

Recuperação de Documentos baseada em Conteúdo

Trabalho de Conclusão de Curso

Engenharia da Computação

Tiago Câmara Bockholt

Orientador: Prof. Dr. Carlos Alexandre Barros de Mello

Co-Orientador: Prof. Dr. Adriano Lorena Inácio de Oliveira

Tiago Câmara Bockholt

**Recuperação de Documentos baseada em
Conteúdo**

Monografia apresentada como requisito parcial para
obtenção do diploma de Bacharel em Engenharia da
Computação pela Escola Politécnica de Pernambuco
– Universidade de Pernambuco.

Recife, dezembro de 2008.

Agradecimentos

Agradeço a Deus pelas oportunidades oferecidas e que me levaram a concluir mais uma etapa da minha vida. Agradeço aos meus Pais pelo carinho, dedicação, paciência e por terem me dado uma excelente educação. Aos meus irmãos por servirem de exemplo de como lidar com os desafios da vida.

Aos professores do curso de Engenharia da Computação. Em especial agradeço a dedicação, dicas e auxílio do Prof. Dr. Carlos Alexandre Barros de Mello durante toda a orientação deste trabalho e também pelas excelentes aulas que tive durante o curso que enriqueceram muito minha formação. Um agradecimento especial ao Prof. Dr. Adriano Lorena Inácio de Oliveira por acreditar no meu potencial durante a iniciação científica e pela co-orientação desta monografia.

Aos meus amigos, em especial aos amigos da faculdade, que tornaram a jornada diária de aulas, seminários e projetos mais alegres. A Deda pelo zelo nas atividades domésticas que me auxiliaram na conquista desse título. Ainda que por último, mas não menos importante, agradeço a Angélica por suportar meu humor nos momentos de estresse e pelo constante carinho e suporte nas fases mais difíceis do curso. A todos citados a cima, deixo aqui a minha admiração e o meu muito obrigado.

Resumo

Sistemas de recuperação de documentos são importantes ferramentas que melhoram a indexação e gerenciamento de grandes coleções de documentos digitais. Durante o desenvolvimento deste trabalho, nós estudamos o estado da arte da recuperação de documentos baseada em conteúdo na área de reconhecimento de padrões. Foi proposto um sistema de recuperação de documentos baseado na extração de várias características de imagens que podem representar as diferentes classes do nosso repositório. Foram realizados testes usando a transformada de Hough para discriminar se um documento é manuscrito ou datilografado. Descritores de textura são usados para extrair características relevantes na segmentação automática do logo. Finalmente, realizamos vários testes para avaliar a precisão da abordagem proposta de recuperação de documentos baseada em conteúdo. Comparamos as taxas de acertos na recuperação de documentos usando diferentes técnicas como k-NN com e sem *relevance feedback* e várias arquiteturas de uma rede neural MLP com algoritmo de aprendizado *backpropagation*.

Abstract

Document Image Retrieval Systems are important tools that improve the indexing and management of large collections of digital documents. During the development of this work, we studied the state-of-the-art of content based document images retrieval in pattern recognition area. We propose a retrieval system based on the extraction of several image features that may represent the different classes of our repository. For manuscript or typed documents, tests were made using the Hough Transform to discern between them. Texture descriptors are used to extract relevant features for the logos automatic segmentation. Finally, we performed several experiments to evaluate the accuracy of the proposed document image retrieval approach. We compared the true positive rates in the document image retrieval using different techniques like k-NN with and without relevance feedback and several architectures of MLP neural networks with backpropagation learning algorithm.

Sumário

Índice de Figuras	v
Índice de Tabelas	vi
Tabela de Símbolos e Siglas	vii
Capítulo 1 Introdução	8
1.1 Organização do Trabalho	9
Capítulo 2 Recuperação de Imagens baseada em Conteúdo - CBIR	10
2.1 Características de Imagens utilizadas nos Sistemas CBIR	14
2.1.1 Características de Cor	14
2.1.2 Características de Forma	14
2.1.3 Características de Textura	15
2.2 Descritores de Texturas	15
2.2.1 Matriz de Co-Ocorrência de Tons de Cinza (<i>GLCM – Gray Level Co-Occurrence Matrix</i>)	16
2.2.2 Histograma da Soma/Diferença (<i>Sum/Difference Histogram</i>)	17
Capítulo 3 Recuperação de Documentos baseada em Conteúdo - DIR	18
3.1 Componentes de um Classificador de Documentos	19
3.1.1 Classificação	19
3.1.2 Arquitetura do Classificador	20
3.1.3 Desempenho dos Resultados	23
3.2 Extração das características de documentos	23
3.2.1 Pré-processamento das imagens	24
3.2.2 Histograma da projeção	24
3.2.3 Mapas de Transição	25
3.3 Técnicas Utilizadas	26
3.3.1 Morfologia Matemática	27
3.3.2 Técnica de Recuperação NN (<i>Nearest Neighbors</i>)	27
3.3.3 <i>Relevance Feedback</i>	28
3.3.4 Rede Neural MLP como técnica de recuperação	29
3.3.5 Transformada de Hough	31
Capítulo 4 Experimentos e Resultados	33

4.1	Base de Imagens de Documentos	<i>iv</i> 34
4.2	Metodologia e Resultados	35
Capítulo 5 Conclusões e Trabalhos Futuros		44
5.1	Contribuições	44
5.2	Trabalhos Futuros	45
Bibliografia		46
Apêndice A		49

Índice de Figuras

Figura 1. Fluxograma de funcionamento básico de um <i>CBIR</i> .	12
Figura 2. Imagens semanticamente relacionadas referenciam umas às outras por características visuais.	12
Figura 3. Imagens semanticamente iguais, mas diferentemente representadas pelo histograma de cor.	13
Figura 4. À direita, a matriz de co-ocorrência é preenchida indicando quantas vezes a cor i e a cor j estão presentes na imagem I (à esquerda) separados por uma distância d igual a 1 [11].	17
Figura 5. Exemplo de varredura da imagem para geração do histograma da soma com deslocamento d_x e d_y .	17
Figura 6. Os três principais componentes de um sistema de recuperação de documentos.	20
Figura 7. Sequência clássica do reconhecimento de documentos onde a presença de texto é dominante.	21
Figura 8. Histogramas da Projeção (vertical e horizontal) de uma imagem de documento histórico.	25
Figura 9. Histogramas que demonstram a computação do Mapa de Transição [22].	26
Figura 10. Arquitetura de uma rede MLP com duas camadas escondidas.	29
Figura 11. Ilustração das direções de propagação do sinal funcional (fase <i>forward</i>) e do erro (fase <i>backward</i>).	30
Figura 12. Quantização do plano de parâmetros para o uso da transformada de Hough.	32
Figura 13. Exemplos da base de imagens de documento utilizada.	34
Figura 14. Parte de dois documentos datilografados pertencentes ao repositório de imagens de documentos utilizado: (b) apresenta traços mais grossos que (a).	35
Figura 15. Interface gráfica do sistema de recuperação de documentos baseada em conteúdo implementada.	36
Figura 16. Diferentes módulos que compõem a ferramenta implementada.	36
Figura 17. Elemento estruturante da erosão na operação de abertura.	37
Figura 18. Resultado obtido pela operação de abertura nos tipos de documentos manuscritos e datilografados usados na base.	38
Figura 19. Resultado obtido pela operação morfológica de abertura nas principais classes de cheques usados na base.	38
Figura 20. Documento datilografado e seu respectivo espaço de Hough.	43
Figura 21. Documento manuscrito e seu respectivo espaço de Hough.	43

Índice de Tabelas

Tabela 1. Representação das classes dos documentos no sistema.	39
Tabela 2. Erro obtido no final do treinamento variando a quantidade de neurônios na camada escondida com taxa de aprendizado 0,1 e 10.000 ciclos de Treinamento.....	40
Tabela 3. Média de acertos de acordo com a quantidade de neurônios na camada escondida com taxa de aprendizado 0,1 e 10.000 ciclos de Treinamento.....	40
Tabela 4. Matriz de confusão para o resultado da rede MLP com taxa de aprendizado 0,1 e 10.000 ciclos de treinamento e 16 neurônios na camada escondida.	41
Tabela 5. Resultados usando k-NN sem o uso de <i>Relevance Feedback</i> . *D é a diferença entre a última distância euclidiana de acerto e a primeira distância euclidiana de erro.	42
Tabela 6. Resultados usando k-NN com o uso de <i>Relevance Feedback</i> . *D é a diferença entre a última distância euclidiana de acerto e a primeira distância euclidiana de erro.	42
Tabela 7. Taxa de acerto usando o espaço de Hough para identificar a natureza do documento.	43
Tabela 8. Média de acertos com 5.000 épocas de treinamento e taxa de aprendizado 0,05.	49
Tabela 9. Evolução do Erro com 5.000 épocas de treinamento e taxa de aprendizado 0,05.	49
Tabela 10. Média de acertos com 5.000 épocas de treinamento e taxa de aprendizado 0,1.	50
Tabela 11. Evolução do Erro com 5.000 épocas de treinamento e taxa de aprendizado 0,1.	50
Tabela 12. Média de acertos com 10.000 épocas de treinamento e taxa de aprendizado 0,1.	50
Tabela 13. Evolução do Erro com 10.000 épocas de treinamento e taxa de aprendizado 0,1.	51
Tabela 14. Média de acertos com 10.000 épocas de treinamento e taxa de aprendizado 0,05.	51
Tabela 15. Evolução do Erro com 10.000 épocas de treinamento e taxa de aprendizado 0,05.	51

Tabela de Símbolos e Siglas

(Dispostos por ordem de aparição)

OCR – *Optical Character Recognition*

CBIR – *Content Based Image Retrieval*

RGB – *Red, Green, Blue*

CMY – *Cyan, Magenta, Yellow*

YIQ – Um modelo de cor usado na transmissão de sinais para televisão. As Cores são separadas em luminância (Y) e 2 valores de cores (I e Q).

HSV – *Hue Saturation Value*

TBIR – *Texture Based Image Retrieval*

GLCM – *Grey Level Co-Ocurrence Matrix*

k- NN – *k - Nearest Neighbors*

NN – *Nearest Neighbors*

MLP – *MultiLayer Perceptron*

MSE – *Mean Square Error*

RBf - DDA – *Radial Basis Function – Dynamic Decay Adjustment*

Capítulo 1

Introdução

Com o significativo aumento na aquisição de equipamentos de digitalização, documentos em formato de imagem digital se tornaram cada vez mais rotineiros e seu uso conveniente. As informações dos documentos que antes ocupavam grandes instalações físicas passaram a ser tratadas como mais uma entidade armazenável em computador originando as chamadas bibliotecas digitais. Dessa forma, imagens de documentos se tornaram cada vez mais úteis, freqüentes e valiosas. Entretanto, junto com a facilidade de ter uma biblioteca digital, tornou-se necessário manter algum tipo de mecanismo de indexação para posterior recuperação de dados. Executar esta indexação manualmente é uma tarefa laboriosa, sensível a interpretações subjetivas e sujeita a erros. Métodos automáticos de indexação e recuperação baseados em algum tipo de característica (cor, forma e textura) são interessantes, pois reduzem a intervenção humana e deixam o sistema mais robusto *i.e* capaz de receber qualquer natureza de imagens de documentos [1].

Partindo da premissa que documentos, de uma forma geral, ajudam a humanidade a manter um registro de ocorrências que promovem a cultura, democratização e a cidadania, a recuperação de documentos baseada em conteúdo se propõe a ser uma ferramenta imprescindível na busca por informações.

Muito mais que apenas recuperar artigos, um sistema de recuperação de documentos baseada em conteúdo permite filtrar os registros mais similares fundamentado nas características inerentes de uma imagem de entrada. Isso significa que um documento em formato digital é apresentado ao sistema e outras imagens similares são recuperadas. Assim, o sistema permite que o usuário não tenha conhecimento prévio sobre qualquer tipo de parâmetro de busca [2]. A recuperação de documentos em meio digital significa não só uma alternativa para uma busca mais eficiente da informação disponibilizada em larga escala, mas também agrega à informação um

valor social propiciado pelo seu uso por uma faixa muito mais larga de usuários, o que, certamente, irá estimular as pesquisas históricas, econômicas e sociais.

Por exemplo, um sistema de recuperação de documentos baseado em conteúdo pode receber uma logomarca de um determinado banco e recuperar da base de dados todos os cheques que são daquele banco. O mesmo acontece para documentos com logomarcas ou cartão postais.

Os principais objetivos desse trabalho consistem em analisar e desenvolver técnicas robustas de recuperação de imagens baseadas em conteúdo aplicadas ao domínio de imagens de documentos sem o uso de *OCR (Optical Character Recognition)*. Com isso, almejamos implementar uma ferramenta que possa comprovar a influência dos diferentes tipos de características de imagens e quais suas nuances dentro do escopo de um sistema de busca de imagens. Propomos também uma abordagem de recuperação de documentos com ou sem figuras em seu conteúdo; recuperação de documentos manuscritos ou datilografados; recuperação de cheques de um determinado banco usando segmentação automática do logo para extração de características.

1.1 Organização do Trabalho

Este trabalho está organizado em 5 capítulos. Neste Capítulo 1 estão expostas as motivações e objetivos do trabalho. No Capítulo 2 são apresentados alguns dos principais aspectos relativos à recuperação de imagens baseada em conteúdo (*CBIR*), abrangendo os principais conceitos sobre tais sistemas, além de considerações sobre o processo de extração de características de imagens.

O Capítulo 3 faz um estudo sobre o estado da arte de sistemas de recuperação de documentos e uma análise geral sobre os classificadores de documentos. São expostos métodos para se descrever computacionalmente documentos e quais técnicas são importantes para recuperação de imagens de documentos.

No Capítulo 4 é mostrado um conjunto de experimentos e resultados relacionados à implementação de uma ferramenta que visa buscar imagens de documentos com base na segmentação automática de possíveis figuras que o documento possa conter. Também são discutidos os resultados obtidos por diferentes técnicas de recuperação, como: vizinho mais próximo com e sem *relevance feedback* e o uso de redes neurais de aprendizado supervisionado. Por fim, o Capítulo 5 expõe conclusões e discute propostas interessantes de trabalhos futuros a esta monografia.

Capítulo 2

Recuperação de Imagens baseada em Conteúdo - CBIR

Sistemas de recuperação de imagens baseada em conteúdo são ferramentas úteis para buscar imagens através de medidas que estão representadas nas próprias imagens e que podem ser computadas de forma a gerar uma interpretação representativa sobre elas. A recuperação de imagens baseada em conteúdo é a busca por imagens com base em características visuais que servem de parâmetros para proporcionar uma procura eficiente e bastante amigável do ponto de vista do usuário já que apenas uma imagem é fornecida como parâmetro para o sistema. As razões para o seu desenvolvimento e sucesso na área de reconhecimento de padrões consistem na ineficiência de métodos tradicionais de busca textuais que provaram ser insuficientes, trabalhosas e extremamente caras computacionalmente, pois demandam um tempo de processamento elevado em comparação com os atuais métodos automáticos de busca por imagens [3]. Tais métodos que indexam e associam uma determinada imagem a uma categoria usando uma palavra-chave ou um índice tem se tornado obsoleto. Em um sistema *CBIR* cada imagem do repositório tem suas características extraídas e comparadas às características da imagem de entrada ou muitas vezes chamada de imagem teste.

Atualmente, com o aumento considerável das bases de dados de imagem e vídeo devido às inovações tecnológicas, tornou-se necessário manter algum tipo de mecanismo de indexação para posterior recuperação de dados. Partindo dessa premissa, a recuperação de imagens de forma automática pode ser uma ferramenta útil para tornar viável a indexação e consulta a grandes coleções de imagens, fotos e documentos digitalizados. A indexação automática, para esses

objetos, proporciona um aumento expressivo na produtividade da busca. Dessa forma, sistemas *CBIR*, por serem métodos automáticos de indexação e recuperação baseada em algum tipo de característica, são interessantes, pois reduzem a intervenção humana e tornam o sistema robusto para qualquer que seja o elemento de busca. A idéia de recuperação de imagens através de outras imagens é uma extensão da idéia de busca de páginas web que contenham no texto as palavras-chave especificadas na pesquisa, por exemplo, Google, Altavista e outros. No caso de imagens, busca-se recuperar em uma base de dados de imagens aquelas que contenham o elemento gráfico procurado como, por exemplo, uma determinada cor, textura e forma. São essas as características visuais de baixo nível que tornaram possível o desenvolvimento de um sistema automático de indexação e recuperação de imagens. Os sistemas *CBIR* utilizam esses elementos ou de suas combinações para filtrar as imagens mais similares dentre um repositório de imagens.

Em um sistema de recuperação por conteúdo algumas etapas são importantes para um eficiente funcionamento de todo mecanismo de busca. O sistema recebe como entrada apenas uma imagem teste que é o elemento de busca a ser procurado em um repositório de imagens que também conterá o resultado da recuperação por conteúdo. Logo após a seleção da imagem de entrada, as características são extraídas com base nos atributos visuais de baixo nível que a imagem contém. Essas características são expressas numericamente e agrupadas em vetores para permitir que classificadores sejam utilizados. Durante a fase de processamento, um conjunto de algoritmos trata de avaliar as imagens mais similares e retorna para o usuário as imagens recuperadas. É na fase de saída que se observa a precisão de todo o conjunto implementado. Observações sobre a semântica de cada imagem recuperada podem ser utilizadas para refinar toda a estratégia de desenvolvimento do sistema. A Figura 1 ilustra como se dá o relacionamento entre as diferentes fases de um projeto de recuperação de imagem baseada em conteúdo.

Para facilitar a realização de consultas e buscas, as imagens podem ser organizadas em estruturas de dados. Características de imagens são representações computacionais e podem não ter uma ordem pré-fixada e ter múltiplos atributos inter-relacionados. Por isso, estruturas de dados flexíveis podem também ser usadas para facilitar o armazenamento e a recuperação em um sistema de recuperação de imagens. Cada estrutura tem vantagens e desvantagens; algumas têm domínios limitados e algumas podem ser utilizadas concorrentemente com outras. Em sistemas de informação visual, a interação do usuário é muito importante em muitas das funções. Há estratégias que utilizam da subjetividade humana para refinamento da busca. Tais estratégias são chamadas de *relevance feedback* [4] .

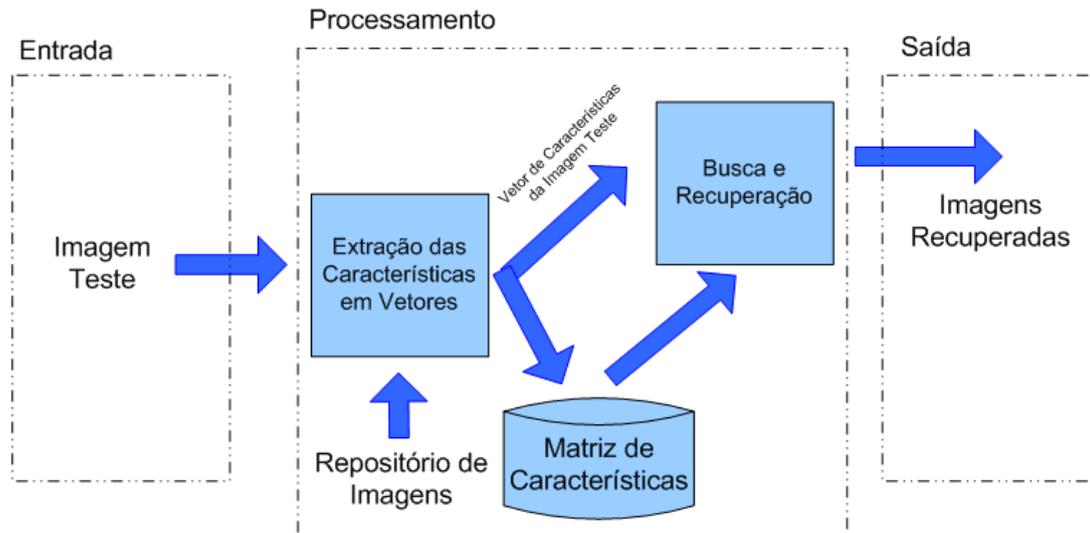


Figura 1. Fluxograma de funcionamento básico de um *CBIR*.

A interface do usuário de um sistema *CBIR* consiste de um processador de consultas e um navegador para proporcionar ferramentas gráficas interativas e mecanismos para consulta e navegação na base de imagens. Mecanismos de consultas convencionais proporcionados pelas interfaces com o usuário são: consulta por palavras chave, por esboço (*sketching*), por exemplo (*Query by example*), navegação por categorias e seleção de características.

Os atuais métodos usados por sistemas *CBIR* são os que agrupam fisicamente as imagens semanticamente relacionadas em um espaço de características visuais. Utilizando características visuais, a recuperação pode ser realizada pela análise das imagens na vizinhança de um elemento de busca no espaço de características. Tal agrupamento pode servir para muitas imagens na base de imagens independentemente se elas forem formadas por fotos de paisagem, imagens geradas através da computação numérica ou pinturas de artistas. Imagens com semânticas diferentes podem, eventualmente, representar uma a outra, pois a amostragem pode ter sido realizada em série, em um mesmo tempo ou pintadas por um mesmo artista em um mesmo período de tempo. A Figura 2 demonstra que fotos de flores capturadas em série em um mesmo período de tempo e lugar podem resultar em características visuais bastante similares como iluminação e composição de cores de cada amostra [5].



Figura 2. Imagens semanticamente relacionadas referenciam umas às outras por características visuais.

Entretanto, nem todas as imagens que possuem a mesma semântica se referenciam quando são representadas por atributos visuais. Imagens semanticamente relacionadas podem ser distribuídas e rotuladas em um mesmo espaço de características. A Figura 3 demonstra que flores podem ser bastante diferentes quando o histograma de cor é utilizado para representar as amostras.

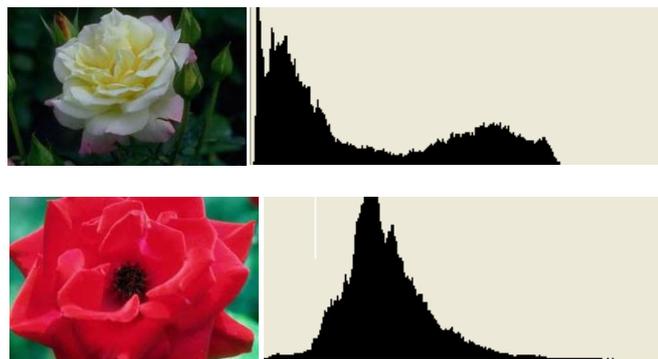


Figura 3. Imagens semanticamente iguais, mas diferentemente representadas pelo histograma de cor.

Apesar de inúmeras vantagens, a recuperação por elementos semânticos tem inúmeras limitações. Uma principal dificuldade está no fato de que a maioria das imagens possui múltiplas interpretações semânticas, ou seja, imagens do mesmo objeto podem ser interpretadas como diferentes objetos. Uma alternativa para contornar esse problema é utilizar várias características, pois imagens similares tendem a ter mais características parecidas do que distintas.

Nas atuais abordagens *CBIR*, assume-se que existe apenas uma classe para cada objeto em questão. Dessa forma, os objetos mesmo que semanticamente diferentes pertencem a apenas uma classe. Por exemplo, as flores ilustradas na Figura 3 pertencerão à mesma classe apesar de serem tipos diferentes de flores.

A principal abordagem para a recuperação de imagens baseada em conteúdo é o que usa exemplos como elementos de busca. Busca por exemplos (*Query by Examples*) rotula uma base de imagens por similaridade com relação à outra imagem fornecida pelo usuário. O sistema é capaz de extrair elementos que descrevem computacionalmente os dados envolvidos na busca, compara com as outras descrições da base de dados e retorna as comparações mais próximas [6]. Na abordagem por exemplos, o sistema apresenta miniaturas de imagens representativas para o usuário.

2.1 Características de Imagens utilizadas nos Sistemas CBIR

Os sistemas *CBIR* podem ser classificados de maneira bem abrangente em duas diferentes categorias: os que utilizam características de baixo nível e os que utilizam características semânticas de alto nível. As características de baixo nível são representações mais gerais e calculadas através da distribuição de pixels na imagem enquanto que as características semânticas são atributos abstratos oriundos, por exemplo, da relevância das cenas graficamente representadas nas imagens. Cor, forma e textura são os três principais grupos de características de baixo nível que são usadas pelos sistemas *CBIR*.

2.1.1 Características de Cor

Dentre as características de baixo nível de imagens, as variações de cores assumem destacada importância, por permitirem a identificação de detalhes de matiz e de regiões dos dados a serem analisados.

Cor é uma propriedade relacionada com a capacidade dos objetos de refletir ondas eletromagnéticas de diferentes comprimentos, num espectro visível que varia aproximadamente de 400 nm até 700 nm. A fim de se mapear numericamente o conjunto de cores do espectro visível, diversos modelos (ou espaços de cores) têm sido propostos e utilizados, cada qual com suas vantagens e aplicações específicas como o RGB (*red, green, blue*), o CMY (*cyan, magenta, yellow*), o YIQ e o HSV (*hue, saturation, value*).

O modelo de cor HSV (*hue, saturation, value*) utiliza descrições de cores que são mais intuitivas do que combinações de cores primárias (RGB) e, por isso, é mais adequado para ser usado na especificação de cores em nível de interface com o usuário.

2.1.2 Características de Forma

A extração dos atributos de forma de um objeto é desafiadora para a criação de um sistema eficiente de *CBIR*. Na recuperação de imagens, as descrições de forma devem ser invariantes à escala, rotação e translação do objeto. Os atributos de forma são bem menos desenvolvidos que os de cor e textura, pois a complexidade de representação das formas dos objetos representados nas imagens exige um estudo mais preciso de como os objetos estão distribuídos na imagem. As regiões ocupadas pelos objetos na imagem têm que ser achados para se obter uma descrição

representativa da forma de um objeto. Muitas vezes é complexo particionar precisamente uma imagem em regiões significativas, usando características de baixo nível devido à variedade de como um objeto 3D poderá ser representado em duas dimensões. Para uma representatividade eficiente devem ser consideradas a complexidade dos objetos contidos nas imagens, a presença de sombras e iluminação não uniforme no momento da amostragem [7].

2.1.3 Características de Textura

Dentre os principais atributos visuais presentes em uma imagem encontra-se a textura. Esse atributo está diretamente relacionado à distribuição e organização dos pixels de uma determinada região da imagem, atuando dessa maneira como uma excelente ferramenta para segmentar regiões ou ainda, como medida a ser utilizada no reconhecimento de padrões [8].

Uma textura se caracteriza pela repetição de um padrão sobre uma região, sendo esse padrão repetido em sua forma exata ou com pequenas variações. Através de sua análise é possível distinguir regiões que apresentem as mesmas características e portando mesmas cores em determinada combinação de bandas. Isso torna a textura um excelente descritor regional, contribuindo para uma melhor precisão dos processos de reconhecimento, descrição e classificação de imagens [8].

A análise de uma imagem em termos de suas texturas permite identificar regiões através desse padrão de cores. Com a análise de texturas é possível se ter uma medida geral de como ocorre o relacionamento entre os pixels de uma dada região.

A área de visão computacional lida com a classificação das texturas presentes em uma imagem ou na identificação de texturas semelhantes de um banco de imagens, originando os sistemas de recuperação de texturas baseada em conteúdo (TBIR - *Texture Based Image Retrieval*).

2.2 Descritores de Texturas

Gonzalez e Woods [9] citam a existência de três principais abordagens para se descrever texturas: a abordagem espectral, a estrutural e a estatística. A abordagem espectral diz respeito a propriedades baseadas no espectro de Fourier, onde pode ser detectada a existência de padrões periódicos ou semi-periódicos. A abordagem estrutural representa textura como sendo formada pela repetição de padrões que obedeçam alguma regra de posicionamento para a sua geração. Já

na abordagem estatística, medidas são utilizadas para descrever texturas. Esta última abordagem é a mais comumente usada e geralmente recomendada para texturas que não apresentam boa regularidade.

A abordagem estatística de descrição de texturas considera o relacionamento, a distribuição e o inter-relacionamento entre os tons de cinza. Os principais descritores que utilizam essa abordagem são a matriz de co-ocorrência de níveis de cinza (*Grey Level Co-Occurrence Matrix - GLCM*) e o Histograma da Soma/Diferença (*Sum/Difference Histogram*), brevemente descritos a seguir.

2.2.1 Matriz de Co-Ocorrência de Tons de Cinza (*GLCM – Gray Level Co-occurrence Matrix*)

A matriz de co-ocorrência de níveis de cinza (*Grey Level Co-Occurrence Matrix - GLCM*) [11] leva em consideração a posição de pixels que possuem níveis de cinza. Uma matriz de co-ocorrência é uma tabulação sobre o número de combinações distintas de valores de intensidade que ocorrem em uma imagem [12].

Para uma dada imagem, cada célula (i, j) da matriz de co-ocorrência funciona como um contador e armazena a frequência com que dois pixels ocorrem na imagem separados por uma distância d , um com a cor i e outro com a cor j , como mostrado na Figura 4. Nessa Figura, temos, à esquerda, a matriz de pixels de uma imagem de exemplo onde aparecem em duas regiões da imagem as cores 1 e 2 separadas por um pixel de distância. Assim, se $d = 1$, a matriz GLCM tem sua posição $(1, 2)$ incrementada para 2, indicando a ocorrência de 2 pixels com cor 1 e 2 separados pela distância d . Dessa forma, cada elemento da GLCM é um contador de quantas vezes a ocorrência entre $d + 1$ cores consecutivas acontece quando a imagem I é percorrida horizontalmente. Quando não há ocorrência entre duas cores é atribuído ao elemento da matriz o valor 0.

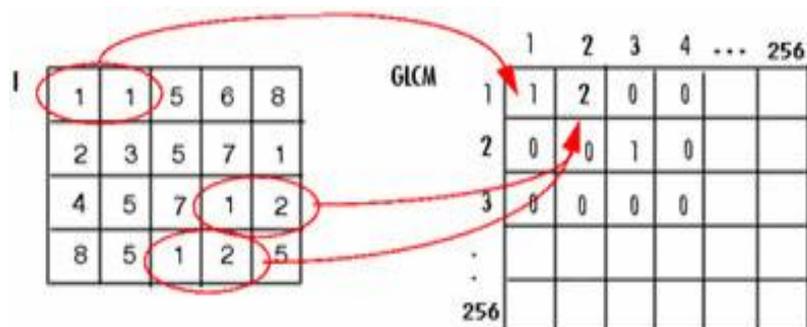


Figura 4. À direita, a matriz de co-ocorrência é preenchida indicando quantas vezes a cor i e a cor j estão presentes na imagem I (à esquerda) separados por uma distância d igual a 1 [12].

O algoritmo da GLCM produz uma matriz com $C \times C$ elementos, onde C é o número de diferentes tons que a imagem pode ter. Com uma imagem de entrada em tons de cinza, a matriz possui dimensões de 256×256 pixels. A geração da matriz GLCM é um processo computacionalmente muito custoso, pois tem que varrer toda a imagem para calcular as ocorrências entre os diferentes tons.

2.2.2 Histograma da Soma/Diferença (*Sum/Difference Histogram*)

De forma análoga à GLCM, o Histograma da Soma/Diferença observa o relacionamento entre cada elemento da matriz de entrada (imagem) e seus vizinhos separados por uma distância. O Histograma da Soma utiliza duas distâncias d_x e d_y . O algoritmo soma todos os pixels separados por d_x e d_y . Como ilustrado na Figura 5, o histograma da soma varre a imagem e, para cada pixel na posição (i, j) , a sua cor é somada a do pixel da posição $(i + d_x, j + d_y)$ e a cor correspondente é incrementada no histograma. Com isso, em uma imagem com 256 tons de cinza, o histograma da soma vai de 0 a 510 (resultado da soma de 255 com 255).

Analogamente, o histograma da diferença calcula a diferença entre os pixels, variando seus valores de -255 a 255. Nesse caso, há uma clara redução no espaço necessário para armazenamento do histograma em comparação com a GLCM. Para imagens com 256 cores, um vetor com 511 posições é suficiente para representar o Histograma da Soma/Diferença. No entanto, em geral, a GLCM apresenta uma informação mais precisa sobre a distribuição de cores na textura original [12].

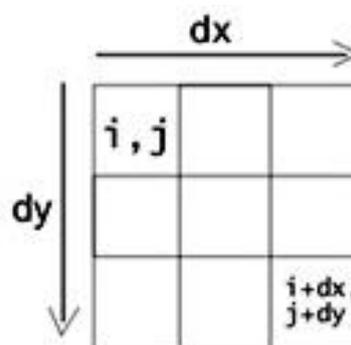


Figura 5. Exemplo de varredura da imagem para geração do histograma da soma com deslocamento d_x e d_y .

Capítulo 3

Recuperação de Documentos baseada em Conteúdo - DIR

A busca e recuperação de grandes coleções de imagens de documentos é uma tarefa desafiadora, especialmente quando existem representações não textuais disponíveis. Para formular e implementar um eficiente sistema de recuperação de documento é necessário buscar imagens usando técnicas que levem em consideração várias coleções de documentos com diferentes qualidades, estilos, tamanhos e fontes.

Acervos digitais são extremamente pobres com relação à qualidade de amostragem dos documentos. Muitos possuem ruídos de poeira em excesso, gotas de tinta agregadas às letras, cortes arbitrários devido à qualidade do papel, degradação do texto impresso devido à má qualidade do papel ou tinta e ainda ruídos originados pela escrita no verso dos documentos.

As imagens dos documentos em bibliotecas digitais estão em diversos idiomas. Por isso, um mecanismo que permita recuperar documentos livre dos significados textuais presentes no texto é interessante, pois permite uma robustez para qualquer que seja o idioma usado na escrita do documento.

Procurar em grandes coleções de imagens de documentos exige uma seqüência de passos bem definida: pré-processamento das imagens, extração de características, comparação e busca de documentos relevantes. Cada um desses passos pode ser computacionalmente caro. Normalmente, um livro pode conter 9.000 palavras e processar todas elas em tempo de execução é inviável. Assim, estratégias de processamento *offline* devem ser utilizadas para diminuir o tempo de busca [13].

A recuperação de documentos baseada em conteúdo (DIR - *Document Image Retrieval*) se apresenta como um caso especial de recuperação de imagens baseada em conteúdo em que o elemento de busca e toda a base de imagens passam a ser documentos digitalizados. Grande parte

das técnicas envolvidas na classificação de imagens é usada também para os documentos sendo que em tal aplicação encontram-se principalmente dificuldades na extração de características, pois a maioria dos documentos é tratada computacionalmente como imagens binárias, *i.e.*, todos os pixels assumem valores 0 ou 1.

A classificação do documento é uma importante tarefa no processo de recuperação de documentos digitais. Ela permite a distribuição e arquivamento automático de acervos e melhora a eficiência da indexação na construção de bibliotecas digitais [14].

3.1 Componentes de um Classificador de Documentos

Existe uma grande diversidade de classificadores de documentos que resolvem uma variedade de problemas de classificação. Tais classificadores diferem principalmente na forma em que os elementos da base são usados para construir modelos de classes de documentos, na escolha de características de documentos e nos algoritmos de reconhecimento.

Os três principais componentes que modelam um classificador de documentos consistem no problema de classificação que é o escopo da complexidade de recuperação de documentos, a arquitetura do classificador e o desempenho dos resultados oriundos da classificação. A Figura 6 mostra de forma hierárquica, genérica e organizada os três principais componentes de um sistema de recuperação de documentos.

3.1.1 Classificação

O problema de classificação consiste na complexidade imposta pelos documentos a serem recuperados por um sistema de busca de documentos. Ele consiste de dois principais aspectos: Os documentos de entrada e o conjunto de classes de documentos que delimitam o escopo do sistema. Os documentos de entrada são aqueles que serão utilizados como base para a busca e com base nela que outros documentos serão recuperados. O conjunto de classes define as possíveis saídas produzidas por um classificador e é usada para rotular os documentos. A maioria dos classificadores utiliza classes definidas pelo usuário baseada na similaridade do conteúdo, forma ou estilo (similaridade visual).

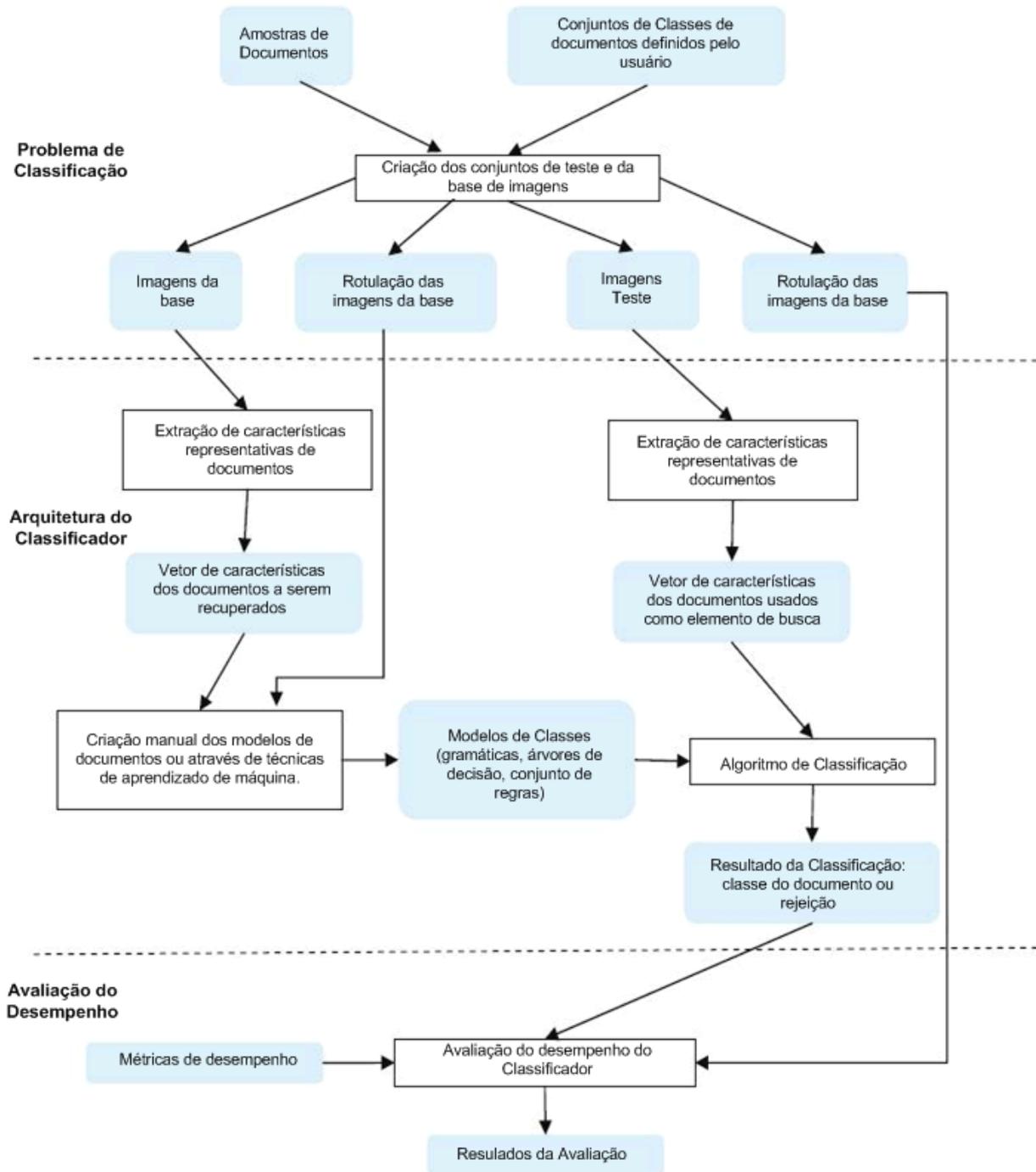


Figura 6. Os três principais componentes de um sistema de recuperação de documentos.

3.1.2 Arquitetura do Classificador

Os quatro principais aspectos para caracterizar uma arquitetura de classificador são as características de documentos, representação de características, modelos de classes e mecanismos de aprendizado [14]. Tais aspectos estão fortemente interligados, ou seja, decisões na formulação

de cada um desses aspectos influenciam diretamente no desempenho do conjunto como um todo. Por exemplo, se os documentos forem representados por um vetor de características de tamanho fixo então modelos estatísticos e algoritmos de classificação são recomendados. A escolha das características a serem abordadas no vetor de características é uma etapa importante no resultado retornado pelo classificador. Algumas abordagens de recuperação de documentos utilizam OCR (*Optical Character Recognition*) [16]. É de praxe o uso de imagens de documentos binarizadas para extração de características a fim de obter uma segmentação mais eficiente de páginas ou eventuais figuras que o documento possa conter [17][18]. A transformação para o documento preto-e-branco facilita a extração de características estruturais produzidas por uma análise lógica e física de como as informações estão dispostas no *background* [19][20]. Inevitavelmente a binarização da imagem causa uma grande perda de informações úteis para representação do documento em um vetor de características.

A Figura 7 mostra uma seqüência clássica do processo de reconhecimento de documentos em que a presença de texto é dominante [20].

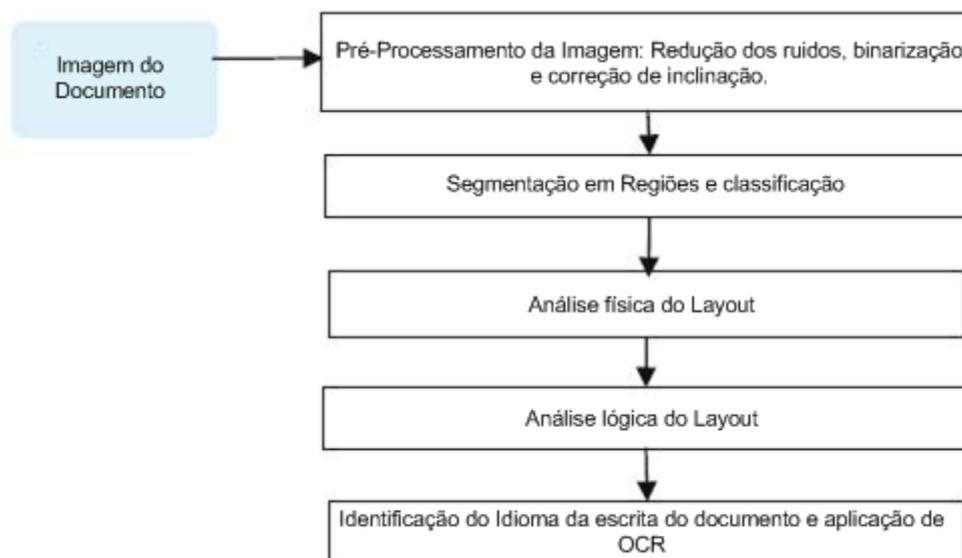


Figura 7. Seqüência clássica do reconhecimento de documentos onde a presença de texto é dominante.

A segmentação em regiões identifica os blocos que estão no documento e que formam porções bem definidas de conteúdo como texto, tabelas e figuras. A análise física do *layout*, também chamada de análise estrutural extrai a estrutura de como as informações estão dispostas

no documento em formato de imagem, além de fornecer uma descrição hierárquica dos objetos presentes baseada na organização geométrica dos componentes de uma imagem de documento [20]. A análise lógica de layout também chamada de rotulação lógica extrai a estrutura coerente dos objetos distribuídos por toda imagem do documento. Por exemplo, a estrutura lógica de uma página de artigo é a hierarquia dos objetos presentes de tal forma que é possível inferir, por exemplo, que os autores são descritos logo após o título, pois esse é padrão adotado para trabalhos científicos [18]. Ou seja, é a ordem pré-estabelecida em que os diversos elementos de conteúdo devem aparecer durante a análise do documento. Existem três principais categorias que definem as características que podem ser extraídas de imagens de documentos: Características de imagens, estruturais e textuais.

Características de imagens são obtidas diretamente da imagem, por exemplo, utilizando a densidade de pontos pretos em uma região extraída de uma imagem segmentada ou analisando o número de linhas horizontais ao decorrer do documento. Quando são extraídas características da imagem como um todo, as características consideradas são globais e ao se analisar regiões específicas representações locais podem ser extraídas. Técnicas e sistemas avançados de recuperação de documentos baseada em conteúdo exigem uma combinação conveniente dessas duas abordagens para que o vetor de característica represente de forma computacional aquilo que semanticamente entende-se como um documento.

As características estruturais são originadas da relação entre os objetos presentes na página e são extraídas das análises física ou lógica do *layout* do documento.

As características textuais tiram proveito da presença de palavras contidas no texto e podem ser computadas através da aplicação de algoritmos de reconhecimento ótico de caracteres. A combinação dessas características deve ser ponderada de acordo com o escopo na qual o documento aborda. Muitas vezes, o uso de apenas uma das categorias de características é suficiente para um resultado expressivo. Há de se analisar, através de resultados empíricos, o tempo computacional e taxa de acerto do sistema para chegar a uma estratégia ótima [21]. Alguns classificadores usam apenas características de imagens, estruturais ou somente textuais; outros utilizam uma combinação deles. Cada uma das categorias tem suas vantagens e desvantagens.

As características lógicas e estruturais são bastante caras computacionalmente, pois precisam de um modelo de domínio lógico específico para cada tipo de documento. Os primeiros sistemas usavam modelos lógicos atribuídos manualmente a cada classe em questão [23].

Atualmente é usada uma estratégia de aprendizagem de tais modelos automaticamente dos padrões rotulados que são apresentados ao sistema [23]. Entretanto, a rotulação de documentos é

extremamente laboriosa já que os significados lógicos devem ser atribuídos aos objetos que se encontram no *layout* físico de cada documento utilizado para extrair características.

A utilização das características essencialmente textuais como frequência e relevância de cada palavra chave ou termos podem ser usadas independentes ou em conjunto com outras características de imagens [24]. As características textuais podem ser extraídas do resultado da aplicação de algoritmos *OCR* os quais podem ser bastante ruidosos e podem desviar informações importantes para o processo de distinção entre outros documentos [24].

3.1.3 Desempenho dos Resultados

O desempenho dos resultados é um componente crítico e importante para um classificador de documentos. Com ele é possível avaliar a robustez dos resultados e as medidas para comparação de múltiplos classificadores. A maior parte dos sistemas de classificação mede a efetividade dos classificadores que é de fato, a habilidade de tomar as decisões de classificação corretas [25]. Várias medidas são usadas para se avaliar o desempenho da classificação como precisão, taxa de correção, taxa de reconhecimento, taxa de erro, taxa de falsos, taxa de rejeitados, *recall* e precisão.

Os classificadores de documentos são costumeiramente bastante difíceis de comparar, pois eles resolvem diferentes problemas com diversas representações computacionais e usando distintos conjuntos de classes com possíveis saídas. Por exemplo, é difícil comparar um classificador que manipula documentos de *layout* fixo (questionários) com um que possui *layout* variável (jornais ou artigos) [26]. Outro enigma é que a quantidade de classes de documentos varia extremamente de acordo com a finalidade a que se propõe o classificador. Por exemplo, nas páginas de uma revista pode haver várias classes de documentos ou a revista, em si, pode ser uma única classe.

3.2 Extração das características de documentos

Para um sistema de recuperação de documentos baseada em conteúdo a extração de características tem um papel fundamental para a representação das imagens dos documentos. Durante o desenvolvimento de um *software* para busca de imagens é de extrema importância que as características sejam não apenas robustas para diferenciar os diversos tipos de ruídos de imagens, mas também aumentem o desempenho dos cálculos feitos durante a classificação.

3.2.1 Pré-processamento das imagens

De uma forma geral, as imagens de um documento podem estar inclinadas e sofrer variações que podem deteriorar as características a serem extraídas. Por isso, documentos digitalizados devem ser primeiramente binarizados e suas inclinações, se houver, têm que ser retiradas.

3.2.2 Histograma da projeção

O Histograma da projeção [21] das imagens dos documentos é útil para distinção entre dois documentos com diferentes estruturas. A técnica de projeção da imagem consiste em contabilizar a quantidade de pixels ao longo de cada linha vertical e horizontal originando um histograma horizontal e outro vertical denotados respectivamente por H_{proj} e V_{proj} . Tais projeções representam a informação contida no documento que contém algum significado semântico, ou seja, as informações que estão no *foreground*. Os histogramas da projeção tanto vertical como horizontal estão ilustrados na Figura 8. Após essa primeira etapa os histogramas devem ser normalizados a fim de diminuir a distribuição dos dados e manter uma escala de trabalho bem definida [22]. A normalização do histograma da projeção é realizada dividindo cada um de seus elementos pela soma total dos elementos, *i.e.*,

$$h_i \leftarrow \frac{h_i}{\sum_i h_i}, \quad v_i \leftarrow \frac{v_i}{\sum_i v_i} \quad (1)$$

Onde $H_{proj} = \{h_1, h_2, \dots, h_m\}$ e $V_{proj} = \{v_1, v_2, \dots, v_n\}$

É comum que durante o processo de digitalização os documentos sejam digitalizados em diferentes sentidos. Para duas imagens digitalizadas do mesmo documento, mas em diferentes sentidos há uma grande variação na computação do histograma da projeção do documento. Por isso, é necessário reverter os histogramas para eliminar tais variações [22]. Para identificar se um histograma necessita de uma reversão o sinal de sua inclinação é analisado. A inclinação de um histograma é definida por S :

$$S = \frac{\mu_3}{\sigma^3} \quad (2)$$

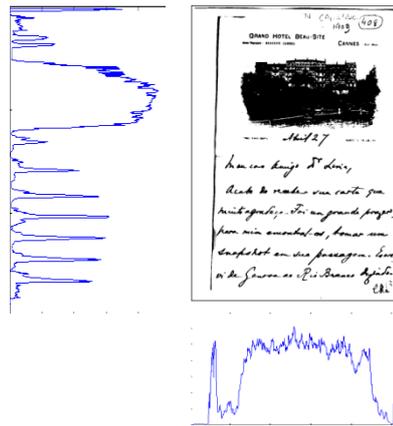


Figura 8. Histogramas da projeção (vertical e horizontal) de uma imagem de documento histórico.

Onde μ_3 é o momento central de terceira ordem de um histograma e σ é o desvio padrão. Se o sinal da inclinação do histograma for negativo então o histograma é revertido. Como última etapa, cada histograma é igualmente particionado em ρ partes e os valores de cada uma dessas partes são somados para resultar em um histograma reduzido. Ou seja:

$$\bar{H}_{proj} = \{\bar{h}_1, \bar{h}_2, \dots, \bar{h}_\rho\}, \bar{V}_{proj} = \{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_\rho\} \quad (3)$$

Finalmente um vetor é formado agrupando-se os dois histogramas reduzidos se a direção do documento for horizontal:

$$PHF = \{\bar{H}_{proj} | \bar{V}_{proj}\} = \{\bar{h}_1, \bar{h}_2, \dots, \bar{h}_\rho | \bar{v}_1, \bar{v}_2, \dots, \bar{v}_\rho\} \quad (4)$$

Senão,

$$PHF = \{\bar{V}_{proj} | \bar{H}_{proj}\} = \{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_\rho | \bar{h}_1, \bar{h}_2, \dots, \bar{h}_\rho\} \quad (5)$$

3.2.3 Mapas de Transição

Mapas de Transição [22] são uma característica que indica a quantidade de vezes que um pixel em uma imagem I varia entre *foreground* e *background* ao decorrer de linhas horizontais ou verticais. Mais especificamente, podem-se definir duas funções binárias $C_x(i, j)$ e $C_y(i, j)$ que tem valor um se interativamente o pixel imediatamente após o pixel atual for diferente.

$$C_x(i, j) = \begin{cases} 1, & \text{se } I(i, j) \neq I(i + 1, j) \\ 0, & \text{caso contrário} \end{cases} \quad (6)$$

$$C_y(i, j) = \begin{cases} 1, & \text{se } I(i, j) \neq I(i, j + 1) \\ 0, & \text{caso contrário} \end{cases} \quad (7)$$

Esse cálculo origina um histograma que pode ser utilizado para extrair características relevantes de um documento. A Figura 9 ilustra os histogramas da computação do Mapa de Transição de uma imagem de documento.



Figura 9. Histogramas que demonstram a computação do Mapa de Transição [22].

Após a obtenção dos dois histogramas H_{mt} (horizontal) e V_{mt} (vertical) a mesma técnica de particionamento usada pelo histograma da projeção é utilizada para transformar em respectivos histogramas reduzidos \bar{H}_{mt} e \bar{V}_{mt} gerando um vetor de características que se o documento for horizontal será denotado por um histograma do mapa de transição (HMT):

$$HMT = \{\bar{H}_{mt} | \bar{V}_{mt}\} = \{\bar{h}_1, \bar{h}_2, \dots, \bar{h}_\rho | \bar{v}_1, \bar{v}_2, \dots, \bar{v}_\rho\} \quad (8)$$

ou, se o documento for vertical por:

$$HMT = \{\bar{V}_{mt} | \bar{H}_{mt}\} = \{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_\rho | \bar{h}_1, \bar{h}_2, \dots, \bar{h}_\rho\} \quad (9)$$

3.3 Técnicas Utilizadas

As subseções a seguir descrevem as técnicas utilizadas para recuperar as imagens dos documentos.

3.3.1 Morfologia Matemática

A operação de abertura [26] é amplamente utilizada para eliminação de ruídos. Ela consiste de uma erosão seguida de uma dilatação. A razão para erodir a imagem é eliminar qualquer forma de ruído que o documento possa conter enquanto que a dilatação tenta recuperar um pouco da informação da figura que estava presente na imagem.

O algoritmo de erosão faz com que o pixel central do elemento estruturante case com o pixel a ser visitado na imagem original; se todos os pixels da região ativa estiverem ativos, o resultado final fica ativo. A erosão é o conjunto de todos os pixels c tal que o elemento estrutural B transladado por c corresponda a um conjunto de pixels pretos em A . Pode ser expresso por:

$$A \ominus B = \{c | (B)_c \subseteq A\} \quad (10)$$

De forma simples, a dilatação faz com que os objetos presentes no documento tornem-se mais largos, pois o pixel central do elemento estruturante casa com o pixel da imagem a ser processada; se qualquer pixel sob a região ativa tiver valor 1, então o pixel correspondente na imagem final assume valor 1 (branco) e se todos forem inativos o pixel na imagem final assume valor 0 (preto).

3.3.2 Técnica de Recuperação NN (*Nearest Neighbors*)

O algoritmo de classificação baseado no vizinho mais próximo (*Nearest Neighbors – NN*) é uma técnica amplamente empregada para reconhecer padrões. O seu funcionamento consiste em identificar os vizinhos mais próximos de um dado vetor entre o conjunto de vetores de treinamento [27].

O k-NN ou k-*Nearest Neighbors*, é conhecido como a regra de classificação dos k vizinhos mais próximos. É um método de classificação de padrões em que um dado padrão é dito pertencer a uma classe de acordo com a quantidade de vizinhos que pertençam a essa classe, conforme um critério de distância Euclidiana. A distância Euclidiana $d(x, y)$ entre dois padrões x e y é dada por:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (11)$$

Onde n é a quantidade total de características, x_i e y_i são, respectivamente, características em particular dos padrões x e y .

O algoritmo k-NN é uma técnica não paramétrica de classificação que tem se mostrada eficiente em reconhecimento de padrões baseado em estatísticas. Ela fornece alta precisão na classificação para uma distribuição incerta e anormal [27].

A fase de treinamento do k-NN consiste apenas no armazenamento de todos os padrões. Em seguida, quando um novo padrão precisa ser classificado, é calculada a distância Euclidiana dele para todos os padrões do conjunto de treinamento e encontram-se os k padrões de treinamento mais próximos do novo padrão. Esse novo padrão é classificado segundo a predominância de uma determinada classe entre os k padrões mais próximos.

A grande vantagem do k-NN, além da simplicidade, é que ele não possui processamento na fase de treinamento. Porém, sua maior desvantagem está em armazenar todos os dados na memória, uma vez que o vetor de atributos que define cada padrão é necessário para o cálculo da distância.

3.3.3 *Relevance Feedback*

Através de vários estudos na área de recuperação de informação (*Information Retrieval*) [29], observou-se que, geralmente, o usuário não consegue recuperar os objetos de seu interesse na primeira tentativa de consulta a um sistema de recuperação baseado em conteúdo. Geralmente o que ocorre é que este usuário faz uma consulta tentativa, e vai refinando-a, alterando-a, de acordo com os resultados que obtém. As consultas subseqüentes passam a retornar cada vez mais documentos relevantes ao usuário, pois ele vai contextualizando melhor a imagem que deseja procurar utilizando novas imagens, e retirando as imagens que desvirtuam sua consulta. Assim, são produzidas novas consultas teoricamente mais precisas e mais úteis (mais objetos relevantes são retornados). A técnica de *relevance feedback* [29] é uma técnica de refinamento, *i.e.* alteração, de uma busca inicial utilizando informações fornecidas pelo usuário sobre a relevância dos documentos previamente retornados (em uma consulta anterior) [29].

Portanto, a realimentação de relevância, do inglês *Relevance Feedback*, é um processo que busca tornar a consulta do usuário mais precisa, recuperando somente os itens relevantes e

descartando os irrelevantes. Usualmente, técnicas de *Relevance Feedback* utilizam a distância Euclidiana ponderada para tentar refinar a busca em sistemas de recuperação de documentos baseada em conteúdo. A distância Euclidiana ponderada $dp(x, y)$ entre dois padrões x e y é dada por:

$$dp(x, y) = \sqrt{\sum_{i=1}^n w_i (x_i - y_i)^2} \quad (12)$$

onde n é a quantidade de características e w_i é o peso atribuído a cada característica do padrão.

3.3.4 Rede Neural MLP como técnica de recuperação

A arquitetura das Redes MLP (*MultiLayer Perceptron*) consiste de uma camada de entrada, geralmente uma ou duas camadas escondidas e uma camada de saída. Cada camada tem um número de unidades de processamento (neurônios) que são inteiramente interconectadas através de conexões ponderadas com os neurônios da camada subsequente como mostrado na Figura 10.

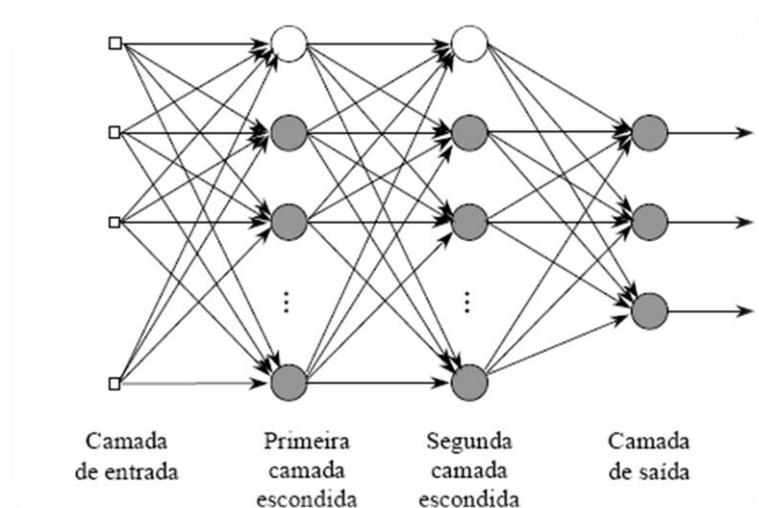


Figura 10. Arquitetura de uma rede MLP com duas camadas escondidas.

O MLP transforma n entradas em l saídas através de funções de ativação não lineares [30]. A saída da rede é determinada pela ativação dos neurônios na camada de saída através de

$$x_o = f\left(\sum_h x_h w_{ho}\right) \quad (13)$$

Onde $f()$ é a função de ativação do h -ésimo neurônio da camada escondida, x_h é a saída do h -ésimo neurônio da camada escondida e w_{ho} é o peso interconectado entre o h -ésimo neurônio da camada escondida e o o -ésimo neurônio da camada de saída. A função de ativação mais comumente usada é a sigmoideal que é expressa através de:

$$x_o = \frac{1}{1 + e^{(-\sum x_h w_{ho})}} \quad (14)$$

Baseado na diferença entre a saída calculada e a saída desejada, um erro é estimado através de:

$$E = \frac{1}{2} \sum_s^N \sum_o^L \left(t_o^{(s)} - x_o^{(s)}\right)^2 \quad (15)$$

Onde N é o número de padrões no conjunto de dados, L é o numero de neurônios na camada de saída dado pela quantidade de classes que estão envolvidas no problema, $t_o^{(s)}$ é a saída calculada e $x_o^{(s)}$ é a saída desejada. O objetivo é reduzir o erro através do ajuste das interconexões dos pesos entre as camadas. Os pesos são ajustados através do algoritmo *backpropagation* [31]. Esse algoritmo necessita de um conjunto de treinamento que consiste em um conjunto de atributos e saída desejada t_o , já que o MLP utiliza treinamento supervisionado. Durante o processo de treinamento, a rede neural começa com um conjunto aleatório de pesos iniciais e o treinamento continua até que um conjunto w_{ih} e w_{oh} seja otimizado de tal forma que um limiar de erro seja atingido entre x_o e t_o [32]. Esses valores ideais são calculados pela retro-propagação do erro ilustrado na Figura 11.

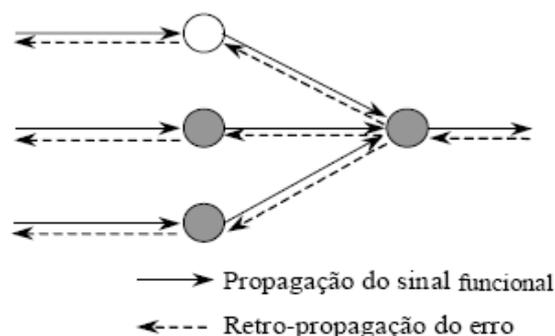


Figura 11. Ilustração das direções de propagação do sinal funcional (fase *forward*) e do erro (fase *backward*).

De acordo com o algoritmo *backpropagation* cada interconexão entre os neurônios é ajustada pela atualização dos valores dos pesos através de:

$$\Delta w_{ho} = -\eta \frac{\partial E}{\partial w_{ho}} = -\eta \delta_o x_h \quad (16)$$

$$\Delta w_{ih} = -\eta \frac{\partial E}{\partial w_{ih}} = -\eta \delta_h x_i \quad (17)$$

Onde E é a função do erro expressa em (15), $\delta_o = x_o'(t_o - x_o)$, $\delta_h = x_h' \sum_o \delta_o w_{ho}$ onde $x_o' = x_o(1 - x_o)$ e $x_h' = x_h(1 - x_h)$ quando a função de ativação sigmoideal é usada.

3.3.5 Transformada de Hough

A Transformada de Hough propõe uma abordagem alternativa para achar subconjuntos de pontos de uma imagem que possam estar alinhados em retas. Se considerarmos o ponto (x_i, y_i) e a equação $y_i = ax_i + b$ teremos infinitas linhas que passam por (x_i, y_i) , mas todas satisfazem a equação $y_i = ax_i + b$ para diferentes valores de a e b . Porém, escrever essa equação como $b = -x_i a + y_i$ e considerar o plano ab (também chamado de espaço de parâmetros) leva à equação de uma única linha dado um par (x_i, y_i) fixo. Além disso, um segundo ponto (x_j, y_j) também possui uma linha no espaço de parâmetros associado a ele, e essa linha intercepta a linha associada a (x_i, y_i) em (a', b') , em que a' é a inclinação e b' é o ponto de intersecção com eixo y da linha que contém (x_i, y_i) e (x_j, y_j) no plano xy .

A vantagem da Transformada de Hough advém da subdivisão do espaço de parâmetros nas chamadas células acumuladoras, como ilustrado na Figura 12 em que (a_{max}, a_{min}) e (b_{max}, b_{min}) são os domínios esperados dos valores da inclinação e do ponto de intersecção. A célula nas coordenadas (i, j) com o valor $A(i, j)$ do acumulador corresponde ao quadrado associado às coordenadas (a_i, b_j) . Inicialmente, essas células são inicializadas com zero. Então para cada ponto (x_k, y_k) no plano da imagem, é variado o parâmetro a dentro de cada valor da subdivisão do eixo a e calculamos o b correspondente usando a equação $b = -x_k a + y_k$. Os b resultantes são então arredondados para o valor permitido mais próximo no eixo b . No final desse

procedimento, um valor M em $A(i, j)$ corresponde a M pontos no plano xy que se localizam sobre a linha $y_i = a_i x + b_j$.

Um problema no uso da equação $y = ax + b$ na representação de linhas é que tanto a inclinação quanto o ponto de intersecção aproximam o infinito na medida em que a linha se torna vertical. Uma maneira de contornar essa dificuldade é através do uso da representação normal de uma linha:

$$x \cos \theta + y \sin \theta = \rho \tag{18}$$

O uso dessa representação na construção de uma tabela de acumuladores é idêntico ao do método discutido anteriormente. Porém, no lugar de linhas retas, curvas senoidais são formadas no plano $\rho\theta$. Como antes, M pontos colineares de uma linha $x \cos \theta_j + y \sin \theta_j = \rho_i$ levam M curvas senoidais que se interceptam em (ρ_i, θ_j) no espaço de parâmetros [9].

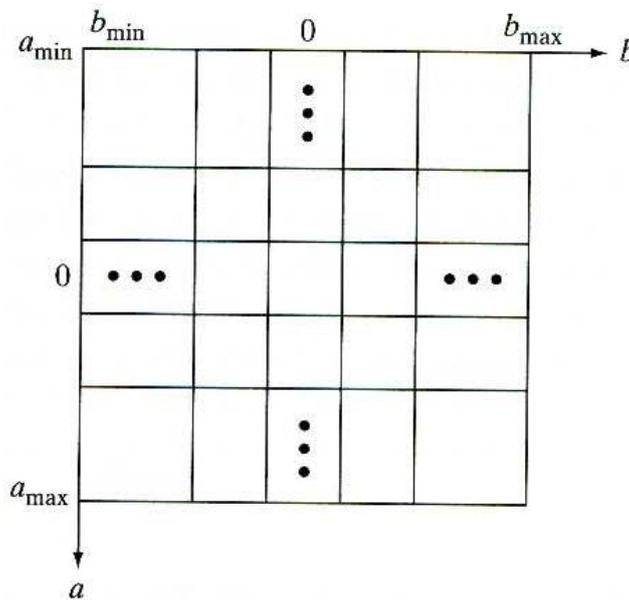


Figura 12. Quantização do plano de parâmetros para o uso na transformada de Hough.

Capítulo 4

Experimentos e Resultados

Neste capítulo é apresentada a metodologia utilizada para avaliar o desempenho de um sistema de recuperação de documentos baseada em conteúdo, um estudo do conjunto de características e das técnicas de recuperação utilizadas assim como uma análise global dos resultados. Os experimentos realizados neste trabalho consistem em desenvolver uma ferramenta que utiliza os principais conceitos de um sistema de recuperação de documentos baseada em conteúdo a fim de utilizar técnicas de reconhecimento de padrões para desenvolver um *software* que utiliza técnicas que pode ser de grande utilidade para acervos digitais, bibliotecas ou empresas que precisem de um mecanismo automatizado de busca de documentos livre de qualquer parâmetro de busca.

Neste trabalho propusemos um sistema de busca para grandes coleções de imagens de documentos. Foram desenvolvidas técnicas de segmentação e extração das figuras de uma imagem de documento assim como algoritmos para reconhecer se um dado documento é, de uma maneira geral, manuscrito ou datilografado. Com a integração dessas duas abordagens foi possível construir um ferramenta genérica o suficiente para receber documentos de várias naturezas.

A estratégia para obter o sistema final consiste de três grandes fases que se interligam de tal forma que o resultado de cada uma influencia diretamente na outra. Na primeira etapa são implementados algoritmos para descobrir se uma imagem de documento possui imagem ou não no seu conteúdo. Na segunda etapa, são identificados os documentos que são datilografados ou manuscritos. Na terceira e última fase são desenvolvidos algoritmos de recuperação para a busca de documentos que utilizam das duas primeiras etapas citadas anteriormente. Com essa abordagem, objetivamos a recuperação de documentos com ou sem imagens; recuperação de

documentos históricos manuscritos ou datilografados e recuperação de cheques de um determinado banco (dado um conjunto de imagens de cheques de diferentes bancos). Mais detalhes sobre a base de imagens são tratados na Seção 4.1.

O sistema de recuperação de documentos baseada em conteúdo proposto foi desenvolvido em basicamente duas principais camadas (*frontend* e *backend*). O *frontend* foi todo desenvolvido em *python 2.5* [33] e consiste dos algoritmos de recuperação utilizados para comparar técnicas de busca de imagens de documentos bem disseminadas na literatura e de uma interface gráfica que ajuda a visualizar os resultados dos testes mais claramente. O *backend* é todo o mecanismo implementado de extração de características e foi todo desenvolvido na linguagem de programação *Matlab®* [34]. O *frontend* do sistema utiliza todas as características extraídas do *backend* para geração de vetores a fim de usar k-NN e MLP como classificadores.

4.1 Base de Imagens de Documentos

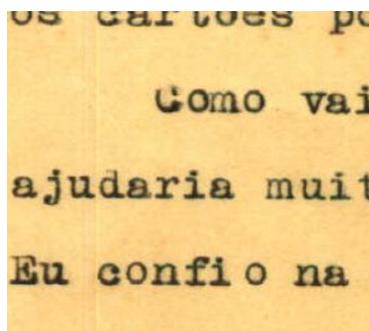
Para realização dos testes foram utilizadas imagens de documentos, como as ilustradas na Figura 13, de diferentes naturezas, como: Cartas manuscritas e datilografadas sem figuras e cheques. Ao todo foram utilizadas 227 imagens de diferentes resoluções e dimensões sendo que 200 delas foram imagens de cheques em tons de cinza de 4 diferentes instituições bancárias, 13 foram imagens de documentos datilografados e 14 cartas manuscritas sem figuras em seu conteúdo.

As inclinações das imagens dos documentos foram desconsideradas, pois isso acarretaria uso de técnicas que fugiam do escopo desse trabalho. Entretanto, todos os ruídos originais como rasuras de correção de erros de grafia, carimbos, escritas indevidas no documento, texto de rascunhos, rotulações providas do armazenamento e ruídos originados pela sujeira no vidro do scanner no momento da digitalização foram mantidos e considerados.

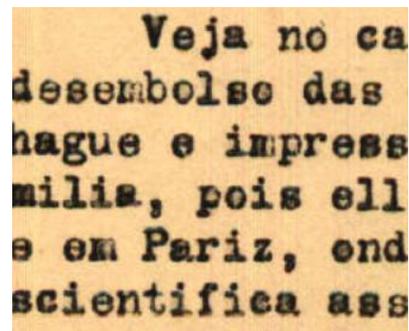


Figura 13. Exemplos da base de imagens de documento utilizada

A qualidade dos documentos existentes no repositório de imagens varia de amostra para amostra. Em alguns documentos datilografados podem ser observados traços mais grossos em alguns caracteres, fato que ocasiona união de caracteres vizinhos e deformação de alguns caracteres. Este problema está ilustrado na Figura 14, onde (b) apresenta traços mais grossos, caracteres unidos e deformados. Por outro lado, a Figura 14(a) apresenta um trecho de imagem com traços mais finos e algumas vezes não contínuos causando outros tipos de deformação.



(a)



(b)

Figura 14. Parte de dois documentos datilografados pertencentes ao repositório de imagens de documentos utilizado: (b) apresenta traços mais grossos que (a).

Este tipo de variação é comum em documentos digitalizados e por isso foi levada em consideração durante todo o desenvolvimento deste trabalho.

4.2 Metodologia e Resultados

Para experimentação e melhor visualização dos testes, a ferramenta desenvolvida foi composta por 3 principais módulos. Um módulo com as técnicas de recuperação, outro com a interface gráfica como ilustrado na Figura 15 e o terceiro com a extração de características que utiliza os atributos computados e armazenadas através de rotinas em Matlab®. Esses módulos foram desenvolvidos pensando em futuramente tornar a ferramenta expansível e bastante robusta para outras características e técnicas de recuperação. A Figura 16 mostra os diferentes módulos que integrados permitem que o sistema recupere imagens de documentos com base em outros documentos.

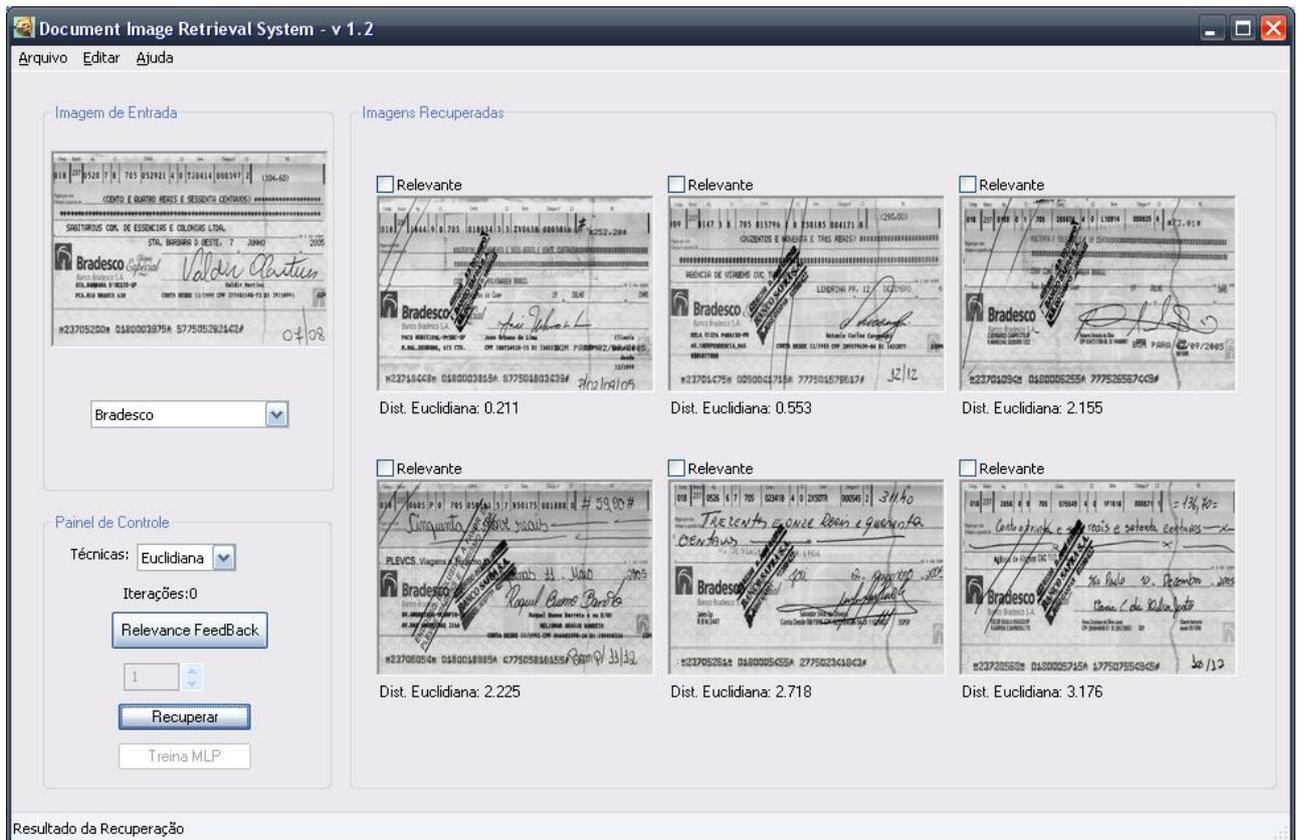


Figura 15. Interface gráfica do sistema de recuperação de documentos baseada em conteúdo implementada.

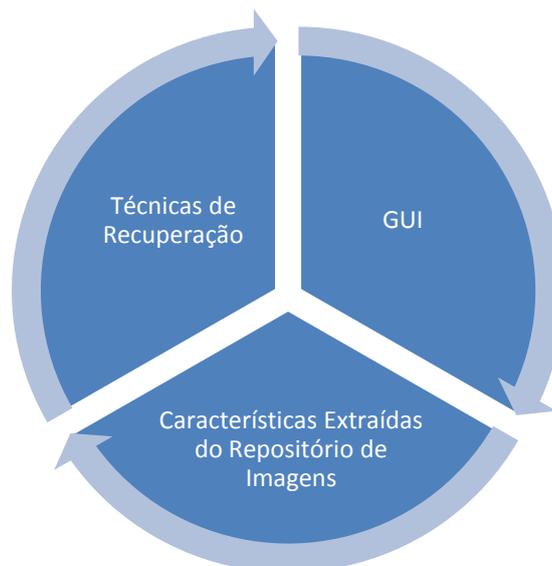


Figura 16. Diferentes módulos que compõem a ferramenta implementada.

A fim de extrair o elemento gráfico contido nos documentos foi implementado um algoritmo que utiliza a operação de abertura descrita pela morfologia matemática apresentada na sub-seção 3.3.1 com dois elementos estruturantes diferentes nas respectivas imagens binárias da base para tentar eliminar tudo que provavelmente não é figura como o texto escrito e outros ruídos diversos. Para as operações de morfologia matemática [26] todas as imagens foram negativadas, pois na erosão desejamos aumentar as áreas de preto e na dilatação aumentar as áreas brancas.

Foi utilizado um elemento estruturante para a erosão com dimensões de 20x30 pixels e para a dilatação foi usado um elemento de dimensões de 50 x 50 composto apenas por zeros. Ambos os elementos foram escolhidos empiricamente para eliminar texto e outros ruídos e realçar as figuras dos documentos. A Figura 17 mostra o elemento estruturante da erosão.

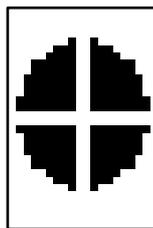


Figura 17. Elemento estruturante da erosão na operação de abertura.

O resultado dessas operações em cima das imagens que continham algum elemento gráfico originou na identificação dos logos, ainda que um pouco degradados, que estavam no conteúdo das imagens dos documentos. Para imagens com ruídos com grande quantidade de pontos pretos, o algoritmo também os considerou como uma possível região de interesse. Nesses casos, a região escolhida para extração de características foi aquela que tinha maior quantidade de pontos pretos.

Não foram computados erros para as imagens utilizadas nos experimentos, ou seja, para os ruídos presentes nas imagens sempre o logo teve maior concentração de pontos pretos que o ruído. Para os documentos que não continham figura no conteúdo foi retornada uma imagem toda em branco. A Figura 18 ilustra o resultado obtido pela operação de abertura realizada nas imagens de documentos manuscritos e datilografados sem figuras no conteúdo e na Figura 19 o resultado da operação de morfologia nas diferentes classes de cheques.

Após a identificação do elemento gráfico na imagem foi utilizada o Histograma da Projeção [21] para extrair as coordenadas da figura segmentada automaticamente. Após a

descoberta das coordenadas, um recorte é feito nas imagens originais para extrair características de texturas para montar um vetor de características que será utilizado por uma técnica de recuperação para buscar as imagens do sistema com base na imagem de entrada.

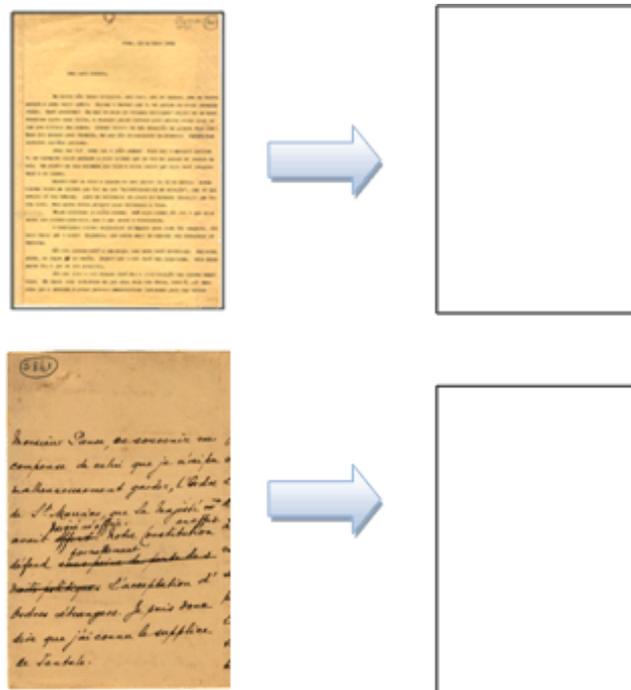


Figura 18. Resultado obtido pela operação de abertura nos tipos de documentos manuscritos e datilografados usados na base.



Figura 19. Resultado obtido pela operação morfológica de abertura nas principais classes de cheques usados na base.

Na construção do vetor de características foram usadas medidas estatísticas (mediana, desvio padrão e média do histograma da soma) assim como 5 atributos da imagem (contraste, correlação, energia, homogeneidade e entropia) que descrevem uma textura definidos por Haralick [19]. Para aplicações de recuperação de documento, as características de textura são informações importantes, pois provêm indicações da distribuição espacial da variação dos tons de cinza nos documentos.

Para descobrir o quão similar cada vetor de cada imagem é um do outro e, enfim, recuperar as imagens foram realizados experimentos com as diferentes técnicas descritas na seção 3.3.

No trabalho proposto, foram realizados testes a fim de comparar a técnica do vizinho mais próximo com redes neurais supervisionadas. Foi utilizada uma arquitetura MLP com algoritmo *backpropagation* [29] e função de ativação sigmoideal como expresso anteriormente na Equação (14) com uma camada de entrada, uma camada escondida e uma camada de saída.

Primeiro, os dados foram particionados em conjunto de treinamento e de teste. Ao todo foram usadas 200 imagens de cheques no formato BMP (*bitmap file format*) em tons de cinzas com largura de 1.350 e altura de 578 pixels, sendo que 180 imagens foram usadas para treinamento e 60 para teste. Foram definidas 4 classes de tal forma que cada uma contribui igualmente com 45 imagens de treinamento e 15 imagens de teste. A Tabela 1 associa cada classe a uma representação.

Tabela 1. Representação das classes dos documentos no sistema.

Classe	Representação
Banco Bradesco	0001
Caixa Econômica	0010
Banco Itaú	0100
Banco Real	1000

O fato de ter escolhido essa representação de classes está no fato de que para cada instituição bancária foi atribuída uma classe que foi representada de forma a expressar o resultado do *winner-takes-all* na classificação. Dessa forma o neurônio de ativação da camada de saída que obtiver o maior valor é atribuído para um e os outros recebem valor zero. Tal rotulação de classes não é usada em nenhum momento quando o k-NN é o classificador do sistema.

Após a extração das características foi construído um arquivo de leitura com os atributos e a respectiva classe de cada imagem de treinamento para obedecer ao treinamento supervisionado inerente ao classificador MLP. Já para os arquivos de teste, foi realizado o mesmo procedimento de geração do vetor de características exceto que no lugar da classe foi utilizado o nome do arquivo para que a interface gráfica pudesse recuperar as imagens.

Após o treinamento da rede MLP, é construída, então, uma lista com os nomes dos arquivos das imagens e suas respectivas saídas fornecida pela rede neural. O processo de

recuperação, então, se baseia apenas na retirada dos nomes dos arquivos que possuem mesma representação de classe da imagem que foi escolhida no painel de imagem de entrada do software.

Alguns testes foram realizados para validar a busca de imagens utilizando a rede neural MLP com algoritmo *backpropagation*. Foram testadas quatro arquiteturas diferentes variando a quantidade dos neurônios da camada escondida (8, 16, 24), ciclos de treinamento (5.000, 10.000) e a taxa de aprendizado (0,05 , 0,1). Foram assumidos 8 neurônios na camada de entrada (quantidade de atributos) e 4 neurônios na camada de saída (quantidade de classes envolvidas) e momentum fixo de 0,1. A melhor arquitetura encontrada foi a que teve 10.000 épocas de treinamento e taxa de aprendizado 0,1 com 16 neurônios na camada escondida. A Tabela 2 mostra os erros de treinamento encontrados ao se variar a quantidade de neurônios na camada escondida com 10.000 épocas de treinamento e taxa de aprendizado 0,1. A Tabela 3 demonstra a média de acertos para as classes com essa mesma arquitetura.

Tabela 2. Erro obtido no final do treinamento variando a quantidade de neurônios na camada escondida com taxa de aprendizado 0,1 e 10.000 ciclos de Treinamento.

Quantidade de neurônios na camada Escondida	8	16	24
Erro de Treinamento	10,565198	8,131427	8,272870

Tabela 3. Média de acertos de acordo com a quantidade de neurônios na camada escondida com taxa de aprendizado 0,1 e 10.000 ciclos de Treinamento.

Classe	Quantidade de neurônios na camada Escondida		
	8	16	24
Bradesco	5	7	4
Caixa	13	15	15
Itaú	14	15	15
Real	15	15	13
Média de Acertos			
(%)	83,33	86,66	81,66

A Tabela 4 mostra a matriz de confusão para a arquitetura com 10.000 épocas de treinamento, taxa de aprendizado 0,1 e com 16 neurônios na camada escondida. No Apêndice A podem ser encontrada as outras tabelas com os testes das outras arquiteturas.

Tabela 4. Matriz de confusão para o resultado da rede MLP com taxa de aprendizado 0,1 e 10.000 ciclos de treinamento e 16 neurônios na camada escondida.

Matriz de Confusão				
	Bradesco	Caixa	Itaú	Real
Bradesco	7	8	0	0
Caixa	0	15	0	0
Itaú	0	0	15	0
Real	0	0	0	15

Também foram realizados testes com o k-NN para validar a representatividade de cada imagem de cheque utilizada. Para abordagem do aplicativo DIR (*Document Image Retrieval*) implementado, considera-se acerto as imagens retornadas para os 15 primeiros vizinhos, ou seja, para as 15 imagens com respectivas distâncias Euclidianas menores. Essa quantidade foi escolhida pelo fato de ser o mesmo número de padrões de mesma classe usada para formar os vetores de características usados no treinamento do k-NN. Dessa forma, pode-se analisar também a distância Euclidiana dos consequentes erros.

Como método de ênfase da busca de imagens de documentos foi implementado um algoritmo baseado na teoria de *relevance feedback* descrita no Capítulo 3, após a primeira busca, os pesos da distância euclidiana ponderada, mostrada na Equação (12), são refinados de acordo com o percentual de semelhança entre as imagens corretamente recuperadas e a imagem teste. Se o valor de cada característica dos documentos relevantes chegar a 90% do valor da característica do padrão de entrada, então o peso correspondente a essa característica será incrementado de um e o processo de busca se repetirá iterativamente até que o resultado satisfaça a busca do usuário.

Na realização dos testes com o k-NN, constatou-se que a taxa de acerto sem RF foi em média de 76,66% como mostrado na Tabela 5 e, após o uso do RF, foi de 78,33% como mostrado na Tabela 6. Além disso, a técnica de *relevance feedback* ofereceu uma diminuição entre as diferenças das distâncias euclidianas das imagens que o algoritmo recupera corretamente. Isso denota que as imagens que pertence à mesma classe da imagem passam a ser

computacionalmente mais próxima da imagem teste. Outra nuance pertinente é que a diferença entre a última distância euclidiana de acerto e a primeira distância Euclidiana de erro aumentou com o uso de *relevance feedback*, mostrando que o sistema passa a discernir mais claramente entre as classes das imagens em questão. Dessa forma, podemos futuramente definir mais facilmente um limiar que retorne o máximo de imagens da mesma categoria.

Tabela 5. Resultados usando k-NN sem o uso de *Relevance Feedback*. *D é a diferença entre a última distância Euclidiana de acerto e a primeira distância Euclidiana de erro.

Classe	Acertos (%)	*D
Bradesco	80	0,572
Caixa	26,66	0,077
Itaú	100	43,026
Real	100	8,183
Média de Acertos		
(%)	76,66	

Tabela 6. Resultados usando k-NN com o uso de *Relevance Feedback*. *D é a diferença entre a última distância Euclidiana de acerto e a primeira distância Euclidiana de erro.

Classe	Acertos (%)	Iterações RF	*D
Bradesco	86,66	25	8,186
Caixa	26,66	25	6,438
Itaú	100	25	359,72
Real	100	25	92,589
Média de Acertos			
(%)	78,33		

Outro tipo de recuperação de documento foi realizada utilizando a classificação entre documentos manuscritos e datilografados. Para tanto, foram realizados experimentos com o espaço de Hough representado por cada imagem. Observamos que para imagens de documentos datilografados as linhas eram mapeadas para pontos (os acumuladores). A Figura 20 ilustra a forma de como as linhas são mapeadas para pontos no espaço de Hough. O mesmo acontece para

documentos manuscritos, só que a inclinação causada pela letra manuscrita faz com que a transformada de Hough interprete linhas no sentido da inclinação da escrita, gerando então outros valores altos nos acumuladores além dos já gerados pelas próprias linhas do documento como ilustrado na Figura 21.

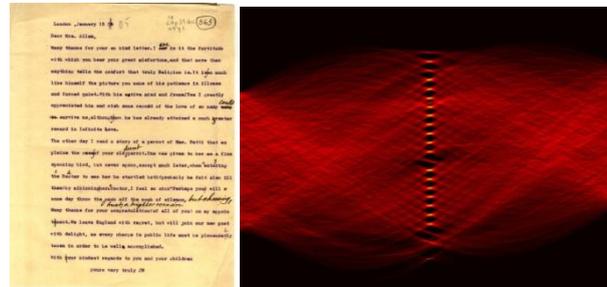


Figura 20. Documento datilografado e seu respectivo espaço de Hough.

Dessa forma, podemos discriminar se um documento é manuscrito ou datilografado. A Tabela 7 mostra a taxa de acerto para os documentos de diferentes naturezas. Os casos de erro foram de documentos manuscritos que possuíam poucas linhas ou de documentos datilografados com ruídos entre as linhas do documento.



Figura 21. Documento manuscrito e seu respectivo espaço de Hough.

Tabela 7. Taxa de acerto usando a transformada de Hough para identificar a natureza do documento.

Natureza do Documento	Acertos (%)
Documentos Manuscritos sem Figuras	71,42
Documentos Datilografados sem Figuras	83,33

Capítulo 5

Conclusões e Trabalhos Futuros

A recuperação de documentos baseada em conteúdo é uma tarefa desafiadora que movimenta a sociedade científica para a pesquisa de novas técnicas de recuperação e extração de características a fim de tornar os sistemas cada vez mais eficientes e com maior taxa de acerto possível. Muitas técnicas existentes são herdadas dos já consolidados sistemas de recuperação de imagens e texturas baseada em conteúdo (*CBIR e TBIR*), mas técnicas de extração de características que cubram totalmente o escopo de imagens de documentos ainda são insuficientes para difundir e concretizar o uso de sistemas capazes de buscar documentos com base em outros documentos.

5.1 Contribuições

Neste trabalho, propusemos e avaliamos um modelo de recuperação de documentos de diversas naturezas baseada na combinação de características de textura obtidas a partir de elementos gráficos presentes em cada imagem de documento e de sua estrutura. A estratégia de segmentar automaticamente as imagens contidas em um documento conseguiu representar bem todo o documento. Esse fato ficou claramente exposto nas taxas de acertos obtidos pelas diferentes técnicas utilizadas.

Sob o ponto de vista prático, a técnica desenvolvida mostra-se potencialmente valiosa para o desenvolvimento de sistemas especializados de busca de documentos servindo como ferramenta imprescindível para acervos e bibliotecas digitais. Ademais, este trabalho mostrou não apenas que imagens de documentos são importantes pelo conteúdo informacional a ser empreendido em sistemas de busca, mas também apresentou técnicas eficazes de explorá-los.

É importante frisar também que os resultados obtidos nesta monografia expressam, de forma computacional, a maneira de quão relevantes são as características extraídas, podendo, de acordo com a base de imagens, apresentar resultados diferentes. O trabalho mostrou a viabilidade da utilização de uma rede neural na tarefa de recuperação de imagens diversas.

A importância deste trabalho se deve, ainda, ao fato de ter apresentado um sistema de visão computacional que pode ser implementado e utilizado em aplicações reais como, por

exemplo, ajudar grandes empresas e instituições que necessitam de técnicas de indexação e busca de documentos digitalizados.

5.2 Trabalhos Futuros

Como trabalhos futuros, podemos citar a evolução da ferramenta desenvolvida para incluir novas funcionalidades como a visualização de todas as imagens recuperadas. A atual implementação mostra apenas as 6 primeiras e o restante é expresso em forma textual através do nome do arquivo. Além do desenvolvimento de uma sofisticação na parte gráfica da ferramenta, seria interessante tornar a extração de características integrada com todo mecanismo de recuperação. Assim, o sistema estaria apto a receber qualquer imagem sem ter que pré-processar separadamente as imagens que se deseja recuperar.

A fim de enriquecer e representar melhor o vetor de características pode-se utilizar também características globais da imagem em conjunto com as características dos logos presentes em cada documento. Assim, teremos uma distribuição mais precisa ainda que mais custosa computacionalmente.

No uso da técnica de recuperação baseada nos vizinhos mais próximos podemos tentar descobrir automaticamente um limiar que separe as imagens corretas das erradas. Pode-se tentar usar técnicas estatísticas como MSE (*Mean Square Error*) para tal finalidade.

Outros testes podem ser realizados com outras técnicas de treinamento supervisionado como o RBF-DDA [35]. Dessa forma, o resultado pode ser comparado para tentar chegar a um classificador que consuma menos tempo e recurso computacional e tenha o maior acerto possível.

Com relação à segmentação automática dos logos, a atual abordagem escolhe a região que possui mais pontos pretos. Isso pode levar a seleção de ruídos que se superem o logo na quantidade de *pixels* pretos. Pode-se analisar semanticamente a representatividade de cada região identificada após a aplicação das técnicas morfológicas de processamento de imagens digitais para tentar escolher aquela que se aproxima mais de um possível logo.

Bibliografia

- [1] Carleton University - Department of Systems and Computer Engineering. Disponível em: <<http://www.sce.carleton.ca/faculty/cuhadar/CBIR/index.htm>>. Acesso em 19 de outubro de 2008.
- [2] DARWISH, K. Retrieving Arabic Printed Document: a Survey, IBM Technology Development Center, Cairo, Egito.
- [3] TRAINA, A.; MARQUES, J.; TRAINA C. Fighting the Semantic Gap on CBIR Systems through New Relevance Feedback Techniques. In: 19th IEEE Intl. Symposium on Computer-Based Medical Systems - CBMS; 2006; Salt Lake City, USA p. 1-6.
- [4] GRIGOROVA, A. et al. Content-Based Image Retrieval by Feature Adaptation and Relevance Feedback, In: IEEE Transactions on Multimedia, vol. 9, N. 6, oct, 2007.
- [5] VELTKAMP, R.; TANASE, M. Content-Based Image Retrieval Systems: A Survey, In: Technical Report UU-CS-2000-34, oct 2000, Utrecht, Alemanha.
- [6] MARINAI, S.; MARINO, E.; CESARINI, F.; SODA, G. A General System for the Retrieval of Document Images from Digital Libraries, In: Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL'04), 2004, California, USA, p.150.
- [7] CASTELLI, V.; BERGMAN, L. D. Image Databases: Search and Retrieval of Digital Imagery. 1. Ed. New York: Wiley-Interscience, 2002. 560 p.
- [8] BACKES, R. A.; MARTINEZ, O. Segmentação de Texturas por Análise de Complexidade. In: FAPESP, Processo N° 04/04572-0, 2006.
- [9] GONZALES, R.C.; WOODS, R.E. Digital Image Processing. 2. Ed. Nova Jersey: Prentice Hall, 2002. 793 p.
- [10] SMITH, B. Fundamentals of Fourier Transform Infrared Spectroscopy. 1. ed. London: CRC Press, 1995. 202 p.
- [11] SCHWARTZ, W.R.; PEDRINI, H. Método para Classificação de Imagens Baseada em Matrizes de Co-ocorrências utilizando Características de Textura. In: Colóquio Brasileiro de Pesquisas Geodésicas, 2003.

- [12] MASCARO, A. Segmentação de Imagens de Mamografias Digitais. 2007. Dissertação de Graduação em Engenharia da Computação, Universidade de Pernambuco - UPE, Recife.
- [13] BALASUBRAMANIAN, A. Retrieval from Document Images Collections In: Lecture Notes in Computer Science 3872, p. 1–12, 2006.
- [14] CHEN N.; BLOSTEIN D. A survey of document image classification: problem statement, classifier architecture and performance evaluation. In: International Journal of Document Analysis and Recognition (IJ DAR'07), 2007, Trier, Alemanha, p. 180.
- [15] JIN X.; FRENCH J. Improving Image Retrieval Effectiveness via Multiple Queries, In: ACM International Workshop on Multimedia Databases, Nov, 2003, New Orleans, USA.
- [16] TRIER, D.; JAIN, A.K.; TAXT, T. Feature extraction methods for character recognition – A survey. Pattern Recognition. In: Technical Report for Michigan State University, jul 1995, Michigan, USA.
- [17] OKUN, O.; DOERMANN, D.; PIETIKÄINEN, M. Page segmentation and zone classification: the state of the art. Technical Report for University of Maryland, nov 1999, Maryland, USA.
- [18] WANG, Y.; PHILLIPS, I.T.; HARALICK, R. A study on the document zone content classification problem. In: Proceedings of the 5th International Workshop on Document Analysis Systems, 2002, Princeton, p. 212–223.
- [19] HARALICK, R.: Document image understanding: geometric and logical layout. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, 1994, Seattle, p. 385–390.
- [20] MAO, S.; ROSENFELD, A.; KANUNGO, T. Document structure analysis algorithms: a literature survey. In: Proceedings of Document Recognition and Retrieval X (IS&T/SPIE electronic imaging), 2003, Santa Clara, USA, p. 197–207.
- [21] GUO, P.; MIAO, Z. Projection Histogram Based Human Posture Recognition. In: 8th International Conference on Signal Processing, Vol. 2, Beijing, China, 2006
- [22] MENG, H.; ZHENG N.; SONG Y. e ZHANG Y. Document image retrieval based on Multiple features Combination, In: Proc. 9th ICDAR, 2007, Paraná, Brazil, p. 1040-1044.
- [23] ESPOSITO, F., MALERBA, D., LISI, F.A.: Machine learning for intelligent processing of printed documents. J. Intell. Inf. Syst. 14(2–3), 175–198, 2000.
- [24] TANG, Y.Y., CHERIET, M., LIU, J., SAID, J.N., SUEN, C.Y.: Document analysis and recognition by computers. In: Handbook of Pattern Recognition and Computer Vision, 2nd edn. World Scientific, Singapore, p. 579–612, 1998.

- [25] CAVNAR, W.; TRENKLE, J. N-gram-based text categorization. In: Proceedings of the 3rd Annual Symposium on Document Analysis and Information Retrieval, 1994, Las Vegas, USA, p. 161–175.
- [26] WU, J.; HSIEH, J.; CHEN, Y. Morphology-based text line extraction. In: Machine Vision and Applications, volume 19, 2007, Heidelberg, Germany, p. 195-207.
- [27] BUCKLEY, C.; SALTON, G. Optimization of Relevance Feedback Weights. In: Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 1995, Seattle, Washington. USA, p. 351 – 357.
- [28] LAM, S. An adaptive approach to document classification and understanding. In: Proceedings of International Association for Pattern Recognition Workshop on Document Analysis Systems, 1994, Kaiserslautern, Germany, p. 231–251.
- [29] RUI, Y. et al. Relevance feedback techniques in interactive content-based image retrieval. In: Storage and Retrieval for Image and Video Databases (SPIE), 1998, San Jose, USA, p. 25 – 36.
- [30] ALTUN, H; GÖKHAN, G. Enhancing Performance of MLP/RBF Neural Classifiers via an Multivariate Data Distribution Scheme. In: Nigde University, Electrical and Electronics Engineering Department. Nigde, Turquia, 2000.
- [31] RUMELHART, D. E. et al. Parallel Distributed Processing. Volume 1: Foundations. The MIT Press. 1986.
- [32] BRAGA, A.; CARVALHO, André; LUDERMIR, Teresa. Redes Neurais Artificiais – Teoria e Aplicações. 1. ed. Rio de Janeiro: LTC, 2000. 262 p.
- [33] Python em: <http://www.python.org/>. Acesso em 20 de dezembro de 2008.
- [34] Matlab em: <http://www.mathworks.com/>. Acesso em 20 de dezembro de 2008.
- [35] OLIVEIRA A. L. I., NETO F. B. L., e MEIRA S. R. L.. Improving RBF-DDA performance on optical character recognition through parameter selection. In: Proc. of the 17th International Conference on Pattern Recognition (ICPR'2004), volume 4, pages 625–628, Cambridge,UK, 2004. IEEE Computer Society Press.

Apêndice A

Tabelas de comparação das diferentes arquiteturas usadas no MLP

Tabela 8. Média de acertos com 5.000 épocas de treinamento e taxa de aprendizado 0,05.

Classe	Quantidade de neurônios na camada Escondida		
	8	16	24
Bradesco	5	4	3
Caixa	15	15	15
Itaú	15	15	15
Real	15	15	15
Média de Acertos			
(%)	83,33	81,66	80,00

Tabela 9. Evolução do Erro com 5.000 épocas de treinamento e taxa de aprendizado 0,05.

Quantidade de neurônios na camada Escondida	8	16	24
	Erro de Treinamento	11,567286	11,391279

Tabela 10. Média de acertos com 5.000 épocas de treinamento e taxa de aprendizado 0,1.

Quantidade de neurônios na camada Escondida			
Classe	8	16	24
Bradesco	5	1	1
Caixa	15	15	15
Itaú	15	15	15
Real	15	15	15
Média de Acertos			
(%)	83,33	76,66	76,66

Tabela 11. Evolução do Erro com 5.000 épocas de treinamento e taxa de aprendizado 0,1.

Quantidade de neurônios na camada Escondida	8	16	24
Erro de Treinamento	11,346390	10,251427	10,555296

Tabela 12. Média de acertos com 10.000 épocas de treinamento e taxa de aprendizado 0,1.

Quantidade de neurônios na camada Escondida			
Classe	8	16	24
Bradesco	5	7	4
Caixa	15	15	15
Itaú	15	15	15
Real	15	15	15
Média de Acertos			
(%)	83,33	86,66	81,66

Tabela 13. Evolução do Erro com 10.000 épocas de treinamento e taxa de aprendizado 0,1.

Quantidade de neurônios na camada Escondida	8	16	24
Erro de Treinamento	10,565198	8,131427	8.272870

Tabela 14. Média de acertos com 10.000 épocas de treinamento e taxa de aprendizado 0,05.

	Quantidade de neurônios na camada Escondida		
Classe	8	16	24
Bradesco	5	6	6
Caixa	15	15	15
Itaú	15	15	15
Real	15	15	15
Média de Acertos (%)	83,33	85,00	85,00

Tabela 15. Evolução do Erro com 10.000 épocas de treinamento e taxa de aprendizado 0,05.

Quantidade de neurônios na camada Escondida	8	16	24
Erro de Treinamento	11,113379	9,793982	9.074900