresentados nos trabalhos de Bunke e seus coautores [Marti & Bunke, 1999], [Marti & Bunke, 2000], [Zimmermann & Bunke, 2002].

Essa base é formada por 1539 imagens de formulários, como o exibido pela Figura 4.1 na Seção 4.1. Cada formulário é constituído de quatro regiões distintas. Na região superior encontram-se o título “Sentence Database” e a identificação do mesmo (Figura 5.2). Em uma região logo abaixo da primeira, encontra-se o texto impresso (Figura 5.3). Mais abaixo, a região com o texto manuscrito (Figura 5.4). E por último, na parte inferior, o nome do escritor da região manuscrita quando esse se identificou (Figura 5.5). Cada região está separada uma da outra por linhas horizontais.

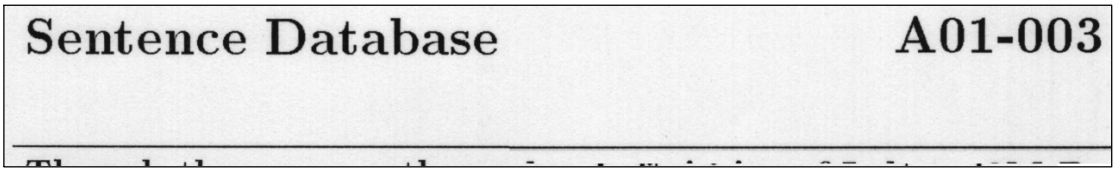


Figura 5.2. Título e identificação de formulário da base AIM.

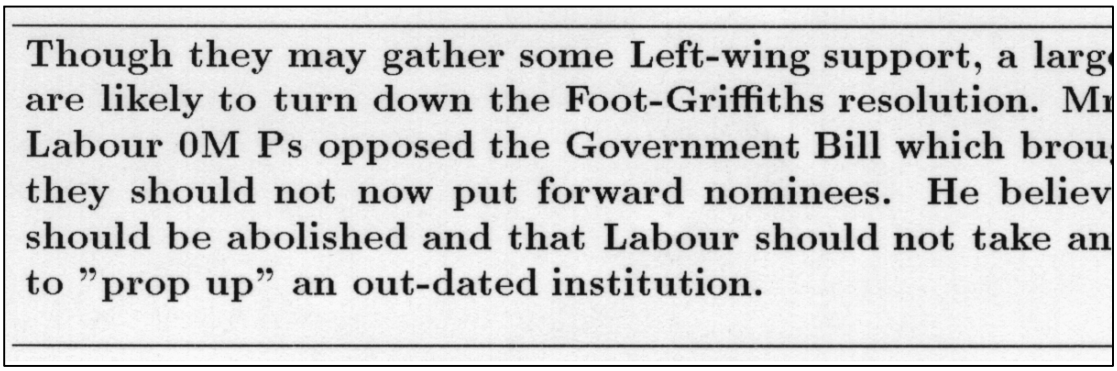


Figura 5.3. Porção de texto impresso de formulário da base AIM.

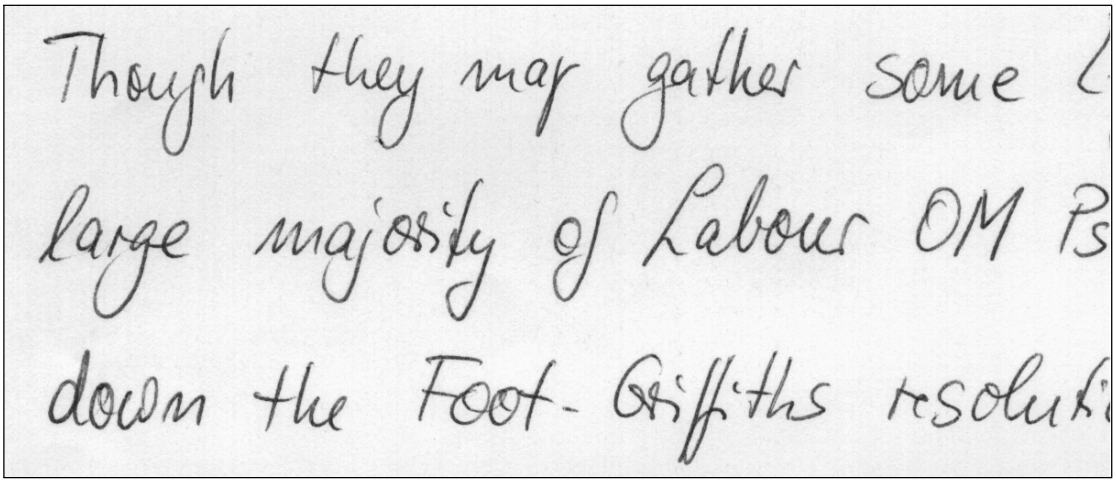


Figura 5.4. Porção de texto manuscrito de formulário da base AIM.

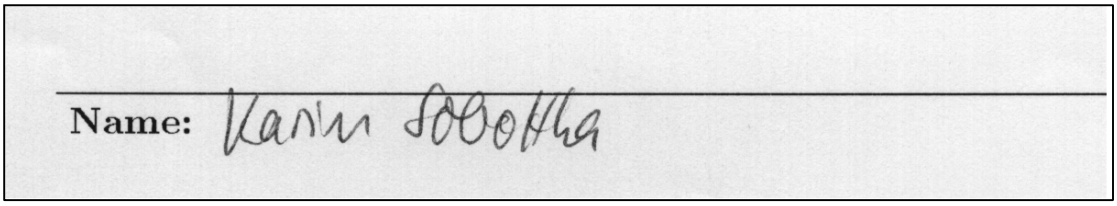


Figura 5.5. Identificação do escritor do formulário na base AIM.

Segundo os autores, essa base foi criada com o objetivo de avaliar metodologias de reconhecimento de textos manuscritos e metodologias de identificação de seus autores. Por isso, nela não existe o *ground truth* em relação à distinção de palavras manuscritas e impressas. Entretanto, a localização dessas palavras, separadas no texto, dispensa a existência do *ground truth* para a avaliação do resultado do teste realizado nessa base. Basta utilizar o posicionamento vertical em relação à página, visto que todas as impressas estão na parte superior e as manuscritas na parte inferior. Para efeito de teste, a região na Figura 5.5, geralmente, não é considerada por apresentar uma palavra impressa e palavras manuscritas na mesma linha de texto.

A região impressa contém fragmentos de 500 textos em inglês com assuntos variados: reportagens, religião, sabedoria popular, ficção científica, romances e humor, produzindo um vocabulário de 10841 palavras. Cada fragmento de texto, com no máximo 50 palavras, foi impresso em um formulário com a estrutura já descrita, e 657 pessoas foram convidadas a copiar, na região em branco, o texto impresso. Assim, foram gerados os 1539 formulários que tiveram suas imagens capturadas com uma resolução de 300dpi em uma definição de escala de cinza de 8 bits conforme mostrado na Figura 4.1 da Seção 4.1.

Essa base foi usada para treinamento, conforme descrito na Seção 5.1, e para testes, gerando os resultados apresentados na Seção 5.3.

Outra base utilizada foi a construída durante a presente dissertação. Diferentemente dos formulários da primeira base, os formulários dessa apresentam palavras impressas e manuscritas em uma mesma linha de texto. A Figura 4.2 da Seção 4.1 mostra um desses formulários. Outra diferença, são as linhas horizontais usadas como base para o preenchimento dos dados manuscritos (Figura 5.6).

Figura 5.6. Linha base para manuscritos.

Essa base é constituída de 121 imagens de formulários preenchidos por 121 pessoas com diferenças de escolaridade, idade e posição social. Um modelo de formulário foi montado e reproduzido em quantidade e essas pessoas foram convidadas para, voluntariamente, preencherem os formulários com dados não necessariamente verdadeiros, porém coerentes. Os formulários preenchidos tiveram suas imagens capturadas em um scanner HP Deskjet F380, com uma resolução de 300dpi em uma definição de escala de cinza de 8 bits no formato BMP. A idéia de não exigir dados pessoais verdadeiros das pessoas decorre do fato de que elas não aceitariam preencher os formulários, caso fossem obrigadas a tal exigência. Mas isso não influenciou no objetivo da tarefa desta dissertação.

A falta de uma base de imagens disponível para pesquisas, de formulários típicos, foi a razão para a criação dessa base de dados. Uma base de formulários contendo, por exemplo: nome, endereço, CPF ou outros dados a serem preenchidos e apresentando palavras impressas e manuscritas em uma mesma linha do texto, onde as manuscritas são preenchidas sobre linhas horizontais base.

Para cada imagem dessa base, foi criado um arquivo de texto contendo sua classificação correta (seu *ground truth*) em relação a palavras impressas e manuscritas. Em cada arquivo (Figura 5.7), cada linha contém as coordenadas dos limites inferior, superior, esquerdo e direito de cada retângulo envoltório formado na imagem. O último valor de cada linha é a classe do conteúdo do retângulo envoltório, ou seja, quando igual a 1 significa que o conteúdo é uma palavra impressa e quando igual a 2 significa que o conteúdo é uma manuscrita.

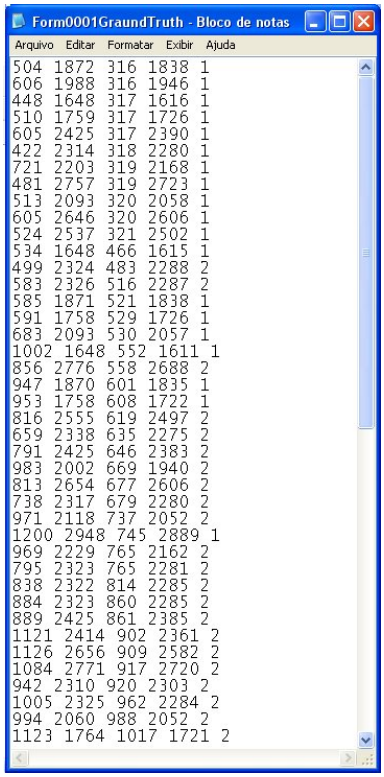


Figura 5.7. Exemplo de arquivo *ground truth*.

5.3 Resultados

Para validação da avaliação dos resultados dos testes realizados nas bases de imagens, foi aplicado o método *K-fold Cross Validation*, abordado na Seção 2.3.6. Os avaliadores da classificação utilizados foram: a acurácia, a precisão, o verdadeiro positivo, o falso positivo, o verdadeiro negativo, o falso negativo, a sensibilidade, a especificidade e o desvio padrão, descritos na Seção 2.3.5.

Em relação à base de dados AIM Database 3.0, 20 imagens foram escolhidas aleatoriamente para formarem os 10 subconjuntos exigidos pelo método *K-fold Cross Validation*, quando o parâmetro *K* é igual a 10. Logo, cada conjunto ficou formado por duas dessas imagens (formulários). Os resultados alcançados estão na Tabela 5.3.

Tabela 5.3. Resultados do teste na base de dados AIM-DB v.3.

	Palavras impressas	**Palavras manuscritas**
Total	1404	2029
Acurácia	97,51%	97,54%
Precisão	96,48%	98,26%
Verdadeiro positivo	1369	1979
Falso positivo	50	35
Verdadeiro negativo	1979	1369
Falso negativo	35	50
Sensibilidade	0,9751	0,9754
Especificidade	0,9754	0,9751
Média das acurácias	97,55%	98,09%
Média das precisões	96,70%	98,10%
Desvio padrão das acurácias	1,617546	2,033507
Desvio padrão das precisões	4,414524	1,385656
Acurácia mínima	91,18%	91,01%
Precisão mínima	81,82%	93,85%
Range das acurácias	8,82%	8,99%
Range das precisões	18,18%	6,15%

Na Tabela 5.3, as linhas "Média das acurácias/precisões" das palavras impressas/manuscritas contêm a média das acurácias/precisões dos dez subconjuntos formados para a aplicação do método *K-fold Cross Validation.* O desvio padrão das acurácias/precisões também é calculado em relação às acurácias e precisões dos dez subconjuntos. O sistema alcançou acurácia e precisão de 100% em 45% das imagens nessa base. Uma tabela do resultado completo do teste encontra-se no apêndice A deste texto.

Em relação à base de imagens de formulários cadastrais (Figura 4.2) construída neste trabalho, 24 imagens foram escolhidas casualmente para formarem os 3 subconjuntos exigidos pelo método *K-fold Cross Validation*, quando o parâmetro *K* é igual a 3. Desse modo, cada conjunto ficou formado por 8 dessas imagens (formulários). Os resultados alcançados estão nas Tabelas 5.4.

Tabela 5.4. Resultados do teste na base de imagens de formulários de cadastro.

	Palavras impressas	**Palavras manuscritas**
Total	600	1329
Acurácia	97,17%	99,47%
Precisão	98,81%	98,73%
Verdadeiro positivo	583	1322
Falso positivo	7	17
Verdadeiro negativo	1322	583
Falso negativo	17	7
Sensibilidade	0,9717	0,9947
Especificidade	0,9947	0,9717
Média das acurácias	97,17%	99,46%
Média das precisões	98,85%	98,75%
Desvio padrão das acurácias	1,892969	0,76153
Desvio padrão das precisões	1,580431	0,801327
Acurácia mínima	88,00%	96,43%
Precisão mínima	92,59%	95,35%
Range das acurácias	12,00%	3,57%
Range das precisões	7,41%	4,65%

As mesmas observações feitas para a Tabela 5.3 em relação às linhas "Média das acurácias/precisões" e "Desvio padrão das acurácias/precisões" são válidas para a Tabela 5.4. Assim como para a primeira base, uma tabela com o resultado completo do teste encontra-se no apêndice A deste texto.

Nas imagens dos documentos analisados, existem apenas dois conjuntos de palavras, o das impressas e o das manuscritas. Como esses conjuntos são complementares um do outro em relação ao total de palavras no texto, consequentemente, os pares de valores verdadeiro/falso positivo de impressos e verdadeiro/falso negativo de manuscritos são iguais, como também os pares verdadeiro/falso positivo de manuscritos e verdadeiro/falso negativo de impressos.

Utilizando um computador equipado com o processador AMD Athlon™ MP 900Mhz, o sistema executou, aproximadamente, 184,04 BI (bilhões de instruções) para cada imagem processada. BI é um conceito de avaliação de desempenho descrito no trabalho de

Brazil [2008]. O valor de 184,04 é obtido do seguinte cálculo: (74*2487)/1000, onde 74 é o tempo médio, em segundos, consumido no processamento da imagem e 2487 é a quantidade de instruções em milhões por segundo a qual o processador é capaz de executar, chamada de MIPS (milhões de instruções por segundo), obtida após a aplicação do teste Processor Arithmetic do software Sandra Lite [Sandra].

Nos testes, a classificação do sistema foi apurada por diversos avaliadores e em todos eles o resultado foi satisfatório. É possível notar que nas duas bases de imagens a taxa de acertos e a precisão foram significativas. Na classificação de palavras impressas, o melhor desempenho do sistema foi na base IAM Database 3.0 e na de manuscritas foi na base de formulários. As conclusões referentes aos resultados estão no próximo capítulo.

6. CONCLUSÕES

Este trabalho apresenta uma metodologia desenvolvida e implementada para a distinção de palavras impressas e manuscritas em uma imagem de documento. E para comprovar a eficiência dessa metodologia, o sistema desenvolvido foi testado em duas bases de imagens, aplicando a técnica *K-fold Cross Validation*. Em uma das bases, com *K=10*, os resultados apurados foram: 97,55% de acurácia média e 96,70% de precisão média na classificação de palavras impressas; 98,09% de acurácia média e 98,10% de precisão média na classificação de palavras manuscritas; 8,82% de range em relação à acurácia e 18,18% em relação à precisão na classificação de palavras impressas; 8,99% de range em relação à acurácia e 6,15% em relação à precisão na classificação de palavras manuscritas; desvio padrão de 1,617546 e de 2,033507 no conjunto das acurácias dos subconjuntos de teste na classificação de palavras impressas e manuscritas, respectivamente; desvio padrão de 4,414524 e de 1,385656 no conjunto das precisões dos subconjuntos de teste na classificação de palavras impressas e manuscritas, nessa ordem; sensibilidade de 0,9751 e 0,9754 na classificação de palavras impressas e manuscritas, respectivamente; especificidade de 0,9754 e 0,9751 na classificação de palavras impressas e manuscritas, nessa ordem. Na segunda base, com *K*=3, os resultados obtidos foram: 97,17% de acurácia média e 98,85% de precisão média na classificação de palavras impressas; 99,46% de acurácia média e 98,75% de precisão média na classificação de palavras manuscritas; 12,00% de range em relação à acurácia e 7,41% em relação à precisão na classificação de palavras impressas; 3,57% de range em relação à acurácia e 4,65% em relação à precisão na classificação de palavras manuscritas; desvio padrão de 1,892969 e de 0,76153 no conjunto das acurácias dos subconjuntos de teste na classificação de palavras impressas e manuscritas, respectivamente; desvio padrão de 1,580431 e de 0,801327 no conjunto das precisões dos subconjuntos de teste

na classificação de palavras impressas e manuscritas, nessa ordem; sensibilidade de 0,9717 e 0,9947 na classificação de palavras impressas e manuscritas, respectivamente; especificidade de 0,9947 e 0,9717 na classificação de palavras impressas e manuscritas, nessa ordem.

Os testes foram realizados em um computador equipado com o processador AMD Athlon™ MP 900Mhz o qual executou, aproximadamente, 184,04 BI (bilhões de instruções) e consumiu 74 segundo, em média, para cada imagem processada.

Outra realização deste trabalho, foi a criação de uma das bases de imagens utilizada para teste. Essa base é constituída de 121 imagens de formulários preenchidos por 121 pessoas de diferentes escolaridades, idades e posições sociais. Para cada uma de suas imagens, foi criado um arquivo de texto contendo seu *ground truth* em relação à classificação das palavras impressas e manuscritas. A base encontra-se disponível na página do trabalho no endereço: http://visual.ic.uff.br/analisededocumentos/pt/index.htm, onde é possível obter também as instruções e termo de uso da mesma.

No Capítulo 3, são descritas várias metodologias de trabalhos anteriores para a distinção de palavras impressas e manuscritas. Algumas dessas, são sensíveis a qualquer inclinação do texto ou são fortemente dependentes do leiaute da página ou das características das letras da escrita em análise. Outras, não admitem que uma mesma linha no documento contenha texto impresso e manuscrito, pois classificam por linha. Algumas outras, utilizam técnicas complexas de classificação e, mesmo assim, suas taxas de acerto são baixas.

Por aplicar a técnica de extração de componentes conectados para segmentar o texto e por realizar a classificação ao nível de palavras, o sistema, aqui desenvolvido, é capaz de tolerar certo grau de inclinação do texto. As características extraídas para descrever impressos e manuscritos são locais, isso significa independência do leiaute do documento para a classificação. Uma mesma linha de texto pode conter palavras impressas e manuscritas, pois a classificação é por palavras. E, utilizando somente regras de classificação, o sistema obteve resultados satisfatórios, já citados no início do capítulo.

De acordo com as tabelas do apêndice A, que mostram os resultados completos dos testes, em uma das bases, o sistema alcançou 100% de acertos em 45% das imagens, com acurácia mínima de 91,18% e precisão mínima de 81,82% em relação às palavras impressas e acurácia mínima de 91,01% e precisão mínima de 93,85% em relação às manuscritas. Na outra base, o sistema alcançou 100% de acertos em 33,33% das imagens, com acurácia mínima de 88,00% e precisão mínima de 92,59% em relação às palavras impressas e acurácia mínima de 96,43% e precisão mínima de 95,35% em relação às manuscritas.

O trabalho de Kavallieratou et al. [2004], citado no Capítulo 3, também utiliza a base de imagens IAM-DB 3.0 para testes. Nesse trabalho, a técnica *K-fold Cross Validation* também é utilizada para validação dos resultados com dez subconjuntos. A acurácia alcançada, segundo os autores, foi de 97,9% no conjunto de teste, porém no artigo não é informado se esse valor é da máxima ou da mínima ou da média. Nessa metodologia, a classificação é feita em relação às linhas do texto, ou seja, o texto é segmentado em nível de linhas e não em nível de palavras e, por isso, os erros foram mais frequentes quando somente uma palavra estava presente nas linhas. Outro ponto negativo é que palavras impressas e manuscritas não podem coexistir em uma mesma linha devido ao tipo de classificação.

Comparando a metodologia proposta com a metodologia do parágrafo anterior, é possível observar uma vantagem em relação à classificação. Aqui ela é feita por palavras e com isso, como já comentado, uma mesma linha pode conter palavras impressas e manuscritas. Em relação à acurácia, não é possível fazer uma comparação precisa, pois o valor de 97,9% no trabalho de Kavallieratou et al. não é devidamente especificado. Contudo, em 45% das imagens da base AIM-DB 3.0 a metodologia proposta alcançou 100% de acurácia e precisão na classificação das palavras impressas e manuscritas. E, considerando todas as imagens da base, a acurácia mínima foi de 91,01%, em relação à classificação das palavras manuscritas em uma determinada imagem, e a precisão mínima foi de 81,82%, em

relação à classificação das palavras impressas em uma outra determinada imagem (Apêndice A).

De acordo com os resultados dos testes, é possível concluir que o sistema desenvolvido na metodologia proposta é confiável para distinguir texto impresso e manuscrito em uma imagem de documento. Por suas características de segmentação e classificação do texto, ele é aplicável à maioria dos tipos de formulários preenchíveis. E a base de imagens construída e disponível é uma contribuição para o desenvolvimento de trabalhos futuros, inclusive de outros pesquisadores.

6.1 Trabalhos futuros

A distinção de texto impresso e manuscrito em uma imagem de documento pode ser realizada de diversas maneiras, variando, independentemente ou em conjunto, as técnicas de pré-processamento aplicadas, a forma de segmentação executada, as características extraídas e as técnicas de classificação utilizadas. Uma continuação imediata deste trabalho seria a utilização de outras técnicas de classificação como redes neurais, classificadores de mínima distância, classificadores SVM (Support Vector Machines) e baseados na lógica fuzzy, para gerar comparações com a técnica atualmente aplicada. Além disso, novas características poderiam ser extraídas dos retângulos envoltórios explorando a regularidade das palavras impressas e a falta dessa nas manuscritas (como o desvio padrão das alturas e larguras dos componentes conectados no interior deles e a simetria dos *loops* desses componentes) e/ou utilizando descritores de textura (aplicando filtros de Gabor, histograma de co-ocorrência e *NXM grams* para análise de padrões). Uma segmentação híbrida do texto poderia ser executada, ou seja, uma segmentação parte top-down e parte bottom-up. A parte top-down segmentaria o texto em linhas, baseada na análise do perfil da projeção horizontal

dos pixels e a parte bottom-up segmentaria as linhas de texto em palavras, unindo componentes conectados, extraídos das mesmas.

Outras propostas para trabalhos futuros, não tão imediatas como as anteriores, poderiam ter como objetivo tornar o sistema capaz de realizar a distinção prevista em uma variedade maior de tipos de documentos. Uma delas seria capacitar o sistema para analisar textos que tenham outras orientações na página, não apenas a horizontal. Uma outra proposta seria tornar o sistema hábil em distinguir texto impresso e manuscrito dentre tabelas, figuras, gráficos e outros elementos, que se apresentam em alguns documentos. Com certa frequência, palavras manuscritas, como anotações, encontram-se no texto impresso do documento. Nesse caso, a tarefa de distinção tem seu grau de dificuldade aumentado. Identificar essas palavras manuscritas também é uma das sugestões para trabalhos futuros.

Várias são as propostas, visto que os documentos apresentam uma variedade ampla de leiaute, onde cada um, como relatado nos trabalhos anteriores citados, determina o caminho percorrido até o objetivo final, a classificação. Então, deixar o sistema o menos especialista possível é uma tarefa desafiadora que exige muita pesquisa e esforço.

Referências Bibliográficas

Aires, S. B. K. [2005]. *"Reconhecimento de Caracteres Manuscritos Baseado em Regiões Perceptivas"*, Pontifícia Universidade Católica do Paraná, Dissertação de Mestrado.

ARFF. Documentação, acessado em 01 de outubro de 2008, disponível em: http://weka.sourceforge.net/wekadoc/index.php/en:ARFF_(3.4.6)#Overview.

Brazil, A. [2008]. *"Path Relinking and AES Cryptography in Color Image Steganography"*, Universidade Federal Fluminense, Instituto de Computação, Dissertação de Mestrado.

Calixto, E. P. [2005]. *"Granulometria morfológica em espaços de cores: Estudos da ordenação espacial"*, Universidade Federal Fluminense, Instituto de Computação, Dissertação de Mestrado.

Cao, Y.; Wu, J. [2004]. *"Dynamics of projective adaptive resonance theory model: the foundation of PART algorithm"*, IEEE Transactions on Neural Networks, v. 15, nº 2, pp. 245 – 260.

Carvalho, J. E. R. [2006]. *"Uma Abordagem de Segmentação de Placas de Automóveis Baseada em Morfologia Matemática"*, Universidade Federal Fluminense, Instituto de Computação, Dissertação de Mestrado.

Castanho, M. J. P; Yamakami, A; Barros, L. C; Vendite, L.L. [2004]. *"Avaliação de um teste em medicina usando uma curva ROC fuzzy"*, Biomatemática, v. 14, pp. 19 – 28.

Conci, A.; Azevedo, E; Leta, F. R. [2008]. *"Computação Gráfica Teoria e Prática"*, vol. 2, ed. Campus.

Cormen, T. H.; Leiserson, C.E.; Rivest, R. L.; Stein, C. [2002]. *"Algoritmos: Teoria e Prática"*, ed. Campus.

DevC++. Documentação, acessado em: 8 de julho de 2008, disponível em: http://wxdsgn.sourceforge.net/

Facon, J. [1996]. *"Morfologia Matemática: Teoria e Exemplos"*, ed. Universitária Champagnhat, PUCPR.

Fan, K. C.; Wang, L. S.; Tu, Y. T. [1998]. *"Classification of Machine-Printed and Handwritten Texts Using Character Block Layout Variance"*, Pattern Recognition, v. 31, nº 9, pp. 1275 – 1284.

Farooq, F.; Sridharan, K.; Govindaraju, V. [2006]. *" Identifying Handwritten Text in Mixed Documents"*, ICPR 2006, 18th International Conference on Pattern Recognition, v. 2, pp. 1142 – 1145.

Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. [1996]. *"From Data Mining to Knowledge Discovery in Databases"*, AI Magazine, v. 17, nº 3, pp. 37–54.

Franke, J.; Oberlander, M. [1993]. *"Writing Style Detection by Statistical Combination of Classifiers in Form Reader Applications"*, Proceedings of the Second International Conference on Document Analysis and Recognition, 20–22 Oct., pp. 581 – 584.

GCC. [1988] Documentação, acessado em: 5 de setembro de 2008, disponível em: http://gcc.gnu.org/

Gllavata, J.; Ewerth, R.; Freisleben, B. [2004]. *"A text detection, localization and segmentation system for OCR in images"*, Proceedings. IEEE Sixth International Symposium on Multimedia Software Engineering, 13–15 dez., pp. 310 – 317.

Gonçalves, E. C. [2005]. *"Regras de Associação e suas Medidas de Interesse Objetivas e Subjetivas"*, Infocomp, Journal of Computer Science, v. 4, nº1, pp. 26 – 35.

Gonzalez, R. C.; Woods, R. E. [1992]. *"Processamento de Imagens Digitais"*, ed. Edgard Blücher, reimpressão: 2003, em português.

Govindan, V. K.; Shivaprasad, A. P. [1990]. *"Character recognition – a review"*, Pattern Recognition, v. 23, pp. 671–683.

Guo, J. K.; Ma, M.Y. [2001]. *"Separating Handwritten Material from Machine Printed Text Using Hidden Markov Models"*, Proceedings. Sixth International Conference on Document Analysis and Recognition, 10–13 Sept., pp. 439 – 443.

Han, J. e Kamber, M. [2001]. *"Data mining: Concept and Techniques"*, ed. Morgan Kaufmann.

Imade, S.; Tatsuta, S.; Wada, T. [1993]. *"Segmentation and Classification for Mixed Text/Image Documents Using Neural Network"*, Proceedings of the Second International Conference on Document Analysis and Recognition, 20–22 Oct., pp. 930 – 934.

Impedovo, S.; Ottaviano L; Occhinegro, S. [1992]. *"Optical character recognition – a survey"*, Int. J. Pattern Recognition And Artificial Intelligence, v. 5, pp. 1–24.

Kapp, M. N., [2004]. *"Reconhecimento de Palavras Manuscritas Utilizando Redes Neurais Artificiais"*, Pontifícia Universidade Católica do Paraná, Dissertação de Mestrado.

Kavallieratou, E.; Stamatatos, S. [2004]. *"Discrimination of Machine-Printed from Handwritten Text Using Simple Structural Characteristics"*, Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, v. 1, 23 – 26 Aug., pp.437 – 440.

Kavallieratou, E.; Stamatatos, S.; Antonopoulou, H. [2004]. *"Machine-Printed from Handwritten Text Discrimination"*, IWFHR-9 2004, Ninth International Workshop on Frontiers in Handwriting Recognition, 26 – 29 Oct., pp. 312 – 316.

Kim, M.-J.; Han, I. [2003]. *"The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms"*, Expert Systems with Applications, v. 25, nº 4, pp. 637– 646.

Koerich, A. L. [2004]. **"***Handwritten Word Recognition Using Markov Models"*, Latin America Transactions, IEEE (Revista IEEE America Latina), v. 2, nº 2, pp. 132 – 141.

Koyama, J.; Kato, M.; Hirose, A. [2008-a]. *"Local-spectrum-based distinction between handwritten and machine-printed characters"* 15th IEEE International Conference on Image Processing, 12–15 Oct., pp. 1021 – 1024.

Koyama, J.; Kato, M.; Hirose, A. [2008-b]. "*Distinction between handwritten and machine-printed characters with no need to locate character or text line position"*, IEEE International Joint Conference on Neural Networks – IJCNN 2008. (IEEE World Congress on Computational Intelligence), 1–8 June, pp. 4044 – 4051.

Kuhnke, K.; Simoncini, L.; Kovacs-V, Z. M. [1995]. *"A System for Machine-Written and Hand-Written Character Distinction"*, Proceedings of the Third International Conference on Document Analysis and Recognition, v. 2, 14 – 16 Aug., pp 811 – 814.

L'Homer, E. [2000]. *"Extraction of strokes in handwritten characters Pattern Recognition"*, v. 33, nº 7, pp. 1147-1160.

Mahadevan, U.; Nagabushnam, R.C. [1995].*"Gap metrics for word separation in handwritten lines"*, Proceedings of the Third International Conference on Document Analysis and Recognition, v. 1, 14–16 Aug., pp. 124 –127.

Manmatha, R.; Rothfeder, J.L. [2005]. *"Scale space approach for automatically segmenting words from historical handwritten documents"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 27, nº 8, pp. 1212 – 1225.

Marti, U.; Bunke, H. [1999].*"A full English sentence database for off-line handwriting recognition"*, ICDAR '99, Proceedings of the Fifth International Conference on Document Analysis and Recognition, 20–22 Sept., pp. 705–708.

Marti, U.; Bunke, H. [2000]. "*Handwritten Sentence Recognition"*, Proceedings. 15th International Conference on Pattern Recognition, v. 3, 3–7 Set., pp. 463–466.

Marti, U.; Bunke, H. [2002]. *"The IAM-database: an English Sentence Database for Off-line Handwriting Recognition"*, Int. Journal on Document Analysis and Recognition, v. 5, nº1, pp. 39–46.

Mital, D. P.; Leng, G. W. [1996]. **"***Text segmentation for automatic document processing"*, Proceedings., 1996 IEEE Conference on Emerging Technologies and Factory Automation, EFTA '96, v. 2, 18–21 Nov., pp. 642 – 648.

Monteiro, L. H. [2002]. *"Utilização de Técnicas de Processamento de Imagens para o Reconhecimento de Placas de Veículos",* Universidade Federal Fluminense, Instituto de Computação, Dissertação de Mestrado.

Nicolas, S.; Paquet, T.; Heutte, L. [2004]. *"Text line segmentation in handwritten document using a production system"*, IWFHR-9 2004. Ninth International Workshop on Frontiers in Handwriting Recognition, 26–29 Oct., pp. 245 – 250.

Nunes, É. O. [2006]. "*Segmentação por Textura em Imagens Multibandas"*, Universidade Federal Fluminense, Instituto de Computação, Tese de Doutorado.

Nuñez, C. C.; Conci, A. [2007]. "*A JavaScript tool to present Mathematical Morphology to beginner"*, ISMM 2007, Proceedings of the 8th International Symposium on Mathematical Morphology, 10–13 Oct., vol 2, pp. 75-76.

OpenGL. Documentação, acessado em: 9 de julho de 2008, disponível em: http://www.opengl.org/documentation/

Otsu, N. [1979]. *"A Threshold Selection Method from Gray-Level Histograms"* IEEE Transactions on Systems, Man and Cybernetics, v. 9, nº 1, pp. 62 – 66.

Pal, U.; Chaudhuri, B.B.; [1999]. *"Automatic separation of machine-printed and hand-written text lines",* ICDAR '99. Proceedings of the Fifth International Conference on Document Analysis and Recognition, 20–22 Sept., pp. 645–648.

Pal, U.; Chaudhuri, B. B. [2001]. *"Machine-printed and Hand-written Text Line Identification"*, Pattern Recognition Letters, v. 22, nº 3 – 4, pp. 431 – 441.

Piatetsky-Shapiro, G. [1990]. *"Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop"*, AI Magazine, v. 11, nº 5, pp. 68 – 70.

Pitas, I. [1995]. "Digital Image Processing Algorithms", ed. Prentice Hall.

Romão, W. [2002]. *"Descoberta de Conhecimento Relevante em Banco De Dados Sobre Ciência E Tecnologia"*, Universidade Federal de Santa Catarina, Dissertação de Mestrado.

Sampaio, M. C. ; Mongiovi, G. ; Carvalho, J. V. [1999]. "*Utilizando Técnicas de Data Mining para o Reconhecimento de Caracteres Manuscritos"*, XIV Simpósio Brasileiro de Banco de Dados, Florianópolis - SC., Anais – pp. 235–249.

Sandra. Acessado em: 20 de fevereiro de 2009, disponível em: http://www.sisoftware.co.uk/

Santos, J. E. B.; Dubuisson, B.; Bortolozzi, F. [2002-a]. *"A Non Contextual Approach for Textual Element Identification on Bank Cheque Images"*, IEEE International Conference on Systems, Man and Cybernetics, v. 4, 6 – 9 Oct., 6 pgs.

Santos, J. E. B.; Dubuisson, B.; Bortolozzi, F. [2002-b]. *"Characterizing and Distinguishing Text in Bank Cheque Images"*, Proceedings. XV Brazilian Symposium on Computer Graphics and Image Processing, 7 – 10 Oct., pp. 203 – 209.

Seixas, F. L.; Martins, A.; Stilben, A. R.; Madeira, D.; Assumpção, R.; Mansur, S.; Victer, S. M.; Mendes, V. B.; Conci, A. [2008]. *"Avaliação dos Métodos para a Segmentação Automática dos Tecidos do Encéfalo em Ressonância Magnética"*, XI Simpósio de

Pesquisa Operacional e Logística da Marinha, Escola de Guerra Naval, SPOLM 2008, 5–6 de agosto, pp. 53–54 (livro de resumos).

Silva, C. M. S. [2003]. *"Utilizando o Processo de Descoberta de Conhecimento em Bancos de Dados para Identificar Candidatos a Padrão de Análise para Bancos de Dados Geográficos"*, Universidade Federal do Rio Grande do Sul, Instituto de Informática, Dissertação de Mestrado.

Srihari, S. N.; Shin Y.C.; Ramanaprasad, V.; Lee, D. S. [1996]. *"A System to Read Names and Addresses on Tax Forms"*, Proceedings of the IEEE, v. 84, nº 7, pp. 1038 – 1049.

Vellasques, E. [2006] *"Classificação de Pontos de Segmentação de Dígitos Manuscritos"*, Pontifícia Universidade Católica do Paraná, Dissertação de Mestrado.

Violante, S.; Smith, R.; Reiss, M. [1995]. *"A Computationally Efficient Technique for Discriminating Between Hand-Written and Printed Text"*, IEEE Colloquium on Document Image Processing and Multimedia Environments, 2 Nov, pp. 17/1 – 17/7.

WEKA. [1999]. Documentação, acessado em: 26 de setembro de 2008, disponível em: http://www.cs.waikato.ac.nz/ml/weka/

Wikipedia. "*Acurácia e precisão"*, acessado em: 7 de novembro de 2008, disponível em: http://en.wikipedia.org/wiki/Accuracy_and_precision

Zheng, Y.; Li, H.; Doermann, D. [2002]. *"The Segmentation and Identification of Handwriting in Noisy Document Images"*, Document Analysis Systems V, Lecture Notes in Computer Science, v. 2423, pp. 95–105.

Zheng, Y.; Li, H.; Doermann, D. [2003]. *"Text Identification in Noisy Document Images Using Markov Random Field"*, Proceedings. Seventh International Conference on Document Analysis and Recognition, 3–6 Aug., v. 1, pp. 599 – 603.

Zheng, Y.; Li, H.; Doermann, D. [2004]. *"Machine Printed Text and Handwriting Identification in Noisy Document Images"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 26, nº 3, pp. 337 – 353.

Zimmermann, M.; Bunke, H. [2002]. *"Automatic Segmentation of the IAM Off-line Database for Handwritten English Text"*, Proceedings. 16th International Conference on Pattern Recognition, v. 4, pp. 35 – 39.

APÊNDICE A

A tabela abaixo mostra o resultado completo do teste feito na base de imagem AIM-DB 3.0 (Figura 3.1). Nele foi utilizada a técnica *k*-fold Cross Validation com 10 subconjuntos (cada um contendo duas imagens) e em 45% dos formulários o sistema alcançou 100% de acertos na classificação de impressos e manuscritos.

Imagem	Total de Impr.	Total de manus.	Impr. class. cor-reta-mente	Manus. class. cor-reta-mente	Class. como impr.	Class. como manus.	Acur. Impr. em %	Acur. manus. em %	Prec. Impr. em %	Prec. manus. em %
a01-3x	79	86	77	86	77	88	97,47	100,00	100,00	97,73
a01102u	68	105	62	100	67	106	91,18	95,24	92,54	94,34
	147	191	139	186	144	194	94,56	97,38	96,53	95,88
a01-132	66	123	64	117	70	119	96,97	95,12	91,43	98,32
a02-078	76	204	76	191	89	191	100,00	93,63	85,39	100,00
	142	327	140	308	159	310	98,59	94,19	88,05	99,35
a03-034	67	80	67	80	67	80	100,00	100,00	100,00	100,00
a04-089	66	115	65	114	66	115	98,48	99,13	98,48	99,13
	133	195	132	194	133	195	99,25	99,49	99,25	99,49
a05-004	72	178	72	162	88	162	100,00	91,01	81,82	100,00
a06-014	67	145	65	144	66	146	97,01	99,31	98,48	98,63
	139	323	137	306	154	308	98,56	94,74	88,96	99,35
b01-033	52	50	51	50	51	51	98,08	100,00	100,00	98,04
b02-105	39	55	38	54	39	55	97,44	98,18	97,44	98,18
	91	105	89	104	90	106	97,80	99,05	98,89	98,11
b03-109	82	94	77	93	78	98	93,90	98,94	98,72	94,90
b04-116	74	86	72	86	72	88	97,30	100,00	100,00	97,73
	156	180	149	179	150	186	95,51	99,44	99,33	96,24
b05-042	93	106	91	106	91	108	97,85	100,00	100,00	98,15
b06-012	54	63	50	61	52	65	92,59	96,83	96,15	93,85
	147	169	141	167	143	173	95,92	98,82	98,60	96,53
b06-079	75	89	75	89	75	89	100,00	100,00	100,00	100,00
c01-014	91	108	89	106	91	108	97,80	98,15	97,80	98,15
	166	197	164	195	166	197	98,80	98,98	98,80	98,98
c02-056	73	89	73	89	73	89	100,00	100,00	100,00	100,00
c03-94d	68	79	66	79	66	81	97,06	100,00	100,00	97,53
	141	168	139	168	139	170	98,58	100,00	100,00	98,82
c04-150	74	90	71	88	73	91	95,95	97,78	97,26	96,70
c06-005	68	84	68	84	68	84	100,00	100,00	100,00	100,00
	142	174	139	172	141	175	97,89	98,85	98,58	98,29
Totais	1404	2029	1369	1979	1419	2014	97,51	97,54	96,48	98,26
		Médias					97,55	98,09	96,70	98,10
		Ranges					8,82	8,99	18,18	6,15
		Desvios padrão					1,617	2,033	4,414	1,385

= mínimo

A tabela abaixo mostra o resultado completo do teste feito na base de imagem de formulários de cadastro (Figura 3.2) construída durante este trabalho. No teste foi utilizada a técnica *k*-fold Cross Validation com três subconjuntos (na tabela K1, K2 e K3) e em 33,33% dos formulários o sistema alcançou 100% de acertos na classificação de impressos e manuscritos.

Imagem	Total de Impr.	Total de manus	Impr. class. cor-reta-mente	Manus class. cor-reta-mente	Class. como impr.	Class. como manus	Acur. Impr. em %	Acur. manus. em %	Prec. Impr. em %	Prec. manus. em %
Form0001	25	50	25	50	25	50	100,00	100,00	100,00	100,00
Form0019	25	65	23	65	23	67	92,00	100,00	100,00	97,01
Form0035	25	52	23	52	23	54	92,00	100,00	100,00	96,30
Form0039	25	57	24	57	24	58	96,00	100,00	100,00	98,28
Form0046	25	62	22	62	22	65	88,00	100,00	100,00	95,38
Form0076	25	50	24	50	24	51	96,00	100,00	100,00	98,04
Form0086	25	72	24	72	24	73	96,00	100,00	100,00	98,63
Form0106	25	42	25	42	25	42	100,00	100,00	100,00	100,00
K1	200	450	190	450	190	460	95,00	100,00	100,00	97,83
Form0002	25	28	24	27	25	28	96,00	96,43	96,00	96,43
Form0008	25	68	25	68	25	68	100,00	100,00	100,00	100,00
Form0034	25	33	25	33	25	33	100,00	100,00	100,00	100,00
Form0037	25	50	25	49	26	49	100,00	98,00	96,15	100,00
Form0045	25	76	25	74	27	74	100,00	97,37	92,59	100,00
Form0052	25	72	24	71	25	72	96,00	98,61	96,00	98,61
Form0070	25	51	25	50	26	50	100,00	98,04	96,15	100,00
Form0082	25	46	24	46	24	47	96,00	100,00	100,00	97,87
K2	200	424	197	418	203	421	98,50	98,58	97,04	99,29
Form0003	25	73	25	73	25	73	100,00	100,00	100,00	100,00
Form0021	25	53	25	53	25	53	100,00	100,00	100,00	100,00
Form0036	25	71	24	71	24	72	96,00	100,00	100,00	98,61
Form0040	25	41	23	41	23	43	92,00	100,00	100,00	95,35
Form0049	25	49	25	48	26	48	100,00	97,96	96,15	100,00
Form0066	25	58	25	58	25	58	100,00	100,00	100,00	100,00
Form0072	25	57	25	57	25	57	100,00	100,00	100,00	100,00
Form0107	25	53	24	53	24	54	96,00	100,00	100,00	98,15
K3	200	455	196	454	197	458	98,00	99,78	99,49	99,13
Totais	600	1329	583	1322	590	1339	97,17	99,47	98,81	98,73
		Médias					97,17	99,46	98,85	98,75
		Ranges					12,00	3,57	7,41	4,65
		Desvios padrão					1,892	0,761	1,580	0,801

= mínimo

APÊNDICE B

Neste apêndice estão as cópias dos e-mails enviados aos autores dos trabalhos, que realizam a distinção de texto impresso e manuscrito, referenciados na Seção1.2. Neles estão as solicitações e as respostas recebidas em relação ao acesso às bases de imagens usadas por eles, em seus respectivos trabalhos, para treinamento e teste dos sistemas.

```
From:       "David Doermann" <doermann@umiacs.umd.edu>
Subject:    RE: Information
Date:       Fri, September 26, 2008 8:54 pm
To:   lsilva@ic.uff.br
----------------------------------------------------------------------

I think this is a code 1, 2, 3 and it should be clear the classification
of each type.

-----

David Doermann
Laboratory for Language and Media Processing
Institute for Advanced Computer Studies
3451 AV Williams Building
University of Maryland
College Park, Maryland 20742

Phone:  301-405-1767          Fax:  301-314-2644
EMail:                                                  doermann@umiacs.umd.edu
</src/compose.php?send_to=doermann%40umiacs.umd.edu>
http://lamp.cfar.umd.edu

-----Original Message-----
From:     lsilva@ic.uff.br     </src/compose.php?send_to=lsilva%40ic.uff.br>
[mailto:lsilva@ic.uff.br </src/compose.php?send_to=lsilva@ic.uff.br>]
Sent: Friday, September 26, 2008 6:39 PM
To: David Doermann
Subject: RE: Information

Dear Sir David,

yes, but now i am trying understand the files txt and zone that are
together. Are there  any guide of use?

Best regards

Thank you very much by all

Lincoln Faria



> Did you get the data?
>
> -----
```

>
> David Doermann
> Laboratory for Language and Media Processing
> Institute for Advanced Computer Studies
> 3451 AV Williams Building
> University of Maryland
> College Park, Maryland 20742
>
> Phone: 301-405-1767 Fax: 443-638-0236
>
EMail:doermann@umiacs.umd.edu</src/compose.php?send_to=doermann%40umiacs.um
d.edu>
> http://lamp.cfar.umd.edu
>
>
>
>
> -----Original Message-----
> From: David Doermann
> Sent: Wednesday, September 24, 2008 2:20 PM
> To: lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>
> Subject: RE: Information
>
> I could not go through it yet, but you can access the layer data at:
>
> www.cfar.umd.edu/~doermann/LayerSeparation_English.rar
>
> -----
>
> David Doermann
> Laboratory for Language and Media Processing
> Institute for Advanced Computer Studies
> 3451 AV Williams Building
> University of Maryland
> College Park, Maryland 20742
>
> Phone: 301-405-1767 Fax: 301-314-2644
> EMail: doermann@umiacs.umd.edu
</src/compose.php?send_to=doermann%40umiacs.umd.edu>
> http://lamp.cfar.umd.edu
>
> -----Original Message-----
> From: lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>
[mailto:lsilva@ic.uff.br </src/compose.php?send_to=lsilva@ic.uff.br>]
> Sent: Tuesday, September 23, 2008 7:40 PM
> To: David Doermann
> Subject: RE: Information
>
> Dear Sir David,
>
> There is any new about the data set?
>
>
> Thank you very much
>
> Lincoln Faria
>
>
>> I have someone starting next week that may be able to put the data
>> together for you. Ill let you know when it is ready.
>>
>> Dave
>>

```
>>
>>
>>
>> -----
>> David Doermann
>> Laboratory for Language and Media Processing
>> Institute for Advanced Computer Studies
>> 3451 AV Williams Building
>> University of Maryland
>> College Park, Maryland 20742
>>
>> Phone:  301-405-1767           Fax:  443-638-0236
>>              EMail:                              doermann@umiacs.umd.edu
</src/compose.php?send_to=doermann%40umiacs.umd.edu>
>> http://lamp.cfar.umd.edu
>>

>> -----Original Message-----
>>  From:  lsilva@ic.uff.br  </src/compose.php?send_to=lsilva%40ic.uff.br>
[mailto:lsilva@ic.uff.br </src/compose.php?send_to=lsilva@ic.uff.br>]
>> Sent: Tuesday, September 09, 2008 8:44 AM
>> To: David Doermann
>> Subject: RE: Information
>>
>>
>> Dear Sir David,
>>
>> Please, I'd like to know if the date is available.
>>
>> Thank you very much
>>
>>   Lincoln Faria
>>
>>> The student who did this project has left.  Let me see if I can find
>> the
>>> data.
>>>
>>> Dave
>>>
>>> -----
>>> David Doermann
>>> Laboratory for Language and Media Processing
>>> Institute for Advanced Computer Studies
>>> 3451 AV Williams Building
>>> University of Maryland
>>> College Park, Maryland 20742
>>>
>>> Phone:  301-405-1767           Fax:  443-638-0236
>>>
EMail:doermann@umiacs.umd.edu</src/compose.php?send_to=doermann%40umiacs.um
d.edu>
>>> http://lamp.cfar.umd.edu
>>>
>>> -----Original Message-----
>>>  From:  lsilva@ic.uff.br  </src/compose.php?send_to=lsilva%40ic.uff.br>
[mailto:lsilva@ic.uff.br </src/compose.php?send_to=lsilva@ic.uff.br>]
>>> Sent: Monday, August 25, 2008 3:36 PM
>>> To: David Doermann
>>> Subject: Information
>>>
>>> Dear Sir,
>>>
>>> I am a graduate (MSc.) student at the UFF (Universidade Federal
```

```
>>> Fluminense
>>> Federal Fluminense University) in Rio de Janeiro, Brazil.
>>> I have read your works and I would like of know how it is possible to
>>> access the database used by you on work "Machine Printed Text
and
>>> Handwriting Identification in Noisy Document Images", becouse I
would
>>> like
>>> of to use the same database into may master degree dissertation on the
>> task
>>> of
>>> identification between Machine Printed and Handwriting Text in
>> document
>>> analysis by  images.
>>>
>>> Thank you very much
>>>
>>> Lincoln Faria
>>>

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

------------------------------------------------------------------------

From:       "Faisal Farooq" <ffarooq2@cedar.buffalo.edu>
Subject:    Re: Database
Date:       Wed, September 24, 2008 3:07 pm
To:   lsilva@ic.uff.br
------------------------------------------------------------------------

Hi Lincoln,

No unfortunately not. The government offices that are responsible for this
unfortunately do not work so fast.
I will update you when I get anything.

Faisal

On    Tue,    Sep    23,    2008    at    7:27    PM,    <lsilva@ic.uff.br
</src/compose.php?send_to=lsilva%40ic.uff.br>> wrote:

> Hi Faisal,
>
> Have you now novelty about the data distribution?
>
> Thank you very much
>
>  Best regards
>
>  Lincoln Faria
>
> > Hi Lincoln,
> >
> > Unfortunately the data distribution legal work is unknown yet. So, I
> > cannot
> > share it with you right now.
> > As soon as I hear anything different I will let you know.
> >
> > Good Luck
> > Faisal
> >
>   >   On   Tue,   Sep   9,   2008   at   11:17   PM,   <lsilva@ic.uff.br
</src/compose.php?send_to=lsilva%40ic.uff.br>> wrote:
```

```
> >
> >> Dear Sir,
> >>
> >> I am a graduate (MSc.) student at the UFF (Universidade Federal
> >> Fluminense
> >> Federal Fluminense University) in Rio de Janeiro, Brazil.
> >> I have read your works and I would like of know how it is possible to
> >> access the database used by you on the work "Identifying handwritten
> >> text
> >> in mixed documents - The 18th International Conference on Pattern
> >> Recognition (ICPR'06)", becouse I would like of to use this database
> >> into
> >> may master degree dissertation on the task of identification between
> >> Machine Printed and Handwriting Text in document
> >> analysis by  images.
> >>
> >> Thank you very much
> >>
> >> Best regards
> >>
> >> Lincoln Faria
-------------------------------------------------------------------------
*Attachments:*
untitled-[2]
Size:       2.1 k
Type:       text/html

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

-------------------------------------------------------------------------
From:       "Umapada Pal" <umapada_pal@yahoo.com>
Subject:    Re: Database
Date:       Mon, September 22, 2008 12:37 am
To:   lsilva@ic.uff.br
-------------------------------------------------------------------------

Dear Lincoln Faria

  Thank you very much for  your interest on your work.

 I am visiting now Japan for more than a month and  I cannot tell you now
the status
of the database.  After going back I shall check it.

 Best wishes --Umapada Pal

--- On Mon, 9/22/08, lsilva@ic.uff.br
</src/compose.php?send_to=lsilva%40ic.uff.br> <lsilva@ic.uff.br
</src/compose.php?send_to=lsilva%40ic.uff.br>> wrote:

From: lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>
<lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>>
Subject: Database
To: umapada@isical.ac.in </src/compose.php?send_to=umapada%40isical.ac.in>
Date: Monday, September 22, 2008, 7:13 AM

Dear Sir,

I am a graduate (MSc.) student at the UFF (Universidade Federal Fluminense
Federal Fluminense University) in Rio de Janeiro, Brazil.
I have read your works and I would like of know how it is possible to
access the database used by you on the work "Automatic Separation of
Machine-Printed and Hand-Written Text Lines", becouse I would like of to
```

use this database into my master degree dissertation on the task of identification between Machine Printed and Handwriting Text in document analysis by images.

Thank you very much

Best regards

Lincoln Faria

This mail is scanned by Ironport

Attachments:
untitled-[2]
Size: 1.5 k
Type: text/html

%%

From: "Kavallieratou Ergina" <kavallieratou@aegean.gr>
Subject: ??: ÁÐ: Database
Date: Wed, September 17, 2008 11:41 am
To: "lsilva@ic.uff.br" <lsilva@ic.uff.br>

You should ask the creators of IAM.

____
???: lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>
[lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>]
????????: ???????, 17 ??????????? 2008 3:45 ??
????: Kavallieratou Ergina
????: Re: ÁÐ: Database

Hello,

Firstly, thank you very much by information above the GRUHD database, but also i would like of know, if possible, how it is measured the accuracy of the classification by using the IAM-DB database.

Thank you very much

Best regards

Lincoln Faria

> Hello Lincoln,
> I reply to this e-mail, so you can see that both papers are mine. In fact
> GRUHD database is not anymore in use for several reasons. Right now I am
> in a different university and we have designed and built another database
> much better and with better software, that you will be able to find in my
> website sometime in this academic year, as soon as it is published first.
> Ergina Kavallieratou

> ____

> Áðü: Stamatatos Efstathios

> ÁðïóôïëÞ: Ôñßôç, 16 Óåðôåìâñßïõ 2008 10:47 ìì

```
> Ðñïò: Kavallieratou Ergina

> ÈÝìá: FW: Database

>> -----Original Message-----

>> From: lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>
[mailto:lsilva@ic.uff.br </src/compose.php?send_to=lsilva@ic.uff.br>]
>> Sent: Tuesday, September 16, 2008 5:45 PM
>> To: Stamatatos Efstathios
>> Subject: Database
>> Dear Sir,
>> I am a graduate (MSc.) student at the UFF (Universidade Federal
>> Fluminense
>> Federal Fluminense University) in Rio de Janeiro, Brazil.
>> I have read your works and I would like of know how it is possible to
>> access the database used by you on the work "Machine-Printed from
>> Handwritten Text Discrimination, Proceedings of the 9th Int'l Workshop
>> on
>> Frontiers in Handwriting Recognition (IWFHR-9 2004)", becouse I would
>> like
>> of to use this database into may master degree dissertation on the task
>> of
>> identification between Machine Printed and Handwriting Text in document
>> analysis by  images.
>> Thank you very much
>> Best regards
>> Lincoln Faria

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%


------------------------------------------------------------------------

From:       "Bidyut Baran Chaudhuri" <bbcisical@gmail.com>
Subject:    Fwd: Database
Date:       Tue, September 16, 2008 8:18 am
To:   "Umapada Pal" <umapada_pal@yahoo.com>,"Umapada Pal"
<umapada@isical.ac.in>
Cc:   lsilva@ic.uff.br
------------------------------------------------------------------------

Umapada

Do you have an electronic copy of the paper? Then please send it to this
researcher. Hope you are attending DAS and finding it interesting. You may
act after the DAS. Best----BBC

---------- Forwarded message ----------
From: lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>
<lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>>
Date: Sep 15, 2008 12:48 AM
Subject: Database
To: bbc@isical.ac.in </src/compose.php?send_to=bbc%40isical.ac.in>


Dear Sir,

I am a graduate (MSc.) student at the UFF (Universidade Federal Fluminense
Federal Fluminense University) in Rio de Janeiro, Brazil.
I have read your works and I would like of know how it is possible to
access the database used by you on the work "Machine-Printed and
```

Hand-Written Text Lines Identifcation, Pattern Recognition Letters 22(2001) 431-441", becouse I would like of to use this database into may master degree dissertation on the task of identification between Machine Printed and Handwriting Text in document analysis by images.

Thank you very much

Best regards

Lincoln Faria

****
Dr. Bidyut B. Chaudhuri FNA, FNAE, FIAPR, FIEEE
Professor and Head
Computer Vision & Pattern Recognition Unit
Indian Statistical Institute
203 B. T. Road
Kolkata 700108
West Bengal, India
Phone: (91) (33) 2575 2852
www.isical.ac.in/~bbc
****

Attachments:
untitled-[2]
Size: 1.7 k
Type: text/html

%%

From: "Ramanaprasad Vemulapati" <raman-v@buffalo.edu>
Subject: Re: Database
Date: Mon, September 15, 2008 7:03 pm
To: lsilva@ic.uff.br

Hello,

I do not have access to the data that you have referenced anymore as it it more than 10 years since I published that work. Also, even if I have access to that data I would not be able to pass it to you as it has confidential tax payer information on them. You may be able obtain similar data from NIST (www.nist.gov).

Sincerely

Ramanaprasad

lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br> wrote:
>
> Dear Sir,
>
> I am a graduate (MSc.) student at the UFF (Universidade Federal
Fluminense
> Federal Fluminense University) in Rio de Janeiro, Brazil.
> I have read your works and I would like of know how it is possible to
> access the database used by you in work "A System to Read Names and
> Addresses on Tax Forms", becouse I would like of to use this database
into
> may master degree dissertation on the task of identification between
> Machine

```
> Printed and Handwriting Text in document analysis by  images.
>
> Thank you very much
>
> Best regards
>
> Lincoln Faria
>

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
-------------------------------------------------------------------------

From:       "Faisal Farooq" <ffarooq2@cedar.buffalo.edu>
Subject:    Re: Database
Date:       Wed, September 10, 2008 1:34 am
To:   lsilva@ic.uff.br
-------------------------------------------------------------------------

Hi Lincoln,

The database was collected in the university at buffalo. Let me verify if
we
can share it openly.
If yes then I will send it to you. If the policy prohibits sharing then
unfortunately I wont be
able to give it to you.

Thanks for your interest in my research. Good luck with your MSc.

Faisal

On Tue, Sep 9, 2008 at 11:17 PM, <lsilva@ic.uff.br
</src/compose.php?send_to=lsilva%40ic.uff.br>> wrote:

> Dear Sir,
>
> I am a graduate (MSc.) student at the UFF (Universidade Federal
Fluminense
> Federal Fluminense University) in Rio de Janeiro, Brazil.
> I have read your works and I would like of know how it is possible to
> access the database used by you on the work "Identifying handwritten text
> in mixed documents - The 18th International Conference on Pattern
> Recognition (ICPR'06)", becouse I would like of to use this database into
> may master degree dissertation on the task of identification between
> Machine Printed and Handwriting Text in document
> analysis by  images.
>
> Thank you very much
>
> Best regards
>
> Lincoln Faria
>
>
>
>
>
>

-------------------------------------------------------------------------
*Attachments:*
untitled-[2]
Size:       1.3 k
```

```
Type:       text/html


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
------------------------------------------------------------------------

From:       "Lambert Schomaker" <L.Schomaker@ai.rug.nl>
Subject:    Re: Information
Date:       Tue, September 9, 2008 1:35 pm
To:   lsilva@ic.uff.br
------------------------------------------------------------------------

Dear Lincoln Faria

We have put the data here:
http://www.ai.rug.nl/~lambert/unipen/ImUnipen/
as a .tgz file.

Please read the "Terms of usage": non-commercial
scientific use, while referring to the correct
IEEE PAMI paper in which this data set was used.

Best regards,
prof. dr. Lambert Schomaker


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

*Subject:  *      RE: Information
*From:  *   "Yefeng Zheng" <zhengyf@umiacs.umd.edu>
*Date:  *   Mon, August 25, 2008 11:25 am
*To:  *     lsilva@ic.uff.br
*Priority:  *     Normal
*Options:  *      View Full Header


Hi Lincoln,

I have graduated from the University of Maryland. Would you please contact
my former
advisor, Dr. David Doermann? His e-mail is doermann@umiacs.umd.edu
</src/compose.php?send_to=doermann%40umiacs.umd.edu>.

Regards,

Yefeng


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

From: lsilva@ic.uff.br </src/compose.php?send_to=lsilva%40ic.uff.br>
[mailto:lsilva@ic.uff.br </src/compose.php?send_to=lsilva@ic.uff.br>]
Sent: Sat 8/23/2008 10:37 AM
To: zhengyf@cfar.umd.edu </src/compose.php?send_to=zhengyf%40cfar.umd.edu>
Subject: Information



Dear Sir,

I am a graduate (MSc.) student at the UFF (Universidade Federal Fluminense
Federal Fluminense University) in Rio de Janeiro, Brazil.
I have read your works and I would like of know how it is possible to
```

access the database used by you on the work "Machine Printed Text and Handwriting Identification in Noisy Document Images", becouse I would like of to use this database into may master degree dissertation on the task of identification between Machine Printed and Handwriting Text in document analysis by images.

Thank you very much

Best regards

Lincoln Faria

%%

*Subject: * Re: Information
*From: * "Marius Bulacu" <M.Bulacu@ai.rug.nl>
*Date: * Fri, August 22, 2008 5:13 am
*To: * lsilva@ic.uff.br
*Priority: * Normal
*Options: * View Full Header

Dear Lincoln Faria,

The Firemaker dataset contains forensic samples and is not available publicly. The ImUnipen dataset is derived from on-line dataset Unipen, which you can obtain from my supervisor:

Prof. Lambert Schomaker
L.Schomaker@ai.rug.nl </src/compose.php?send_to=L.Schomaker%40ai.rug.nl>

He is responsible for maintaining and distributing the Unipen dataset.

Good luck with your research!

Best wishes,

Marius Bulacu

%%

Olá Lincoln,

Desculpe não poder ajudar, mas como o Prof. Jacques falou... não é fácil conseguir dados (cheques de verdade) junto aos bancos.

Segue minha Tese.

Obrigada,
- Ocultar texto das mensagens anteriores -

Cinthia

2008/11/4 Lincoln Faria <lincoln.faria@gmail.com>:
> Olá Prof.ª Cinthia,
>

> Tudo bem, mas de qualquer forma muito obrigado por tudo. Quanto a tese
> ainda não tenho. Tentei pegar na página da PUC, mas lá o PDF não esta
> disponível. Em que link eu consigo este PDF.
>
> Atenciosamente
> Lincoln
>
> Em 04/11/08, Cinthia Freitas<almendracinthia@gmail.com> escreveu:
>> Olá Lincoln,
>>
>> Não temos manuscritos e pré-impresso juntos. Temos somente
>> manuscritos. Você tem a minha Tese de Doutorado? Lá são explicados
>> todos os procedimentos utilizados na colheita dos dados manuscritos.
>> Recomendo que você dê uma olhadinha, para entender melhor o que temos.
>>
>> Desculpe, não poder ajudar.... mas só temos manuscritos.
>>
>> Abraços,
>>
>> Cinthia
>>
>> 2008/11/4 Lincoln Faria <lincoln.faria@gmail.com>:
>>> Olá Prof.ª Cinthia, tudo bem?
>>>
>>> Então vcs não têm uma base com imagens que contenham impressos e
>>> manuscritos mesmo que sejam dados hipotéticos, que vcs tenham usados?
>>>
>>> Mesmo assim agradeço pelo atendimento a nossa solicitação.
>>>
>>> Obrigado
>>> Lincoln
>>>
>>> Em 31/10/08, Cinthia Freitas<almendracinthia@gmail.com> escreveu:
>>>> Olá Lincoln,
>>>>
>>>> As bases utilizadas por mim, pelo Prof. Luiz Eduardo, Marisa, Prof.
>>>> Alceu e Prof. Justino não possuem o tipo de informação que você
>>>> necessita. Criamos bases independentes de formulário, ou de fundo ou
>>>> de informação pré-impressa.
>>>>
>>>> Estou copiando esta mensagem para o Profs. Jacques Facon e Alessandro
>>>> Koerich, pois talvez eles disponham de bases com as informações que
>>>> você necessita (fundo + pré-impresso + manuscrito).
>>>>
>>>> Obrigada,
>>>>
>>>> A+
>>>>
>>>> Cinthia
>>>>
>>>> 2008/10/30 Lincoln Faria <lincoln.faria@gmail.com>:
>>>>> Oi Prof.ª Cinthia,
>>>>>
>>>>> Desculpe-me mais uma vez lhe importunar, mas gostaria de saber se há
>>>>> novidades sobre os dados, ou seja, vocês possuem o tipo de imagens
>>>>> descrito no e-mail anterior?
>>>>>
>>>>> Obrigado
>>>>> Lincoln
>>>>>
>>>>> Em 18/10/08, Lincoln Faria<lincoln.faria@gmail.com> escreveu:
>>>>>> Oi Profª. Cinthia,
>>>>>>

>>>>>> O sistema que desenvolvemos tem por objetivo distingüi em uma mesma
>>>>>> imagem de documento o que é texto impresso e o que é texto
manuscrito.
>>>>>> Então, estamos interessados em imagens de documentos que contenha os
>>>>>> dois tipos de textos como formulários preenchidos ou cheques
>>>>>> preenchidos.
>>>>>>
>>>>>> Obrigado
>>>>>> Lincoln
>>>>>>
>>>>>> Em 17/10/08, Cinthia Freitas<almendracinthia@gmail.com> escreveu:
>>>>>>> Olá Lincoln,
>>>>>>>
>>>>>>> Desde que o Prof. Flávio, conversou comigo sobre a base de dados é
>>>>>>> que
>>>>>>> estou tentando organizar isto.
>>>>>>>
>>>>>>> O problema é que não temos uma base única. Cada sub-item do cheque
>>>>>>> constitui uma base de imagens: valor numérico, palavras do extenso,
>>>>>>> data e assinaturas.
>>>>>>>
>>>>>>> Estas bases foram montadas a partir de situações hipotéticas e não
>>>>>>> constituem cheque de verdade, pois as instituições financeiras tem
>>>>>>> reservas para não fornecer este tipo de documento.
>>>>>>>
>>>>>>> O Prof. Alessandro Koerich, durante o seu Mestrado, utilizou uma
base
>>>>>>> de cheques bancários (sem preenchimento).
>>>>>>>
>>>>>>> Qual o seu interesse? Dígitos, palavras, datas, etc? ou o
forumulário
>>>>>>> propriamente dito?
>>>>>>>
>>>>>>> Aguardo retorno.
>>>>>>>
>>>>>>> Obrigada,
>>>>>>>
>>>>>>> Cinthia Freitas
>>>>>>>
>>>>>>> 2008/10/16 Lincoln Faria <lincoln.faria@gmail.com>:
>>>>>>>>> Boa tarde Profª. Cinthia.
>>>>>>>>>
>>>>>>>>> Sou aluno da Profª. Aura Conci do IC-UFF e estou lhe escrevendo,
>>>>>>>>> pois
>>>>>>>>> o Prof°. Flávio me pediu que entrasse em contato consigo a
respeito
>>>>>>>>> do
>>>>>>>>> envio da base de imagens de cheques de bancos usadas em seus
>>>>>>>>> trabalhos.
>>>>>>>>>
>>>>>>>>> Antecipadamente, muito obrigado.
>>>>>>>>>
>>>>>>>>> Atenciosamente
>>>>>>>>> Lincoln
>>>>>>>>>
>>>>>>>>> Em 16/10/08, Flavio Bortolozzi<flavio.bortolozzi@opet.com.br>
>>>>>>>>> escreveu:
>>>>>>>>>>> Boa tarde Lincoln
>>>>>>>>>>>
>>>>>>>>>>> Ontem a tarde conversei com a professora Cinthia, por favor faça
o
>>>>>>>>>>> contato
>>>>>>>>>>> com ela no e-mail que estou enviando com cópia para ela.

```
>>>>>>>>> Fico no aguardo da confirmação.
>>>>>>>>> Atenciosamente
>>>>>>>>>
>>>>>>>>> Prof. Flávio
>>>>>>>>>
>>>>>>>>>
>>>>>>>>> _____
>>>>>>>>> De: Lincoln Faria [lincoln.faria@gmail.com]
>>>>>>>>> Enviado: quarta-feira, 15 de outubro de 2008 8:32
>>>>>>>>> Para: Flavio Bortolozzi
>>>>>>>>> Cc: aura.conci.uff@gmail.com
>>>>>>>>> Assunto: Resp.: Database - Prof. Flavio
>>>>>>>>>
>>>>>>>>> Boa noite Prof. Flávio,
>>>>>>>>>
>>>>>>>>> Como o senhor me pediu, estou lhe escrevendo para ter notícias
>>>>>>>>> sobre
>>>>>>>>> as imagens de cheques.
>>>>>>>>>
>>>>>>>>> Antecipadamente, muito obrigado.
>>>>>>>>>
>>>>>>>>> Atenciosamente
>>>>>>>>>    Lincoln
>>>>>>>>>
>>>>>>>>> Em 11/10/08, Flavio Bortolozzi<flavio.bortolozzi@opet.com.br>
>>>>>>>>> escreveu:
>>>>>>>>>>> Boa noite Lincoln
>>>>>>>>>>>
>>>>>>>>>>> Eu pensei que a Professora Cinthia já tinha providenciado.
>>>>>>>>>>> Vou cobrar dela na segunda feira. Por favor me cobre na terça
que
>>>>>>>>>>> te
>>>>>>>>>>> dou
>>>>>>>>>>> um
>>>>>>>>>>> retorno.
>>>>>>>>>>> Atenciosamente
>>>>>>>>>>> Prof. Flávio
>>>>>>>>>>> _____
>>>>>>>>>>> De: Lincoln Faria [lincoln.faria@gmail.com]
>>>>>>>>>>> Enviado: terça-feira, 7 de outubro de 2008 7:58
>>>>>>>>>>> Para: Flavio Bortolozzi
>>>>>>>>>>> Cc: aura.conci.uff@gmail.com
>>>>>>>>>>> Assunto: Re: Database - Prof. Flavio
>>>>>>>>>>>
>>>>>>>>>>> Prof. Flávio, boa tarde.
>>>>>>>>>>>
>>>>>>>>>>> Desculpe - me por lhe escrever novamente, pois sei que o senhor
é
>>>>>>>>>>> uma
>>>>>>>>>>> pessoa muito ocupada, mas gostaria de ter notícias sobre a base
de
>>>>>>>>>>> imagens de cheques de banco. Estamos na fase de testes do
sistema
>>>>>>>>>>> e
>>>>>>>>>>> gostariamos de verificar os resultados sobre esta base na qual
>>>>>>>>>>> seus
>>>>>>>>>>> trabalhos foram testados.
>>>>>>>>>>>
>>>>>>>>>>> Antecipadamente, muito obrigado.
>>>>>>>>>>>
>>>>>>>>>>> Atenciosamente
>>>>>>>>>>>    Lincoln
>>>>>>>>>>>
```

```
>>>>>>>>>>> Em 25/09/08, Flavio Bortolozzi<flavio.bortolozzi@opet.com.br>
>>>>>>>>>> escreveu:
>>>>>>>>>>> Boa noite Cinthia e Luis
>>>>>>>>>>>
>>>>>>>>>>> Peço-lhe o favor se possível de ver como podemos enviar ao
>>>>>>>>>>> Lincoln
>>>>>>>>>>> a
>>>>>>>>>>> base
>>>>>>>>>>> de
>>>>>>>>>>> cheques solicitada, pelo Lincoln e pela Profª. Aura Conci.
>>>>>>>>>>> Agradeço antecipadamente
>>>>>>>>>>>
>>>>>>>>>>> Flávio
>>>>>>>>>>>
>>>>>>>>>>> ps - Lincoln aproveite para dar um abraço na Aura.
>>>>>>>>>>>
_________________________________
>>>>>>>>>>>
>>>>>>>>>>> De: Lincoln Faria [mailto:lincoln.faria@gmail.com]
>>>>>>>>>>> Enviada: qua 24/9/2008 20:52
>>>>>>>>>>> Para: Flavio Bortolozzi
>>>>>>>>>>> Assunto: Re: Database - Prof. Flavio
>>>>>>>>>>>
>>>>>>>>>>> Prof. Flávio, bom dia!
>>>>>>>>>>>
>>>>>>>>>>> Provavelmente o Senhor tenha chegado recentemente da França e
eu
>>>>>>>>>>> já
>>>>>>>>>>> estou
>>>>>>>>>>> escrevendo lhe importunando, mas gostaria de saber se o envio
do
>>>>>>>>>>> banco de dados
>>>>>>>>>>> de
>>>>>>>>>>> cheques já se encontra viável.
>>>>>>>>>>>
>>>>>>>>>>> Atenciosamente
>>>>>>>>>>>   Lincoln
>>>>>>>>>>>
>>>>>>>>>>> 2008/9/12 Flavio Bortolozzi <flavio.bortolozzi@opet.com.br>
>>>>>>>>>>>
>>>>>>>>>>>       Bom dia Lincoln
>>>>>>>>>>>
>>>>>>>>>>>       Pode sim, eu estou na Franca ate dia 23. Se puder me
envie
>>>>>>>>>>> envie
>>>>>>>>>>> um
>>>>>>>>>>> e-mail
>>>>>>>>>>> desois do dia 23 que combinamos como enviar o banco de cheques.
>>>>>>>>>>>       Atenciosamente
>>>>>>>>>>>
>>>>>>>>>>>       Prof. Flavio Bortolozzi
>>>>>>>>>>>       _________________________________
>>>>>>>>>>>
>>>>>>>>>>>       Von: Lincoln Faria [mailto:lincoln.faria@gmail.com]
>>>>>>>>>>>       Gesendet: Do 11/9/2008 20:18
>>>>>>>>>>>       An: flavio.bortolozzi@pucpr.br
>>>>>>>>>>>       Cc: flavio.bortolozzi@pq.cnpq.br
>>>>>>>>>>>       Betreff: Database
>>>>>>>>>>>
>>>>>>>>>>>
>>>>>>>>>>>       Prof. Flávio Bortolozzi,
>>>>>>>>>>>
>>>>>>>>>>>       Sou aluno de mestrado da UFF orientado pela Profª. Aura
```

```
>>>>>>>>>>> Conci
>>>>>>>>>>> e
>>>>>>>>>>> estou
>>>>>>>>>>> desenvolvendo um sistema que distingue entre texto impresso e
>>>>>>>>>>> texto
>>>>>>>>>>> manuscrito em uma imagem de documento e li o resumo do
trabalho:
>>>>>>>>>>>
>>>>>>>>>>>       Distinguishing between Handwritten and Machine Printed
Text
>>>>>>>>>>> in
>>>>>>>>>>> Bank
>>>>>>>>>>> Cheque
>>>>>>>>>>> Images
>>>>>>>>>>>
>>>>>>>>>>>       Book Series      Lecture Notes in Computer Science
>>>>>>>>>>>
<https://commerce.metapress.com/content/105633/?p=8694f10d1d1540d5af3d96f7d
bc38603&pi=0>
>>>>>>>>>>>       Publisher        Springer Berlin / Heidelberg
>>>>>>>>>>>       ISSN     0302-9743 (Print) 1611-3349 (Online)
>>>>>>>>>>>       Volume   Volume 2423/2002
>>>>>>>>>>>       Book     Document Analysis Systems V
>>>>>>>>>>>
<https://commerce.metapress.com/content/7b5y7mjl43kj/?p=8694f10d1d1540d5af3
d96f7dbc38603&pi=0>
>>>>>>>>>>>       DOI      10.1007/3-540-45869-7
>>>>>>>>>>>       Copyright        2002
>>>>>>>>>>>       ISBN     978-3-540-44068-0
>>>>>>>>>>>       DOI      10.1007/3-540-45869-7_7
>>>>>>>>>>>       Pages    69-76
>>>>>>>>>>>       Subject Collection       Computer Science
>>>>>>>>>>> <https://commerce.metapress.com/computer-science/>
>>>>>>>>>>>       SpringerLink Date        Tuesday, January 01, 2002
>>>>>>>>>>>
>>>>>>>>>>>       Gostaria de saber como posso ter acesso ao banco de dados
>>>>>>>>>>> (imagens
>>>>>>>>>>> de
>>>>>>>>>>> cheques) usados neste trabalho para que eu possa fazer testes
no
>>>>>>>>>>> meu
>>>>>>>>>>> sistema
>>>>>>>>>>> e como ter acesso ao texto  completo deste artigo?
>>>>>>>>>>>
>>>>>>>>>>>       Atenciosamente
>>>>>>>>>>>         Lincoln
>>>>>>>>>>>
<http://mail.opet.com.br/Exchange/flaviobortolozzi/Rascunhos/images/cleardo
t.gif>
>>>>>>>>>>>

            Cinthia-PhD-2001-pdf.zip
2494K Examinar e baixar
```