

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Desenvolvimento de métodos
para extração, comparação e análise
de características intrínsecas de imagens médicas,
visando à recuperação perceptual por conteúdo

Joaquim Cezar Felipe

São Carlos - SP

Desenvolvimento de métodos
para extração, comparação e análise
de características intrínsecas de imagens médicas,
visando à recuperação perceptual por conteúdo

Joaquim Cezar Felipe

Orientadora: *Profa. Dra. Agma Juci Machado Traina*

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional.

USP – São Carlos

Dezembro de 2005

RESUMO

A possibilidade de recuperar e comparar imagens usando as suas características visuais intrínsecas é um recurso valioso para responder a consultas por similaridade em imagens médicas. Desse modo, a agregação desses recursos aos Sistemas de Arquivamento e Comunicação de Imagens (*Picture Archiving and Communication Systems* - PACS) vêm potencializar a utilidade e importância destes no contexto de atividades tais como ensino e treinamento de novos radiologistas, estudos de casos e auxílio ao diagnóstico de forma geral, uma vez que as consultas por similaridade permitem que casos parecidos possam ser facilmente recuperados.

O trabalho apresentado nesta tese possui duas vertentes. Primeiro, ele apresenta novos métodos de extração e de características, com o objetivo de obter a essência das imagens, considerando um critério específico. Os atributos obtidos pelos algoritmos de extração são armazenados em vetores de características para posteriormente serem utilizados para indexar e recuperar as imagens baseando-se em seu conteúdo, para responder a consultas por similaridade. Há uma relação próxima entre os vetores de características e as funções de distância utilizadas para compará-los. Assim, a segunda parte deste trabalho trata da proposta, análise e comparação de novas famílias de funções de distância. As funções de distância propostas têm por objetivo tratar o problema do *gap* semântico, o qual representa o principal obstáculo das funções de distância tradicionais, derivadas da família L_p , quando processam consultas por similaridade.

As principais contribuições desta tese incluem o desenvolvimento de novos métodos de extração e comparação de características de imagens, que operam sobre os três principais descritores de baixo nível de imagens: distribuição de cor, textura e forma. Os experimentos realizados mostraram que os ganhos em precisão são maiores para os métodos propostos, quando comparados com algoritmos tradicionais. No que diz respeito às famílias de funções de distância propostas (WAID e SAID), pelos resultados iniciais obtidos, podemos afirmar que eles são bastante promissores no sentido de se aproximarem da expectativa do usuário, no momento de comparar imagens.

Os resultados obtidos com esse trabalho podem ser futuramente integrados aos PACS. Particularmente, pretendemos acrescentar novos algoritmos e métodos ao cbPACS, que consiste em um sistema PACS em construção, desenvolvido em uma colaboração entre o Grupo de Bases de Dados e Imagens (GBDI) do Instituto de Ciências Matemáticas e de Computação - USP e o Centro de Ciências da Imagens e Física Médica (CCIFM) da Faculdade de Medicina de Ribeirão Preto - USP.

Palavras-chave: recuperação de imagens baseada em conteúdo, consultas por similaridade, imagens médicas, diagnóstico auxiliado por computador.

ABSTRACT

The ability of retrieving and comparing images using their inherent pictorial information is a valuable asset to answer similarity queries over medical images. Thus, having such resources added in Picture Archiving and Communication Systems (PACS) increase their applicability and importance in the context of teaching and training new radiologists on diagnosing, since that similar cases can be easily retrieved. Similarity queries also play an important role on gathering close images, what allows to perform case studies, as well as to aid on diagnosing.

The work presented in this thesis is twofold. First, it presents new feature extraction techniques, which aim at obtaining the essence of the images regarding a given criteria. The features obtained by the algorithms are stored in feature vectors and employed to index and retrieve the images by content, in order to answer similarity queries. There is a close relationship among feature vectors and the distance function employed to compare them. Thus, the second, part of this work concerns the comparison, analysis and proposal of new families of distance functions to compare the features extracted from the images. The distance functions proposed intend to deal with the *semantic gap* problem, which is the main drawback of the traditional distance functions derived from the L_p metrics when processing similarity queries.

The main contributions of this thesis include the development of new image feature extractors that works on the three aspects of raw image data (color distribution, texture and shape). The experiments have shown that the gain in precision are higher for all the feature extractors proposed, when comparing with the state-of-the-art algorithms. Regarding the two families of distance functions WAID and SAID proposed, by the initial experiments performed we can claim that they are very promising on preserving the user expectation when comparing images.

The results provided by this work can be straightforwardly integrated to PACS. Particularly, we intend to add the new algorithms and methods to cbPACS, which is under joined development between the Image Data Base Group of Instituto de Ciências Matemáticas e de Computação of USP and Centro de Ciências de Imagens e Física Médica of Faculdade de Medicina de Ribeirão Preto of USP.

Keywords: content-based image retrieval, picture archiving and communication system, similarity query, medical images, computer-aided diagnosis.

SUMÁRIO

1. INTRODUÇÃO	13
1.1. Considerações Iniciais	2
1.2. Motivação	3
1.3. Objetivos e Contribuições	5
1.4. Organização da Tese	6
2. SISTEMAS PACS E RECUPERAÇÃO POR CONTEÚDO EM IMAGENS MÉDICAS	9
2.1. Introdução	9
2.2. Sistemas PACS	12
2.2.1. Aquisição de Imagens	13
2.2.2. Rede para Transmissão de Imagens e Dados	14
2.2.3. Armazenamento de Imagens	14
2.2.4. Exibição de Imagens	15
2.2.5. Considerações Adicionais	16
2.3. Recuperação de Imagens por Conteúdo	17
2.3.1. Características de Bancos de Dados de Imagens	20
2.3.2. Tipos de Consultas	23
2.4. Recuperação por Conteúdo de Imagens Médicas	24
2.4.1. Processamento e Análise de Imagens Médicas	24
2.4.2. Incorporação de CBIR aos Sistemas PACS	25
2.4.3. Exemplo de CBIR / PACS	25
2.5. Desafios	28
2.6. Considerações Finais	29
3. FUNÇÕES DE DISTÂNCIA E MÉTODOS COMPARATIVOS	31
3.1. Introdução	31
3.2. A Distância Minkowski (L_p)	32
3.2.1. Distância Minkowski Ponderada	35
3.3. Outras Distâncias de Uso Consolidado	36
3.3.1. Distância Mahalanobis	36
3.3.2. Divergência de Kullback-leibler e Divergência de Jeffrey	37
3.3.3. Medida χ^2	37
3.3.4. Interseção de Histogramas	37
3.4. Outras Propostas de Distância	38
3.4.1. Considerações sobre Tipos de Dados	39
3.4.2. Funções de Distância Métricas	39
3.5. Funções de Distância Típicas para Descritores Específicos	41
3.6. Natureza Estatística das Funções de Distância	42
3.7. Comparações Experimentais entre Funções de Distância	43
3.8. Métodos Comparativos	44
3.8.1. Pré-Processamento dos Dados	45
3.8.2. Normalização	45
3.8.3. Redução da Dimensionalidade	46
3.8.4. Exemplos de Métodos Comparativos	49
3.9. Considerações Finais	56
4. MÉTODO PARA COMPARAÇÃO OTIMIZADA ENTRE HISTOGRAMAS	58
4.1. Introdução	58
4.2. Definição do Problema	58
4.3. O Método Desenvolvido	61
4.3.1. Características Globais de Histogramas	61
4.3.2. <i>Global Warp Distance</i>	63

4.4. Resultados Experimentais	65
4.4.1. Experimento 1	65
4.4.2. Experimento 2	67
4.4.3. Experimento 3	70
4.5. Conclusões	71
5. RECUPERAÇÃO POR SIMILARIDADE E CLASSIFICAÇÃO DE IMAGENS MÉDICAS USANDO DESCRITORES DE FORMA E APLICANDO REDUÇÃO DE DIMENSIONALIDADE	74
5.1. Introdução	74
5.2. Definição do Problema	75
5.3. O Método Desenvolvido	76
5.3.1. Etapa 1: Pré-processamento e extração de momentos	76
5.3.2. Etapa 2: Determinação dos momentos relevantes.....	77
5.3.3. Etapa 3: Determinação dos momentos representativos.....	77
5.4. Abordagem Experimental	78
5.4.1. Testes com Segmentação	79
5.4.2. Pré-processamento das Images	80
5.4.3. Momentos Relevantes e Momentos Representativos.....	81
5.4.4. Resultados	83
5.5. Conclusões	84
6. DESCRITORES DE TEXTURA PARA CARACTERIZAÇÃO DE TECIDOS	87
6.1. Introdução	87
6.2. O Método Desenvolvido	89
6.3. O Descritor Proposto	90
6.4. Resultados Experimentais	91
6.4.1. Ferramenta para Extração de Características e Consulta aos Vizinhos Próximos	92
6.4.2. Experimento 1	93
6.4.3. Experimento 2	95
6.4.3. Experimento 3	956
6.5. Conclusões	97
7. UMA NOVA FAMÍLIA DE DISTÂNCIAS PERCEPTUAIS	99
7.1. Introdução	99
7.2. Definição do Problema	99
7.3. Distâncias Propostas: AID (<i>Attribute Interaction Distances</i>)	102
7.3.1. Grau de Interação	103
7.3.2. Expressões Gerais das Famílias AID	104
7.3.3. Estendendo as Expressões das Famílias AID para Dimensões Superiores.....	107
7.3.4. Composição das Famílias AID com a Família Lp	109
7.3.5. Procedimento para Determinar a Função de Distância mais Adequada	111
7.4. Resultados Experimentais	112
7.4.1. Similaridade Perceptual de Cores	112
7.4.2. Similaridade Perceptual de Textura para Imagens Médicas	115
7.4.3. Similaridade Perceptual de Imagens Médicas com Atributos Variados	118
7.5. Conclusões	119
8. CONCLUSÃO	121
8.1. Considerações Iniciais	121
8.2. Principais Contribuições e Resultados	121
8.3. Publicações	123
8.4. Propostas de Trabalhos Futuros	124
8.5. Considerações Finais	127

APÊNDICE A - PROCESSAMENTO DE IMAGENS.....	141
A.1. Introdução	141
A.2. Descritores	141
A.2.1. Cor.....	142
A.2.2. Textura	144
A.2.3. Forma	148
A.3. Operações em Imagens.....	150
A.3.1. Realce de Contraste (<i>Stretching</i>).....	150
A.3.2. Filtragem	150
A.3.3. Operações Algébricas	152
A.3.4. Segmentação	152
A.3.5. Morfologia Matemática.....	154
A.4. Transformadas.....	156
A.4.1. Transformada Discreta de Fourier	157
A.4.2. Transformada Discreta do Co-Seno	158
A.4.3. Wavelets.....	158
A.5. Momentos	160
A.5.1. Momentos Centrais	161
A.5.2. Momentos de Zernike	162
A.6. Considerações Finais	163

Lista de Figuras

Figura 2.1. Integração entre HIS, RIS e PACS	10
Figura 2.2. Componentes de um PACS	12
Figura 2.3. Processos e fluxos de dados envolvidos em um ambiente CBIR.....	19
Figura 2.4. Sistema de integração entre PACS e CBIR.....	26
Figura 2.5. Interface do SRIS-HC	27
Figura 2.6. Arquitetura geral do CIRCE	28
Figura 3.1. Lugar geométrico definido pela distância L_2	33
Figura 3.2. Lugar geométrico definido pela distância L_1	34
Figura 3.3. Lugar geométrico definido pela distância L_∞	34
Figura 3.4. Lugar geométrico definido pelos diferentes valores de p , no espaço 3D.....	35
Figura 3.5. Outros poliedros gerados pela extensão dos operadores de Minkowski	35
Figura 3.6. Efeito centauro: incongruência da desigualdade triangular.....	40
Figura 3.7. Exemplo de histograma métrico	50
Figura 3.8. Comparação entre séries: a) convencional; b) com <i>Dynamic Time Warping</i>	51
Figura 3.9. Imagem original e seus dois níveis de decomposição <i>Wavelet</i>	54
Figura 3.10. Imagem original (a), segmentada em 4 classes (b) e classes separadas (c - f) ...	56
Figura 4.1. (a) Imagem original e (b) com alterações de brilho e contraste	60
Figura 4.2. Histogramas típicos de regiões específicas do corpo humano	60
Figura 4.3. Características estatísticas e globais dos histogramas	62
Figura 4.4. Dois histogramas e suas respectivas matrizes de ajuste	64
Figura 4.5. Curvas de revocação versus precisão de consultas om HD, MD e GWD.....	67
Figura 4.6. Resultados de consulta aos 20 vizinhos mais próximos: (a) MD e (b) MWD	68
Figura 4.7. Resultados de consultas: (a) MD, (b) MWD e (c) GWD	69
Figura 4.8. Seqüências obtidas com o radiologista 1 (R1), a distância GWD e MD.....	70
Figura 5.1. Nódulos tumorais de mama típicos: benignos e malignos	75
Figura 5.2. Etapas para definição de atributos e CBIR.....	77
Figura 5.3. Zernike Extractor executando uma consulta aos k vizinhos mais próximos.....	79
Figura 5.4. Imagens originais; segmentadas pelo LDA; segmentadas pelo EM/MPM.....	80
Figura 5.5. Apresentação parcial da árvore de decisão gerada pelo algoritmo C4.5	82
Figura 5.6. Composição da dimensão fractal parcial	82
Figura 5.7. Precisão versus revocação: 256, 38, 18 e 16 momentos.....	83

Figure 5.8. Precisão versus revocação: 18, 17, 16 e 15 momentos.....	85
Figura 6.1. Passos do método de comparação entre imagens	89
Figura 6.2. Quatro possíveis distâncias, obtidas a partir do rodízio	90
Figura 6.3. Composição do Histograma de Contraste	91
Figura 6.4. Exemplos de ROIs utilizadas nos experimentos	92
Figura 6.5. Interface da ferramenta <i>Texture Extractor</i>	93
Figura 6.6. Revocação versus precisão da combinação de descritores.....	95
Figura 7.1. Exemplos de interações entre atributos	101
Figura 7.2. Exemplo de alta e baixa proporções de diferença	102
Figura 7.3. Curvas de WAID e SAID	103
Figura 7.4. Lugar geométrico dos pontos localizados à distância d , para valores de n	104
Figura 7.5. Condições para determinar a expressão geral da família SAID	105
Figura 7.6. Condições para determinar a expressão geral da família WAID.....	106
Figura 7.7. Comparações entre L1 e L2 (a) e entre L^∞ e L2 (b)	109
Figura 7.8. Composição da família L_p com as famílias AID	110
Figura 7.9. Variações da distância de C a Q, com as famílias L_p e AID	111
Figura 7.10. Lâminas de cores. A cor central (Q) é a referência	113
Figure 7.11. Imagens que apresentam o mesmo valor de L^∞ em relação à imagem Q.....	115
Figura 7.12. Exemplos de imagens usadas no conjunto de testes com atributos de textura..	115
Figura 7.13. Valores de precisão para avaliar o desempenho das funções de distância.....	116
Figura 7.14. Valores de precisão para determinar a melhor função de distância SAID.....	117
Figura 7.15. Valores de precisão utilizando 4 atributos de textura.....	118
Figura 7.16. Revocação versus precisão para as diferentes funções de distância.....	119
Figura A.1. Imagem de TC e seu histograma de intensidades.....	144
Figura A.2. Texturas da base de dados Brodatz	144
Figura A.3. Texturas de tecidos do corpo humano	145
Figura A.4. Imagem e matriz de co-ocorrência	151
Figura A.5. Imagem da Figura A.1 após processo de realce de contraste	151
Figura A.6. "Janelamento" para realce de contraste da imagem da Figura A.1.....	151
Figura A.7. Imagem da Figura A.1 acrescida de ruído e com filtro da mediana	152
Figura A.8. (a) Imagem original; (b) Imagem com bordas detectadas.	153
Figura A.9. (a) Imagem original; (b) Imagem segmentada em regiões	154
Figura A.10. Operações de dilatação e erosão	156
Figura A.11. Representação espacial das séries polinomiais de Zernike até 6ª ordem	163

Lista de Tabelas

Tabela 3.1. Matriz de classificação cruzada entre dois objetos	39
Tabela 4.1. Valores de precisão das sequências dos radiologistas em relação às sequências das funções de distância MD e GWD	71
Tabela 6.1. Resultados de revocação versus precisão (áreas) para cada descritor	94
Tabela 7.1. Distâncias das imagens C1 e C2 à imagem Q	114
Tabela A.1. Descritores de textura	147

Lista de Abreviaturas

ACR	<i>American College of Radiology</i>
AID	<i>Attribute Interaction Distance</i>
BI-RADS	<i>Breast Imaging Reporting and Data System</i>
BLOB	<i>Binary Large Object</i>
CAD	<i>Computer-aided Diagnosis</i> (Diagnóstico Auxiliado por Computador)
CBIR	<i>Content-based Image Retrieval</i> (Recuperação de Imagens Baseada em Conteúdo)
cbPACS	<i>Content-based PACS</i>
CCIFM	Centro de Ciências das Imagens e Física Médica
CIRCE	<i>Content-based Image Retrieval Core Engine</i>
DBLP	<i>Database Systems and Logic Programming</i>
DCT	<i>Discrete Cossine Transform</i> (Transformada Discreta do Co-Seno)
DDSM	<i>Digital Database for Screening Mammography</i>
DFT	<i>Discrete Fourier Transform</i> (Transformada Discreta de Fourier)
DICOM	<i>Digital Imaging and Communication in Medicine</i>
DKL	Divergência de Kullback-Leibler
DWT	<i>Discrete Wavelet Transform</i> (Transformada Discreta de Wavelets)
EMD	<i>Earth Mover's Distance</i>
FD-ASE	<i>Attribute Significance Estimator based on the Fractal Dimension</i>
FFT	<i>Fast Fourier Transform</i> (Transformada Rápida de Fourier)
FMRP	Faculdade de Medicina de Ribeirão Preto
GBDI	Grupo de Bases de Dados e Imagens
GWD	<i>Global Warp Distance</i>
HC	Hospital das Clínicas
HD	<i>Histogram Distance</i> (Distância entre Histogramas)
HIS	<i>Hospital Information System</i> (Sistema de Informações)

Hospitales)

HL7	<i>Health Level 7</i>
ICMC	Instituto de Ciências Matemáticas e de Computação
INCOR	Instituto do Coração
LDA	Linha Divisora de Águas (<i>watershed</i>)
MAM	<i>Metric Access Method</i> (Método de Acesso Métrico)
MD	<i>Metric Distance</i> (Distância Métrica)
MM	Morfologia Matemática
MPEG	<i>Moving Picture Experts Group</i>
MRF	<i>Markov Random Field</i>
MWD	<i>Metric Warp Distance</i>
PACS	<i>Picture Archiving and Communication System</i> (Sistema de Arquivamento e Comunicação de Imagens)
PEP	Prontuário Eletrônico do Paciente
RAID	<i>Redundant Array of Independent Disks</i>
RD	Radiografia Digital
RIS	<i>Radiology Information System</i> (Sistema de Informações Radiológicas)
RM	Ressonância Magnética
ROI	<i>Region of Interest</i> (região de interesse)
SAID	<i>Strong Attribute Interaction Distances</i>
SAM	<i>Spatial Access Method</i> (Método de Acesso Espacial)
SGBD	Sistema Gerenciador de Bases de Dados
SRIS-HC	Sistema de Recuperação de Imagens Similares – Hospital das Clínicas
StARMiner	<i>Statistical Association Rule Miner</i>
TC	Tomografia Computadorizada
TCP/IP	<i>Transmission Control Protocol/Internet Protocol</i>
USP	Universidade de São Paulo
WAID	<i>Weak Attribute Interaction Distance</i>
WD	<i>Warp Distance</i> (Distância Warp)

Capítulo 1

1. INTRODUÇÃO

Neste Capítulo apresentamos o contexto, a motivação e os desafios que deram origem ao desenvolvimento do projeto de doutorado que culmina com a construção e defesa desta tese. Os principais objetivos são discutidos e algumas das contribuições almeçadas são apresentadas, finalizando com a descrição da organização da tese.

1.1. Considerações Iniciais

A inserção das Ciências da Computação e suas vertentes junto às áreas de atuação das Ciências Médicas é um processo que vem se consolidando ao longo das duas últimas décadas. Inicialmente, métodos e técnicas computacionais eram adaptados para utilização em ambientes restritos (como o administrativo ou de pesquisa, por exemplo), com a finalidade única de reduzir o esforço humano nas atividades rotineiras de hospitais, centros médicos e unidades de ensino e pesquisa em saúde. Com o tempo, a participação da Informática foi ganhando vulto, no sentido de estabelecer um envolvimento efetivo com o controle e a integração de processos e lidando com especificidades inerentes aos ambientes médicos, de modo que, naturalmente, começaram a surgir áreas de pesquisa e desenvolvimento de processos computacionais especificamente voltados às áreas da saúde.

No escopo de pesquisa e formação temos atualmente, no Brasil e no exterior, diversos programas de pós-graduação e cursos de graduação direcionados à integração entre as Ciências da Saúde e as Ciências da Computação, em áreas tais como Informática Biomédica, Engenharia Biomédica, Informática em Saúde e Informática Médica, todas voltadas à formação de profissionais e ao desenvolvimento e à aplicação de métodos computacionais a ambientes de saúde.

Em termos de sistemas computacionais, estabeleceram-se projetos e propostas focando áreas distintas, os quais, por força do inter-relacionamento existente entre estas áreas, também apresentam diversos pontos de tangência e de interseção. Os Sistemas de Informações Hospitalares (*Hospital Information System - HIS*) abrangem todo o funcionamento de um hospital, embarcando ou relacionando-se diretamente com outros sistemas específicos, tais como o Prontuário Eletrônico do Paciente (PEP), os Sistemas de Administração Hospitalar, os Sistemas de Informações Radiológicas (*Radiology Information System - RIS*) e os Sistemas de

Arquivamento e Comunicação de Imagens (*Picture Archiving and Communication System - PACS*). Os sistemas PACS, por sua vez, têm por função básica gerenciar os processos de armazenamento e exibição das imagens médicas digitais, obtidas pelos equipamentos especializados de geração de imagens no ambiente de radiologia. Estes sistemas são utilizados diretamente pelos usuários radiologistas, sendo que, para poder dar suporte de modo eficiente às tarefas relacionadas ao uso das imagens médicas, eles devem prever um "interfaceamento" com outros sistemas, a fim de poderem propiciar um intercâmbio de informações (sobre pacientes e exames, por exemplo), que resulta na disponibilização dessas informações para os usuários de forma integrada ao acesso às imagens relacionadas.

Além disso, na prática, a usabilidade dos sistemas PACS depara-se com um obstáculo que merece atenção: o grande volume de imagens produzidas diariamente no ambiente hospitalar gera uma alta dificuldade para a manipulação racional das mesmas, no sentido de se executarem buscas às bases de imagens, baseando-se em critérios de similaridade, com o objetivo de realizar tarefas tais como o diagnóstico clínico e o estudo de casos. Para auxiliar na solução desse problema, as técnicas de Recuperação de Imagens Baseada em Conteúdo (*Content-based Image Retrieval - CBIR*) apresentam-se como um campo promissor, no sentido de que a sua integração aos sistemas PACS poderão propiciar aos mesmos um conjunto de recursos de alto poder de discriminação das imagens, disponibilizando aos usuários uma ferramenta poderosa e útil para auxiliar, com propriedade, nos procedimentos de análise e avaliação de imagens médicas.

1.2. Motivação

Uma pergunta que surge quando se fala em consultas por similaridade de imagens é se, na avaliação dos resultados, eles correspondem ao que era esperado pelo usuário. Isto é, como saber, por exemplo, ao solicitar as imagens mais semelhantes a uma dada imagem, se a proximidade ou similaridade retornada pelo sistema realmente mapeia o que é desejado pelo usuário? Essa diferença entre um resultado obtido e o resultado que o usuário esperava é o que se convencionou chamar de "gap semântico". Nos sistemas que implementam técnicas de CBIR, os mecanismos de tratamento de consultas por similaridade devem aproximar-se, em um nível de precisão aceitável, dos critérios envolvidos na percepção humana (do especialista do domínio).

Muitas vezes os próprios médicos divergem entre si quanto ao ponto de vista semântico referente à análise de imagens médicas. Esse é um dos problemas mais conhecidos e discutidos sobre o diagnóstico por imagens. Imagens patologicamente semelhantes podem ter sido classificadas divergentemente. Um sistema de recuperação de imagens convenientemente calibrado poderá auxiliar no diagnóstico, fornecendo ao médico um conjunto de casos semelhantes ao do paciente em questão, juntamente com os diagnósticos efetuados, já que estes também estão associados à imagem no banco de dados. Desta forma, o médico poderá analisar comparativamente o tipo de tratamento dado ao paciente e sua resposta ao mesmo. Esse tipo de abordagem não poderá substituir o médico na análise da imagem, mas permitirá a consulta ao banco de imagens de modo eficiente e otimizado. Atualmente é impossível realizar-se esse tipo de pesquisa sem um conhecimento prévio sobre casos clínicos semelhantes (ou seja, sem que o médico saiba muito bem onde procurar referências) e, considerando o grande volume de exames realizados em hospitais, essa é uma tarefa bastante árdua.

Em CBIR as imagens são descritas por atributos (vetores de características) que as representam sob um determinado aspecto, sendo que a similaridade entre duas imagens é avaliada por meio da aplicação de uma função de distância aos respectivos vetores de características. Nessa estrutura, dois novos desafios se apresentam: a escolha de um conjunto de atributos que capturem a essência das imagens por meio de uma quantidade sucinta de valores e a escolha de uma função de distância capaz de avaliar a similaridade da forma mais próxima possível da percepção humana.

Na grande maioria dos sistemas CBIR, utilizam-se como funções de distância, de maneira generalista, a família de distâncias baseada nas normas L_p , mais especificamente as distâncias L_1 , L_2 e L_∞ . Porém, tais funções nem sempre se mostram eficazes para representar diretamente o sentimento que o analista humano tem ao comparar duas imagens. Essas funções são gerais e não captam aspectos complexos de percepção, que envolvem a sobreposição de diferentes características, com diferentes relevâncias, para cada situação particular das imagens em questão. Assim, os resultados de comparações entre imagens deverão ser mais apropriados e efetivos, se for feito uso de métodos ajustados às características mais específicas do domínio de aplicação - no caso, as imagens médicas - e também às diferentes tarefas que são comumente executadas num ambiente PACS.

Muito se tem buscado e muito há ainda por fazer, no sentido de tornar os métodos de consulta por similaridade de imagens médicas suficientemente confiáveis e efetivos, para que seu uso seja aceito e difundido no dia-a-dia do ambiente de radiologia digital.

1.3. Objetivos e Contribuições

O objetivo geral deste trabalho é contribuir com os sistemas de busca por similaridade entre imagens, mais especificamente os sistemas PACS, que manipulam imagens médicas, no sentido de desenvolver e avaliar métodos comparativos e funções de distância que reduzam o *gap* semântico, buscando compatibilizar seus resultados com a expectativa do especialista no seu processo de avaliação de imagens.

A base de representação de imagens por meio de suas características intrínsecas concentra-se em três grandes categorias de descritores que podem ser obtidos automaticamente: distribuição de cores, textura e forma. Sendo assim, pretendemos explorar essas três áreas, da seguinte forma, no contexto das imagens médicas:

- Propor, implementar e avaliar um método comparativo que aperfeiçoe a comparação entre histogramas de intensidade, objetivando reduzir ocorrências de resultados falsos positivos e falsos negativos;
- Propor, implementar e avaliar um método comparativo que utilize um descritor de forma, executando redução da dimensionalidade do conjunto de dados e permitindo que imagens sejam classificadas a partir da aplicação deste método comparativo;
- Propor, implementar e avaliar um método comparativo, juntamente com um descritor associado a características de textura da imagem. O método deverá ser invariante a operações de rotação da imagem.
- Implementar ferramentas computacionais apropriadas para extração das características relacionadas aos três descritores de baixo nível de imagens e também para permitir aos usuários realizar consultas aos vizinhos mais próximos, para cada um dos métodos propostos;

No contexto das funções de distância, uma nova família de funções será discutida, formalizada matematicamente e avaliada. Estas funções são sensíveis à proporção de interação entre os atributos que são utilizados para representar as imagens comparadas, de modo que a sua escolha criteriosa poderá melhorar a eficiência na geração de resultados próximos aos dos radiologistas.

Com essas contribuições, pretendemos estar auxiliando na evolução dos sistemas PACS, no sentido de agregar aos mesmos novas funcionalidades e novas opções de trabalho. Desse modo, o sistema computacional de auxílio ao diagnóstico e estudo de casos deve alcançar um índice de aceitação e usabilidade mais efetiva, contribuindo realmente para a melhoria de tais procedimentos.

Com a finalidade de divulgar estas contribuições e, ao mesmo tempo, receber opiniões e sugestões da comunidade científica, apresentamos trabalhos em 12 conferências, distribuídos da seguinte forma: em eventos internacionais, 5 artigos completos e 2 resumos; em eventos nacionais, 3 artigos completos e 2 resumos. No Capítulo 8, essas publicações são apresentadas.

1.4. Organização da Tese

O conteúdo desta tese está estruturado da seguinte forma:

O Capítulo 1 apresentou os desafios inerentes ao contexto deste trabalho, bem como a motivação para tentar resolvê-los e as contribuições deste trabalho com o desenvolvimento de métodos comparativos e de funções de distância apropriados.

O Capítulo 2 apresenta o contexto de aplicação deste trabalho, descrevendo os sistemas PACS e CBIR, destacando a interação entre eles.

O Capítulo 3 apresenta uma revisão sobre funções de distância e métodos comparativos, servindo como referência para a utilização das mesmas ao longo do trabalho.

O Capítulo 4 apresenta um novo método comparativo, a *Global Warp Distance*, usado para otimizar a comparação entre histogramas de intensidade. São mostrados também os resultados de experimentos realizados.

O Capítulo 5 traz uma descrição de um novo método comparativo que utiliza momentos de Zernike para reter informações sobre formato das imagens, além de técnicas para promover a redução da dimensionalidade. Experimentos realizados permitem avaliar o método.

No Capítulo 6, é descrito um novo método comparativo entre vetores de características baseadas em co-ocorrência de valores de cinza. Além disso, um novo descritor é proposto, sendo que o seu desempenho é testado e avaliado.

O Capítulo 7 apresenta uma nova família de distâncias cujo valor semântico se comprova com experimentos realizados com a ajuda de avaliações de radiologistas.

No Capítulo 8 são discutidos os resultados do trabalho desenvolvido, bem como são apresentadas as principais conclusões, assim como a indicação de linhas de futuras pesquisas.

Finalmente, o Apêndice A contém um resumo dos fundamentos sobre Processamento de Imagens, com ênfase nas técnicas que foram usadas pelos métodos propostos.

Capítulo 2

2. SISTEMAS PACS E RECUPERAÇÃO POR CONTEÚDO EM IMAGENS MÉDICAS

Neste Capítulo apresentaremos os conceitos e as características básicas inerentes ao contexto no qual concentram-se as contribuições do presente projeto de doutorado. Esse contexto compreende o ambiente de manipulação e análise de imagens médicas, normalmente contemplado pelos sistemas PACS, assim como as técnicas que compõem o processo de Recuperação de Imagens por Conteúdo (CBIR), almejando a integração de tais técnicas a esse ambiente.

2.1. Introdução

Os Sistemas de Informações Hospitalares (*Hospital Information System – HIS*) dão suporte computacional à maioria dos procedimentos operacionais de um hospital, processando, além de informações administrativas, informações relativas a pacientes, a exames realizados e a procedimentos médicos adotados. No ambiente de radiologia, os Sistemas de Informações Radiológicas (*Radiology Information System – RIS*) controlam os processos básicos, tais como geração e agendamento de procedimentos, manutenção da base de dados dos pacientes e geração de laudos e relatórios [Bushberg, J. T. '02] [Marques, P. M. d. A. '00]. Para o armazenamento, gerenciamento e disponibilização de imagens médicas, são desenvolvidos os Sistemas de Arquivamento e Transmissão de Imagens (*Picture Archiving and Communication System PACS*).

Esses três sistemas (HIS, RIS e PACS) apresentam interações de recursos e de funcionalidades. A sua correta integração no ambiente computacional de um hospital é de suma importância para o bom andamento do processo informatizado das atividades hospitalares [Morioka, C. A. '05]. A Figura 2.1 apresenta o inter-relacionamento básico entre eles.

Desde os primórdios da informática médica, uma das grandes preocupações em relação ao HIS concentra-se na logística da informação [Lehmann, T. M. '03b], no sentido de agregar recursos, do ponto de vista funcional, permitindo-lhe:

- a) apresentar informações de forma adequada;

- b) gravar e distribuir informações de forma que estas estejam disponíveis no lugar certo e no momento certo;
- c) se necessário, prover suporte à decisão.

Considerando as atividades de diagnóstico por imagens, realizadas pelos radiologistas, espera-se dos PACS que estes forneçam imagens de forma rápida, eficiente e com uma qualidade adequada para o diagnóstico. Apesar dos avanços recentes em tecnologias de *hardware*, tanto para os dispositivos de aquisição quanto para os dispositivos de exibição das imagens, os quais contribuem para aumentar a confiabilidade dos diagnósticos realizados em ambiente sem filme (*filmless*), os recursos para a recuperação das imagens armazenadas em PACS básicos restringem-se às informações associadas aos padrões relativos aos protocolos de comunicação de imagens adotados (DICOM, por exemplo), que incluem descrições textuais padronizadas sobre estudo, paciente, região do corpo e parâmetros técnicos relacionados à modalidade da imagem.

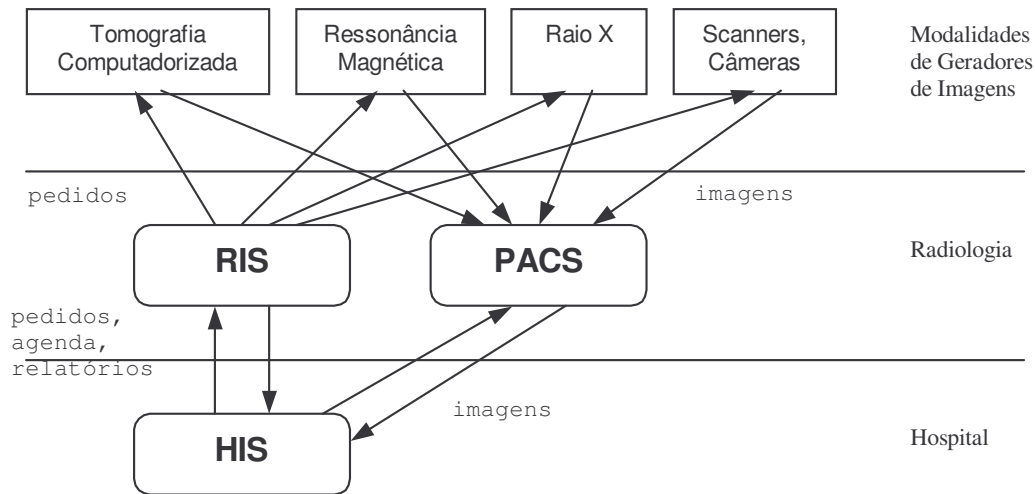


Figura 2.1. Integração entre HIS, RIS e PACS [Müller, H. '04]

A associação de anotações textuais às imagens pode ajudar a melhorar o processo de recuperação das mesmas, porém, o uso dessas anotações normalmente implica em algumas limitações importantes: a) a sua geração é trabalhosa e demorada; b) as anotações podem apresentar grandes variações de um usuário para outro; c) o seu conteúdo normalmente depende de um emprego ou contexto específico; d) geralmente descrevem de forma insuficiente a grande variedade de detalhes locais apresentados pelas imagens médicas.

Como os sistemas PACS armazenam e organizam grandes volumes de imagens provenientes de exames efetuados sobre os pacientes, um recurso de grande importância, a ser incorporado aos mesmos, seria a possibilidade de recuperar tais imagens através de **consultas**

por similaridade, baseadas no conteúdo intrínseco das mesmas. Por exemplo, pode ser necessário ou muito útil "*recuperar as 10 imagens obtidas em exames de ressonância magnética de coração que sejam similares a uma imagem dada*". Neste caso, o cruzamento das informações é independente das chaves de busca usuais, baseadas em dados numéricos e textuais, as quais associam exames e pacientes em um HIS tradicional. O termo "conteúdo intrínseco", nesse caso, diz respeito às características visuais próprias das imagens, tais como cor, textura e forma. A possibilidade de recuperação de dados baseada em seu conteúdo é um dos recursos atualmente mais almejados para incorporação aos sistemas PACS. Através desse recurso, podem ser encontrados registros de pacientes cujas imagens associadas apresentam aspectos de similaridade, além de se obterem correlações utilizando sintomas previamente registrados, tratamentos efetuados, etc [Marsh, A. '97]. Assim, um sistema PACS torna-se potencialmente uma ferramenta útil no auxílio ao diagnóstico médico [Shiraishi, J. '03].

A similaridade entre imagens pode ser medida de várias maneiras. Atributos relativos a descritores tais como forma, cor, textura bem como a posição espacial de elementos e regiões, podem ser extraídos de imagens armazenadas e são informações que podem servir como base para cálculos envolvendo similaridade [Aslandogan, Y. A. '99]. Pode-se então pesquisar e recuperar imagens utilizando comparações sobre essas informações e propriedades das imagens. Esse processo é chamado de Recuperação de Imagens por Conteúdo (CBIR). Técnicas de CBIR lidam com o conteúdo visual intrínseco das imagens para indexá-las e recuperá-las de forma automática [Bueno, J. M. '02c] [Lehmann, T. M. '03b] [Zhang, H. '02], na maioria dos casos, sem exigir a intervenção do usuário.

Portanto, o alto volume de imagens digitais produzidas e armazenadas diariamente nos hospitais, associado à alta complexidade apresentada pelas imagens médicas, credencia os PACS para agregarem recursos de CBIR, com o objetivo de prover mecanismos automáticos para a sua recuperação de forma que os médicos tenham um suporte à consulta à base de imagens, que lhes permita análises e avaliações eficientes para a definição do diagnóstico. Diversos trabalhos têm sido desenvolvidos no sentido de agregar funcionalidades de CBIR aos PACS [Bueno, J. M. '02a] [Antani, S. K. '04] [Rosa, N. A. '02] [Müller, H. '04] [Marques, P. M. A. '02] [Lehmann, T. M. '03b] [Lee, C.-S. '99] [Kak, A. '02] [Lehmann, T. M. '03a].

2.2. Sistemas PACS

Segundo a definição da associação americana NEMA (*National Electrical Manufactures' Association*), um PACS é um sistema que consiste de quatro componentes principais:

- interfaces para dispositivos de geração de imagens digitais (e, possivelmente, dispositivos para digitalização de imagens em filme);
- módulos para arquivamento e gerenciamento dos arquivos de imagens;
- estações de trabalho para acesso visual às imagens armazenadas;
- uma rede de comunicação para integração de todos os componentes.

A Figura 2.2 apresenta uma arquitetura que agrega esses componentes.

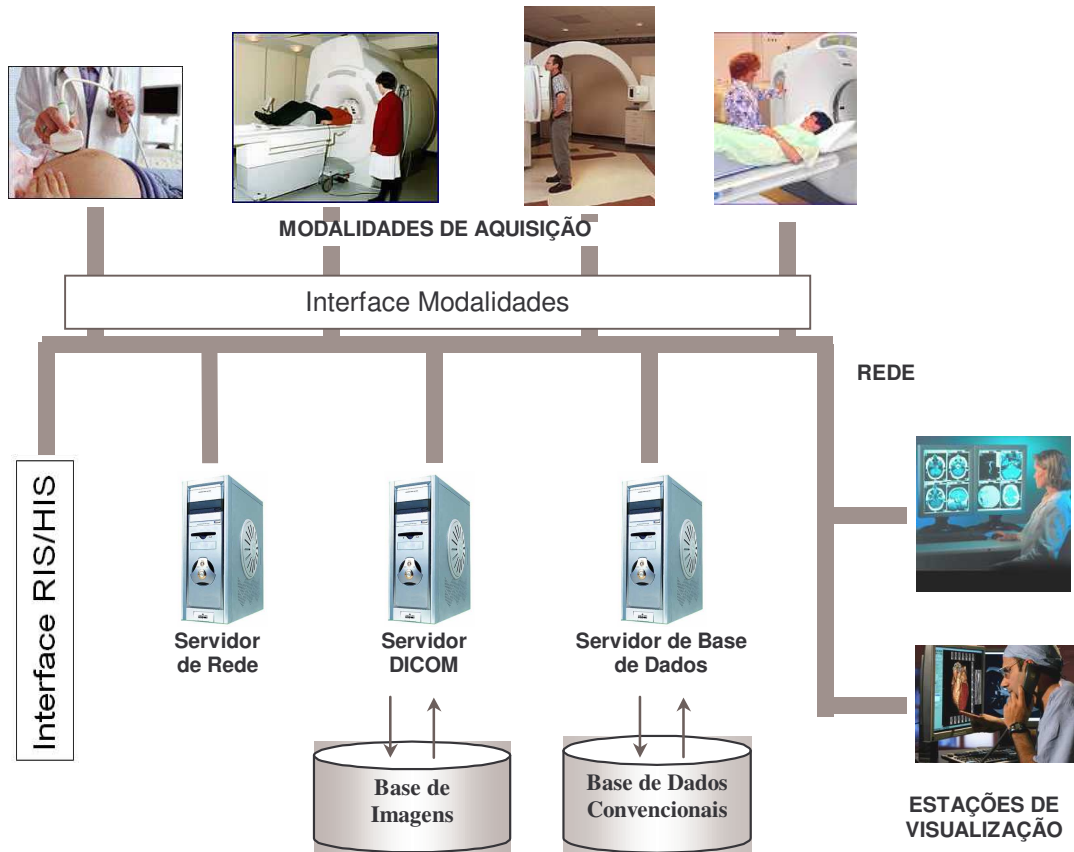


Figura 2.2. Componentes de um PACS

Os sistemas PACS variam em tamanho e abrangência. Um PACS pode ser dedicado a apenas uma modalidade em um setor médico específico, tal como o departamento de medicina nuclear ou o setor de ultra-som. Nesse caso, todo o sistema poderá ser conectado por uma única rede. Essa modalidade de PACS é geralmente chamada de **mini-PACS** [Bushberg,

J. T. '02]. Por outro lado, um PACS pode incorporar todas as modalidades em um sistema composto por diversos centros médicos. Nesse caso, o gerenciamento do sistema torna-se bastante complexo, demandando o uso de tecnologias avançadas de comunicação de dados.

2.2.1. Aquisição de Imagens

Os dispositivos de aquisição produzem imagens em diversas modalidades: Tomografia Computadorizada (TC), Ressonância Magnética (RM), Radiografia Digital (RD), Fluoroscopia Digital, Ultra-som e Angiografia Digital. Embora todos esses dispositivos sejam capazes de gerar imagens em formato digital, a grande maioria das imagens de radiografia continuam sendo adquiridas no formato convencional, ou seja, por meio de filmes [Siegel, E. L. '99].

Atualmente, a grande maioria dos equipamentos de modalidades digitais apresentam uma interface que utiliza o padrão de comunicação DICOM (*Digital Imaging and Communication in Medicine*) [NEMA '03]. O padrão DICOM especifica formatos para: (a) um conjunto de protocolos de comunicação entre equipamentos em uma rede; (b) a sintaxe e a semântica de comandos e informações associadas que podem ser compartilhados com esses protocolos; (c) uma série de serviços para armazenamento de mídia, assim como um formato e uma estrutura de diretórios para facilitar o acesso às imagens e as informações pertinentes às mesmas. DICOM especifica um formato de arquivo que permite a inclusão de informações relacionadas no cabeçalho do arquivo binário da imagem, para conjuntos de dados, tais como "pacientes", "estudos" e "imagens" [Hludov, S. '99]. Usualmente, DICOM é empregado como um protocolo de camada superior, acima do protocolo TCP/IP¹. Uma propriedade importante do protocolo DICOM é a capacidade de preservar o nível de resolução espacial da imagem, utilizado pelo equipamento de aquisição.

A implementação de um PACS deve prever um conjunto de interfaces que funcionem com eficiência. Essas interfaces são o meio pelo qual o PACS "conversa" com os outros componentes do cenário (Figura 2.2): a interface com as modalidades, para a troca de imagens e respectivas informações e a interface com os sistemas HIS/RIS, para transferir informações textuais sobre pacientes, exames, etc.

¹ TCP/IP (*Transmission Control Protocol/Internet Protocol*) é um conjunto de protocolos de comunicação usados para conectar servidores na Internet.

2.2.2. Rede para Transmissão de Imagens e Dados

As redes de comunicação têm a função de ligar os componentes do PACS entre si e também com outros sistemas. A rede deverá prover imagens para as estações de trabalho do próprio PACS, ou então para estações remotas de visualização, localizadas em unidades de emergência ou centros de terapia intensiva.

A arquitetura a partir da qual a rede será projetada pode ser central ou distribuída [Siegel, E. L. '99]. Numa arquitetura central, as imagens são disponibilizadas a partir de dispositivos de armazenamento compartilhados de "curto período", ou seja, durante um período específico de tempo, tipicamente em uma estrutura de tecnologia RAID², disponibilizando as imagens em todas as estações de trabalho, durante esse período. Numa arquitetura distribuída, as imagens ficam armazenadas em múltiplos servidores, distribuídos ao longo do ambiente hospitalar.

As interfaces com os sistemas RIS e HIS são também muito importantes, pois é através delas que informações típicas de cada um dos sistemas são integradas e disponibilizadas para todos [Bushberg, J. T. '02]. Por exemplo, informações sobre os pacientes ficam disponíveis para as estações de trabalho do PACS e imagens ficam disponíveis no Prontuário Eletrônico do Paciente. Outra vantagem importante dessa integração é a garantia da integridade de identidade do paciente, que se obtém com a manutenção de identificadores únicos para pacientes, exames e imagens.

A comunicação entre HIS, RIS e PACS é freqüentemente implementada através do uso de um padrão chamado *Health Level 7* (HL7), criado para a troca eletrônica de informações médicas, tais como informações administrativas e dados de exames laboratoriais. O acoplamento entre HL7 e DICOM permite, por exemplo, que a lista de informações sobre o paciente – oriunda do HIS ou do RIS – chegue até os dispositivos de imagens.

2.2.3. Armazenamento de Imagens

Uma estrutura de armazenamento comum nos PACS consiste de um esquema hierárquico, no qual imagens adquiridas recentemente são disponibilizadas em conjuntos de discos rígidos e, depois de um tempo específico, são transferidas para dispositivos de armazenamento mais lentos, porém de maior capacidade, tais como discos óticos e fitas magnéticas.

² RAID (*Redundant Array of Independent Disks*) é uma categoria de composição de discos rígidos que explora combinações de dois ou mais discos, objetivando desempenho e/ou tolerância a falhas.

A quantidade de espaço necessário para o armazenamento depende das modalidades disponibilizadas pelo PACS e pelo volume de trabalho do ambiente. É possível, superficialmente, estimar esse espaço. Um departamento de medicina nuclear de médio porte pode gerar alguns GBytes de imagens não comprimidas em um ano, de forma que uma estrutura simples de discos magnéticos seria suficiente para um mini-PACS servindo esse departamento. Por outro lado, imagens não comprimidas de TC, RM e RD, em um departamento de radiologia de tamanho médio, podem exigir vários GBytes de espaço por dia e vários TBytes por ano, necessitando de sistemas de arquivos bem mais extensos e complexos [Bushberg, J. T. '02].

Segundo [Furuie, S. S. '99], o armazenamento pode ser classificado como:

- *On-line*: provê acesso quase imediato aos dados. Tipicamente formado por conjuntos de discos magnéticos com capacidade de até centenas de GBytes, suportando armazenamento de 2 meses de exames para imagens com compressão sem perda;
- *Near-line*: provê tempo de acesso em torno de 1 minuto, sem intervenção humana. Formado por *jukeboxes* de discos óticos ou fitas magnéticas. Discos óticos permitem acesso mais rápido, porém fitas magnéticas são mais baratas. Os arquivos são descarregados para os dispositivos *on-line*. Armazenam, aproximadamente, imagens geradas nos últimos 3 anos;
- *Off-line*: acesso manual, com intervenção de um operador, para carregamento pré-agendado para os dispositivos *on-line*. Geralmente são usados discos óticos ou fitas magnéticas armazenados em estantes.

O controle do armazenamento pode ser realizado por meio de Sistemas Gerenciadores de Bases de Dados (SGBDs), que mantêm o gerenciamento de localização e movimentação das imagens e dos respectivos estados (*on-line*, *near-line* ou *off-line*).

2.2.4. Exibição de Imagens

Estações de trabalho computacionais equipadas com monitores de vídeo têm sido cada vez mais freqüentemente utilizadas para a visualização de imagens médicas, em ambientes de geração ou revisão de laudos. As diferentes modalidades radiológicas impõem características específicas para as estações de trabalho. Imagens de RD (Radiografia Digital), por exemplo, exigem alta resolução e maiores dimensões dos monitores em preto e branco (níveis de cinza),

enquanto exames de ultra-som, medicina nuclear e angiografia exigem monitores coloridos com capacidade de exibir seqüências filmadas.

Monitores coloridos comumente possuem o formato de 1280 x 1024 ou 1600 x 1200 pixels. Um formato geralmente encontrado em monitores de níveis de cinza usados para exibir radiografias digitais é 2048 x 2560 pixels, também conhecido como monitor de 5 Mpixels. Uma vez que imagens de RD grandes são tipicamente armazenadas em formatos de aproximadamente 2 x 2 Kpixels ou 2 x 2.5 Kpixels, resultando em aproximadamente 5 Mpixels, este formato permite a exibição de toda a imagem com resolução próxima da máxima. Estações de trabalho para avaliação de imagens são normalmente equipadas com pelo menos dois monitores para permitir a comparação simultânea de duas imagens com fidelidade total. Para estações de trabalho de medicina nuclear ou ultra-som, um único monitor colorido é suficiente.

A tendência que se observa é de aumento contínuo na velocidade de placas gráficas e de flexibilidade de operações das estações utilizadas para exibição de imagens, permitindo, assim, manipulação mais rápida das mesmas. Além disso, os computadores pessoais e até mesmo dispositivos móveis começam a ser usados como dispositivos de visualização em ambientes distribuídos e móveis, executando os aplicativos sob diferentes sistemas operacionais [Morioka, C. A. '05].

2.2.5. Considerações Adicionais

Infelizmente, a maioria dos sistemas PACS comerciais disponíveis são extremamente caros, além de não contemplarem todas as necessidades dos centros médicos [Sung, M. Y. '00]. Dessa forma, o que tem sido feito em termos de implementação prática resume-se no desenvolvimento de soluções locais, muitas vezes restritas. Alguns centros médicos integrados a unidades de pesquisa optam, via de regra, por adquirir apenas alguns módulos do sistema, desenvolvendo e adaptando outros módulos e criando soluções próprias conforme sua necessidade. Assim, o custo é diminuído e as características podem ser adaptadas às necessidades específicas de cada ambiente.

Os sistemas PACS apresentam diversas dificuldades no que diz respeito a implementação, funcionamento e manutenção. Dentre elas, podemos citar: alto custo; alta complexidade; perda de precisão quando imagens convencionais são digitalizadas; limitações de tamanho e resolução dos monitores de exibição; necessidade de enorme capacidade de armazenamento de dados; necessidade de pessoal técnico especializado e relutância dos

médicos em aceitar o ambiente computacional para trabalhar com imagens. Entretanto, as vantagens são muitas e são relevantes [Bushberg, J. T. '02]:

- Acesso (quase) imediato a imagens, dentro do centro médico e também em locais remotos;
- Possibilidade de mais de um usuário visualizar as mesmas imagens simultaneamente;
- Possibilidade de realce das imagens, por meio do uso de técnicas de processamento de imagens;
- Redução dos custos com material radiológico;
- Redução do espaço físico de armazenamento das imagens;
- Redução do pessoal de suporte físico;
- Possibilidade do uso de programas computacionais para otimizar o armazenamento, a recuperação e a análise das imagens, em tarefas tão importantes quanto a detecção de anomalias e o diagnóstico médico.

Logicamente, a estratégia a ser adotada durante a implementação de um PACS deverá buscar a minimização das dificuldades e a viabilização das vantagens. Os benefícios aumentam à medida que aumenta a experiência com o desenvolvimento desses sistemas. E aumentam também à medida que se desenvolvem e se aperfeiçoam os programas para auxiliar o armazenamento, a recuperação e a análise das imagens. Este desenvolvimento e este aperfeiçoamento constituem o objetivo principal que motivou os estudos e trabalhos apresentados nesta tese.

2.3. Recuperação de Imagens por Conteúdo

As tecnologias atuais nos permitem gerar, examinar, transmitir, armazenar e manipular grandes quantidades de imagens digitais. Na prática, em muitos sistemas, a análise das imagens é realizada com base em descrições textuais associadas. Apesar de ser útil, há vários problemas com essa abordagem, como discutido na Seção 2.1. Em outras palavras, além das consultas baseadas em textos simples, deve-se permitir que o usuário pesquise imagens em grandes bases de dados, usando esboços, *layouts* ou descrições da estrutura intrínseca das imagens, tais como texturas, cores, imagens-exemplo e outras informações icônicas e gráficas. Esta técnica de pesquisa é chamada de Recuperação de

Imagens por Conteúdo (CBIR) [Vailaya, A. '01] [Antani, S. K. '04] [Smeulders, A. W. M. '00] [Saccavini, C. '01] [Figueiredo, J. C. B. '01b].

Em diversas aplicações, a combinação de texto com informações de conteúdo para a recuperação das imagens é uma idéia que merece ser considerada [Antani, S. K. '05] [Inoue, M. '05]. Porém, mesmo nesse caso, o desenvolvimento de métodos eficientes para lidar com características intrínsecas e para realizar comparações por similaridade é crucial para se atingir bons resultados.

CBIR apresenta diferenças significativas quando comparada às consultas tradicionais: não há comparação exata, mas sim aproximações baseadas em critérios de similaridade. Em outras palavras, técnicas de CBIR servem como filtros de informações e simplesmente reduzem a busca para o usuário que irá, ao final, descartar recuperações falsas ou irá visualmente percorrer as imagens recuperadas e selecionar aquelas que ele deseja.

As características descritivas do conteúdo intrínseco das imagens são obtidas a partir de técnicas das áreas de Visão Computacional e Processamento de Imagens. No Apêndice A essas características visuais intrínsecas serão abordadas de forma detalhada, juntamente as técnicas clássicas de Processamento de Imagens.

A Figura 2.3 apresenta os principais processos e fluxos de dados envolvidos em um ambiente CBIR. A imagem de consulta passa por um processo de extração de características, a partir do cálculo de propriedades locais, no qual é gerada uma assinatura (vetor de características). Esta assinatura é então submetida a uma busca por similaridade junto à estrutura de indexação que contém as assinaturas de todas as imagens armazenadas na base de imagens. Os identificadores das imagens resultantes da busca são utilizados então para recuperar essas imagens da base de imagens, podendo as mesmas, assim, ser apresentadas para o usuário.

Os esforços empregados nos últimos anos por numerosos projetos de pesquisas direcionados à recuperação de imagens por conteúdo geraram diversos sistemas, acadêmicos e comerciais, dentre os quais podemos citar: QBIC [Niblack, W. '93], Chabot [Ogle, V. E. '95], Virage [Bach, J. R. '96], Photobook [Pentland, A. '96], Excalibur [Feder, J. '96], VisualSEEK [Smith, J. R. '97], ImageRover [Sclaroff, S. '97], SIMPLIcity [Li, J. '00], CLIMS [Kao, O. '00], BlobWorld [Carson, C. '02], eID [Stan, D. '03] e SRIS-HC [Traina, A. J. M. '03b], entre outros. Estes sistemas permitem ao usuário formular consultas usando combinações de atributos de baixo nível das imagens. As consultas são especificadas explicitamente fornecendo-se os valores desejados de atributos ou implicitamente especificando-se uma imagem de busca. Alguns sistemas também usam a organização das características da

imagem, de forma que a similaridade é determinada não apenas pela existência de certas características, mas também pela sua disposição absoluta ou relativa na imagem [Cohen, S. '99] [Li, J. '00] [Stehling, R. O. '02]. Os primeiros sistemas focavam no mecanismo de pesquisa (dada uma consulta, encontrar as melhores combinações) e não usavam consultas realizadas anteriormente para melhor entender o que o usuário está procurando. Sistemas mais recentes permitem ao usuário refinar a busca indicando a relevância (ou irrelevância) de imagens no conjunto obtido (*relevance feedback*) [Rui, Y. '97; Zhang, H. '02]. O ser humano tende a utilizar conceitos de alto nível no seu dia-a-dia, fato que acaba por contrastar com o uso de características de baixo nível, tais como as discutidas aqui, produzindo o que é chamado de "gap semântico".

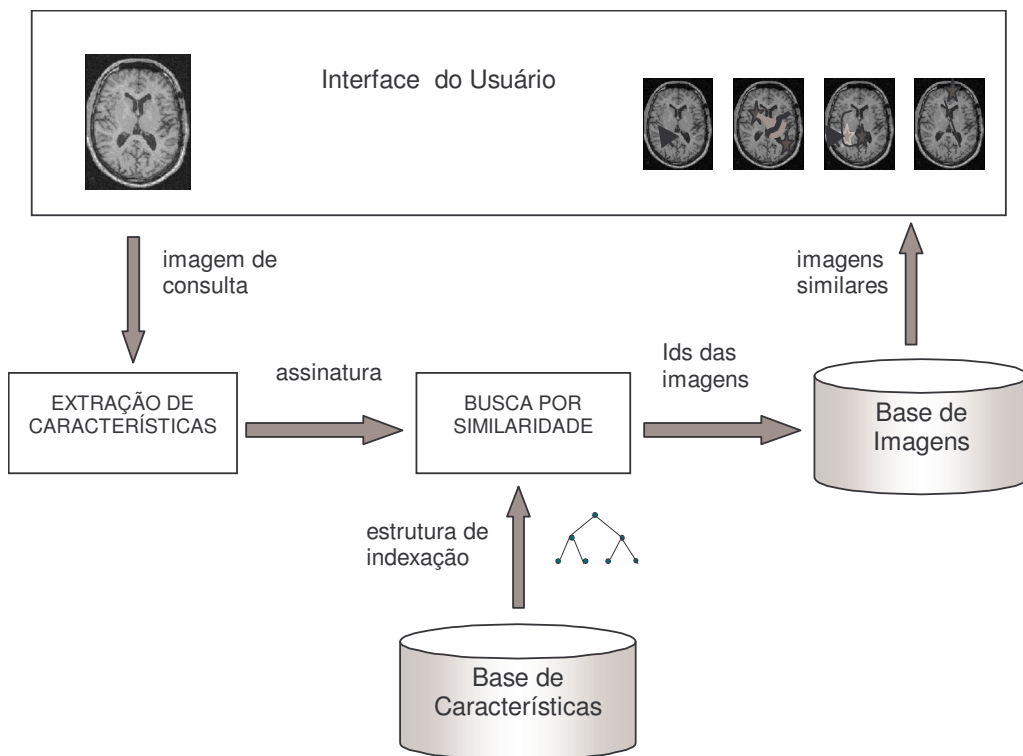


Figura 2.3. Processos e fluxos de dados envolvidos em um ambiente CBIR

Um desafio importante para a abordagem de CBIR é determinar um conjunto de atributos que: (a) descrevam o conteúdo de uma imagem; (b) aceitem alguma medida de similaridade e (c) possam formar a base de uma estrutura de indexação na coleção de imagens. Esses conjuntos de atributos irão compor os vetores de características que representarão a imagem. Entretanto, há uma incompatibilidade entre esses vetores e as ferramentas de indexação: a dimensionalidade dos vetores é frequentemente alta, levando a

maioria dos métodos de indexação a explosões exponenciais, num processo conhecido como "maldição da alta dimensionalidade" [Kom, F. '01] [Beyer, K. '99]. Tais questões constituem um foco de gradual interesse por parte da comunidade de pesquisadores da área de Banco de Dados.

Existe uma diversidade de métodos comparativos e de funções de distância usados para se determinar a similaridade entre duas imagens. A escolha do método e da função de distância a serem utilizados depende de uma série de fatores, tais como o contexto, as características focadas e as propriedades das imagens. No Capítulo 3 as funções de distância e os métodos comparativos serão abordados de forma detalhada.

Além da busca por conteúdo, outras importantes tarefas relacionadas à recuperação e à mineração de informações também baseiam-se em critérios de similaridade para obter seus resultados. Entre estas, podemos destacar: (a) Classificação: dado um objeto Q, direcioná-lo para uma classe, dentre duas ou mais classes pré-definidas; (b) Agrupamento: determinar grupos de itens similares de acordo com algum critério preestabelecido. Até mesmo quando o sistema simplesmente realiza uma varredura no banco de dados em busca de imagens desconhecidas, sem o uso de uma especificação ou de uma imagem de busca, a similaridade pode ser utilizada para a organização de apresentação das imagens recuperadas, através da ordenação das imagens similares.

2.3.1. Características de Bancos de Dados de Imagens

Características específicas de bancos de dados de imagens e, de forma geral, de dados multimídia, quando comparados com bancos de dados relacionais convencionais, incluem:

- O tamanho de cada item de dado é elevado: imagens coloridas são da ordem de alguns MBytes; uma hora de vídeo com qualidade de vídeo-cassete, de MPEG comprimido é da ordem de 1 GByte; um único exame de TC (Tomografia Computadorizada) normalmente contém centenas de imagens representando cortes seqüenciais. Isso impõe exigências severas não apenas em armazenamento, mas também em recuperação de dados (resultados finais de consultas, inspeções, etc).
- Recuperação baseada em conteúdo e similaridade pode ser exigida em adição à recuperação textual. O número de atributos pode ser alto (da ordem de 100 ou mais). A função de distância não é necessariamente Euclidiana, freqüentemente envolvendo interdependência entre atributos.

- Uso de bases de dados distintas para informações textuais e imagens, para otimizar o desempenho da recuperação de dados.
- A interface de consulta deve ser o mais amigável possível (ou seja, permitir ao usuário selecionar cor, forma e textura através de meios gráficos) e também possibilitar retorno com relevância e navegação dirigida pelo usuário.
- O tempo de resposta da consulta deve ser rápido, já que a mesma é tipicamente interativa.
- Alterações são raras. A base de dados frequentemente possui uma natureza de arquivo (ou seja, "*append only*") e pode até mesmo ser estática (por exemplo, uma coleção de imagens distribuída em CD-ROM ou disco ótico).

Essas características sugerem que, para muitas aplicações, podemos pré-computar e armazenar os atributos [Smeulders, A. W. M. '00], a fim de poder executar consultas baseadas nos atributos, comparando as imagens de forma indireta. Além disso, com bases de dados estáticas podemos fazer um monitoramento fino de nossos métodos de indexação sem nos preocuparmos com inserções e com reorganizações.

Bases de imagens armazenam as imagens propriamente ditas, representadas em formato binário, após a sua geração em forma digital ou após o processo de digitalização, no caso de imagens originalmente "análogas".

O armazenamento de imagens pode acontecer de duas maneiras distintas: por meio de um SGBD ou por meio de uma hierarquia de diretórios. No primeiro caso, as imagens são armazenadas como um conjunto de bits, usando-se o tipo de dado *Binary Large Object* (BLOB). No segundo caso, cada imagem fica armazenada em um arquivo, que por sua vez se aloja em um diretório específico, acessado pelo sistema através do seu nome. Nesse caso, o sistema tem que realizar todo o controle de localização, acesso e distribuição das imagens nos respectivos arquivos e diretórios.

Indexação em Bancos de Dados de Imagens

As estruturas de indexação são ferramentas fundamentais que habilitam os sistemas gerenciadores de bases de dados a eficientemente armazenar e recuperar informações de um grande volume de dados. As primeiras estruturas de indexação foram desenvolvidas para operar em dados que atendessem à relação de ordem total (tais como a família de árvores B-tree [Comer, D. '79]), ou que tivessem relativamente poucos dados (estruturas de *hashing* [Samet, H. '95], por exemplo). As primeiras estruturas de indexação destinadas a suportar dados espaciais foram propostas para espaços de poucas dimensões. Estas estruturas são

amplamente descritas na literatura, e são chamadas de Métodos de Acesso Espaciais (*Spatial Access Methods* – SAM). O desenvolvimento dessas estruturas iniciou-se com o trabalho pioneiro sobre as R-Trees [Guttman, A. '84] e prosseguiu com numerosas variações e adaptações [Beckmann, N. '90] [Hellerstein, J. M. '95]. Em [Gaede, V. '98] encontra-se uma descrição e comparação geral sobre estruturas espaciais.

No entanto, para dados em espaços de altas dimensões ou mesmo adimensionais, nenhuma dessas estruturas mostrou-se adequada. Estruturas de dados para espaços métricos (que englobam tanto dados espaciais com dimensão definida quanto dados adimensionais) foram propostas para suportar buscas por similaridade.

Formalmente um **espaço métrico** é um par $M=(D, d)$, onde D é o domínio do vetor de atributos - que são as chaves de indexação, e $d()$ é uma função distância que satisfaz as seguintes propriedades:

1. **Simetria:** $d(O_1, O_2) = d(O_2, O_1)$;
2. **Não negatividade:** $0 < d(O_1, O_2) < 4, O_1 \neq O_2$;
3. **Auto-similaridade:** $d(O_i, O_i) = 0 \quad O_i$;
4. **Desigualdade triangular:** $d(O_1, O_2) \leq d(O_1, O_3) + d(O_3, O_2)$.

onde: O_1, O_2, O_3 e O_i são objetos pertencentes ao espaço D .

Quando a função de distância satisfaz essas propriedades, dizemos tratar-se de uma **distância métrica**.

As estruturas métricas, ou Métodos de Acesso Métricos (*Metric Access Methods* – MAM), suportam naturalmente consultas por similaridade além de se mostrarem eficientes para dados de dimensões altas. Desse modo, tem-se mostrado bastante apropriado utilizar um MAM para indexar imagens ou, mais propriamente, os atributos que foram extraídos das imagens, suportando busca por similaridade. Utilizando os atributos extraídos previamente da imagem, um MAM constrói a estrutura de índices calculando as distâncias entre elas. A partir da primeira proposta de indexação de dados métricos [Burkhard, W. A. '73], diversos avanços têm sido alcançados no sentido de se obter estruturas eficientes e compactas [Shasha, D. '90] [Uhlmann, J. K. '91] [Ciaccia, P. '97] [Traina-Jr, C. '00a]. Descrições gerais sobre as principais técnicas de indexação para espaços métricos são apresentadas em [Chávez, E. '01] e em [Hjalton, G. '03].

2.3.2. Tipos de Consultas

As características de imagens que são investigadas nesta tese são: distribuição de cor, textura e forma. Estas características possuem aplicabilidade ampla e podem ser obtidas de forma automática ou semi-automática. Tanto para uma cena quanto para um objeto ou uma região de interesse da imagem, o usuário pode formular uma consulta sobre qualquer uma dessas propriedades ou uma combinação lógica das mesmas.

Todas as consultas partem do princípio da similaridade. Como exemplo, se um usuário está interessado em recuperar uma cena de praia, ele pode formular uma consulta contendo características de distribuição de cores (por exemplo, 35% de branco e 65% de azul na área da imagem) e texturas (presença de textura de areia). O sistema CBIR irá retornar imagens com essas propriedades. Os resultados irão indicar cenas de praia, juntamente com falsos resultados (imagens que por acaso possuem distribuição de cores semelhante). Isto não representa um problema para o usuário, já que o sistema visual humano é ótimo em focalizar rapidamente itens de interesse e descartar padrões indesejados, desde que não haja muitos destes. Porém, tais resultados espúrios podem diminuir o grau de aceitação do sistema pelos usuários.

Além da definição dos atributos, o usuário deverá também especificar o grau de cobertura da consulta sobre os resultados obtidos. Essa cobertura poderá determinar a quantidade de imagens a serem recuperadas, ou então o nível de similaridade das mesmas, gerando assim dois tipos de consultas: consulta por abrangência e consulta aos vizinhos mais próximos.

Consultas por Abrangência

Dados: um conjunto de objetos $O = \{O_1, O_2, \dots, O_n\}$ pertencentes a um domínio D , uma função de distância $d()$, um objeto de consulta $Q \in D$ e uma distância de busca máxima $r(Q)$, a consulta por abrangência

$$\text{range}(Q, r(Q)) = \{ O_i \mid O_i \in D \text{ e } d(O_i, Q) \leq r(Q) \},$$

seleciona todos os objetos O_i do conjunto de dados tal que eles estejam a uma distância menor ou igual a $r(Q)$ do objeto de busca. Um exemplo deste tipo de consulta seria: “Encontre as estrelas que estão até 10 anos-luz de distância do Sol”. Nesse caso o objeto de consulta é "Sol", o domínio D é o conjunto de estrelas do Universo, e o raio de busca (distância máxima) é 10 anos-luz. A distância utilizada é a medida astronômica que mede o espaço em anos-luz.

Consultas aos Vizinhos mais Próximos

Dados um conjunto de objetos $O = \{O_1, O_2, \dots, O_n\}$ de um domínio D , uma função de distância métrica $d()$, um objeto de consulta $Q \in D$ e um número inteiro $k \geq 1$, a consulta aos k Vizinhos mais Próximos (k -NN) seleciona os k objetos do conjunto de dados que estão mais próximos de Q . Isto é,

$$k\text{-NN}(Q) = \{A_i \mid A_i \in D, A_i \notin \{Q\}, |A| = k \text{ e } A_i \in A, O_i \in D - A, d(Q, A_i) \leq d(Q, O_i)\}.$$

Em caso de empate na maior distância obtida, pode-se selecionar apenas o número necessário de objetos que satisfazem a regra, ou criar uma lista de empates. Utilizando o exemplo anterior, a consulta “Selecione as 5 estrelas mais próximas do Sol” é uma busca **5-NN** para o objeto "Sol" sobre o mesmo domínio de objetos da consulta-exemplo anterior.

2.4. Recuperação por Conteúdo de Imagens Médicas

CBIR aplicado a imagens médicas apresenta peculiaridades e especificidades inerentes a dois aspectos: a) o estado da arte em processamento e análise das imagens médicas e b) a integração de CBIR ao ambiente computacional hospitalar, ou seja aos sistemas PACS.

2.4.1. Processamento e Análise de Imagens Médicas

O processamento e a análise computadorizada de imagens médicas engloba diversas áreas de estudo potenciais, dentre as quais destacam-se a aquisição, reconstrução tridimensional, realce, compressão, armazenamento, indexação, análise e visualização de imagens. No que diz respeito à análise de imagens médicas, algumas técnicas podem ser citadas: segmentação, *matching*, detecção de alterações em seqüências, reconhecimento e classificação de texturas, reconhecimento e classificação de formas, estimativa de área ou volume e medição de parâmetros anatômicos e fisiológicos. Devido à grande amplitude de escopos de estudo, a maioria das pesquisas desenvolvidas tentam concentrar-se fortemente em conjuntos específicos de imagens, assim como em objetivos clínicos ou biológicos bem definidos.

Em [Duncan, J. S. '00] é realizada uma avaliação do progresso apresentado pelo campo da análise de imagens médicas nas últimas décadas, identificando questões importantes que se encontram ainda em aberto. Atualmente, tecnologias computacionais (*hardware* e *software*) avançadas facilitam o trabalho baseado em imagens, assim como visualizações mais realistas. Renderização e reconstrução de contornos e volumes complexos em tempo real, suportadas

por *hardware*, permitem a exploração e a análise interativa de grandes montantes de imagens médicas. Existe uma variedade de equipamentos comerciais, disponibilizando recursos simples de visualização e análise. Plataformas de *software* e *hardware* de tele-presença para cirurgias à distância, guiadas por imagens, encontram-se em pleno desenvolvimento [Sorid, D. '00]. Sistemas de realidade virtual acoplados a interfaces altamente imersivas possibilitam a realização de jornadas interativas pelo corpo humano. Esses recursos têm sido de grande utilidade para o ensino e o treinamento em medicina.

2.4.2. Incorporação de CBIR aos Sistemas PACS

A recuperação eficiente de imagens é uma das metas mais importantes a serem atingidas em sistemas PACS. Com o auxílio de um sistema desse tipo, a manipulação das imagens de exames de pacientes e seus exames poderá vir a ser totalmente realizada de forma digital [Bueno, J. M. '02c].

A Figura 2.4, adaptada de [Lehmann, T. M. '03a], apresenta uma proposta de sistema para integração entre PACS e CBIR. Nela podemos ver os principais componentes dos ambientes PACS e CBIR em diferentes níveis, juntamente com as interfaces de aplicação, programação, banco de dados e comunicação.

CBIR no domínio médico é mais desafiador do que num domínio de imagens de propósito geral. A principal razão está no fato de que as características importantes de imagens médicas encontram-se normalmente em regiões localizadas, porém de natureza difusa e que não podem ser assumidas como fazendo parte de alguma classe previamente definida. Isto torna a extração e a comparação das características muito mais complexa, porque os atributos extraídos das imagens deverão descrever seus detalhes, assim como permitir a recuperação rápida de imagens similares.

2.4.3. Exemplo de CBIR / PACS

Os sistemas de consulta que têm sido desenvolvidos, em geral, possuem duas características: são voltados para áreas específicas, tais como técnicas de análise de mamografias; e incorporam algum tipo de anotação ou então de interação com o usuário.

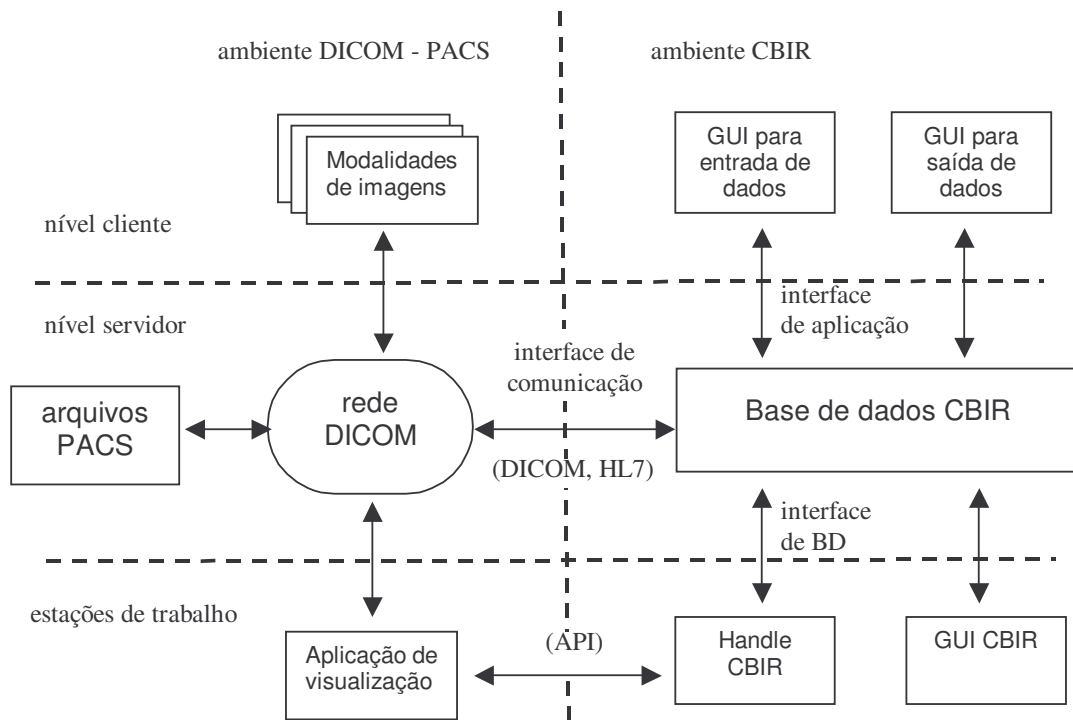


Figura 2.4. Sistema de integração entre PACS e CBIR

Em [Rosa, N. A. '02] é apresentado um sistema CBIR intitulado **SRIS-HC** (Sistema de Recuperação de Imagens Similares – Hospital das Clínicas), que foi desenvolvido com a finalidade de demonstrar a viabilidade de recuperação de imagens por conteúdo em um ambiente hospitalar real. Esse sistema é parte integrante do projeto de um sistema PACS chamado **cbPACS** [Bueno, J. M. '02a], cujo desenvolvimento vem sendo realizado através de pesquisa comum entre o Grupo de Bases de Dados e Imagens (GBDI) do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo (USP) e o Centro de Ciências das Imagens e Física Médica (CCIFM) da Faculdade de Medicina de Ribeirão Preto (FMRP) da USP.

A partir do Sistema de Laudo Eletrônico do Serviço de Radiodiagnóstico do Hospital das Clínicas (HC) da FMRP, desenvolvido numa parceria entre o CCIFM e o Centro de Informações e Análises do HCFMRP, o SRIS-HC foi implementado mantendo total integração com o padrão dos módulos de consultas do Sistema de Laudo Eletrônico. O SRIS-HC poderá atuar como uma ferramenta de apoio ao diagnóstico, permitindo a recuperação de imagens associadas a todas as informações relacionadas ao exame do paciente, contidas no Sistema de Laudo Eletrônico, tais como informações do paciente, dados textuais dos exames, laudos médicos e diagnósticos. A interface de consulta do SRIS-HC é mostrada na Figura 2.5.

Duas técnicas de extração de características de imagens foram adotadas no SRIS-HC: o histograma tradicional e o histograma métrico, um novo descritor construído sobre os histogramas tradicionais de imagens (ver Seção A.8.4). A estrutura de indexação utilizada foi a árvore métrica Slim-Tree [Traina-Jr, C. '00a]. O sistema suporta os dois tipos principais de consultas por similaridade: por abrangência e aos vizinhos mais próximos.

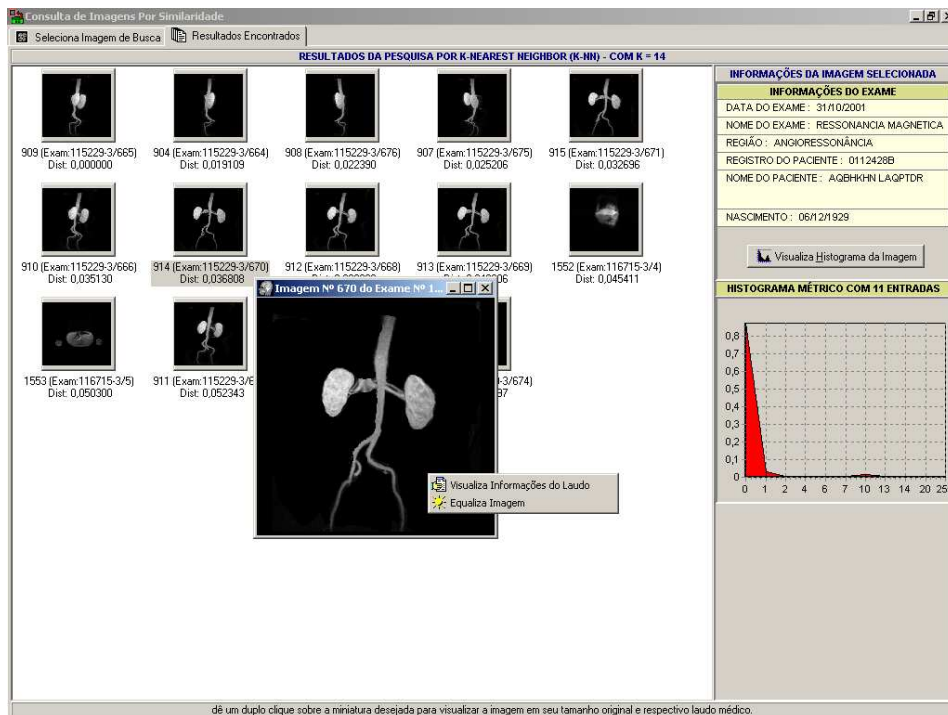


Figura 2.5. Interface do SRIS-HC, mostrando os resultados de uma consulta por vizinhança próxima [Rosa, N. A. '02]

Além deste sistema, também dentro da parceria entre o GBDI e o CCIFM, um outro sistema que se integra a este está em desenvolvimento. Esse sistema, denominado CIRCE (*Content-based Image Retrieval Core Engine*) [Araujo, M. R. B. '02] [Figueiredo, J. M. '05], tem como objetivo estender as funcionalidade de um SGBD relacional para prover funcionalidades de busca por conteúdo em imagens. A Figura 2.6 mostra a sua arquitetura geral. Os módulos principais do CIRCE são: Interpretador de Consultas (IC), Extrator de Parâmetros (XP) e Métodos de Acesso (MA). O módulo IC inclui a extensão da linguagem SQL para suportar imagem como mais um tipo de dado, permitindo que o usuário especifique atributos de relações como sendo desse tipo. O módulo MA atua como interface do CIRCE com estruturas de acesso, como é o caso da estrutura Slim-Tree. O módulo XP atua no sentido de obter das imagens dados que servirão de parâmetros para os métodos de acesso.

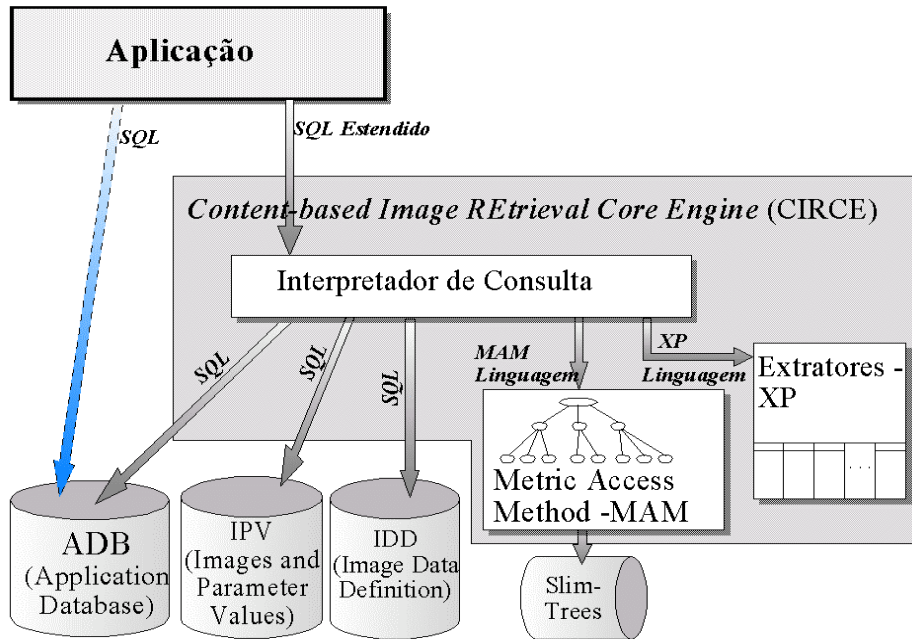


Figura 2.6. Arquitetura geral do CIRCE [Figueiredo, J. M. '05]

2.5. Desafios

Na prática de CBIR sobre imagens médicas, muitas dificuldades ainda permanecem para serem resolvidas, tanto no que concerne à definição das necessidades práticas da comunidade de usuários (médicos e cientistas), quanto no desenvolvimento de substratos teóricos que possam alicerçar a pesquisa científica.

Um ponto importante a ser trabalhado é a ampliação dos métodos e técnicas apresentados até então como eficazes apenas sob escopos limitados.

O uso de modelos físicos para prever fenômenos temporais, tais como deformações de órgãos, evolução de tumores ou taxa de fluxo sanguíneo, também apresenta-se como um campo ainda em estágio inicial e que demanda grandes esforços até atingir resultados maciços. Nesse caso, o uso das técnicas de séries temporais será uma importante alavanca para gerar resultados.

Outro aspecto que deve ser considerado prioritário é a integração de tarefas de análise de imagens, que naturalmente estão correlacionadas, mas que normalmente são tratadas isoladamente.

Finalmente, identificamos a necessidade de se desenvolver, dentro da comunidade de pesquisa, abordagens apropriadas e integradas para validação e avaliação de experimentos e métodos. Faltam conjuntos de imagens catalogadas e classificadas disponíveis para testes. E

falta o desenvolvimento de metodologias eficazes de avaliação de resultados, de forma que algoritmos novos possam ser devidamente comparados, sem distorções locais.

2.6. Considerações Finais

Neste capítulo foram apresentados os conceitos e definições de CBIR, PACS e consultas por similaridade, bem como a sua integração. Analisando tal contextualização, foi possível identificar e discutir os desafios e algumas das principais questões ainda abertas nestas áreas.

Capítulo 3

3. FUNÇÕES DE DISTÂNCIA E MÉTODOS COMPARATIVOS

Neste Capítulo, apresentamos uma revisão sobre funções de distância e métodos comparativos para busca por similaridade de imagens, abordando propriedades e aplicações dos mesmos.

3.1. Introdução

Um sistema computacional que efetue o armazenamento e a manipulação de imagens deve possuir três componentes fundamentais, que atuam de forma integrada: (a) um conjunto de extratores de atributos (ou descritores) das imagens; (b) um conjunto de métodos comparativos - cada qual utilizando uma função de distância - que sejam compatíveis com os descritores e (c) uma ou mais estruturas de indexação.

Um **extrator** de atributos é um programa computacional responsável por efetuar uma análise de baixo nível no conteúdo da imagem, calcular propriedades locais e gerar, de forma genérica, uma **assinatura** (ou vetor de características) da mesma. Normalmente, os sistemas CBIR disponibilizam um conjunto de diferentes extratores, cada qual responsável por gerar uma diferente assinatura, relacionada a um diferente descritor original da imagem, partindo de uma preparação específica sobre os dados originais.

Os **métodos comparativos** são procedimentos que definem de que forma as assinaturas serão comparadas – podendo, inclusive, já estar parcialmente embutidos no próprio método de geração das mesmas – a fim de se obter a distância entre as imagens consideradas. A formulação matemática utilizada em cada método para se calcular a distância, a partir das assinaturas, é a **função de distância**. No momento de comparar um par de assinaturas, a mesma função de distância pode ser utilizada de diferentes maneiras, ou seja, com diferentes métodos comparativos. Em Seções subseqüentes esses conceitos serão exemplificados.

Os algoritmos de indexação geram e organizam estruturas de dados (geralmente, árvores) onde são armazenadas as assinaturas das imagens de forma a permitir comparações entre as mesmas. Considerando uma mesma base de dados, para cada diferente assinatura e cada diferente método comparativo, tem-se uma árvore específica. Tal mecanismo permite a recuperação eficiente da informação, segundo os parâmetros explicitados pelo usuário.

A especificação de um método comparativo inclui a definição de uma função de distância específica a ser utilizada pelo mesmo, sempre almejando a obtenção dos melhores resultados, considerada a sua semântica no contexto de aplicação do sistema.

Quando analisamos a maioria dos métodos comparativos propostos e apresentados na literatura, verificamos ser verdadeira a afirmação de [Vasconcelos, N. '00], segundo a qual parece ainda não haver um entendimento claro sobre o sentido matemático e sobre os efeitos gerados pelas diferentes funções de distância, assim como sobre suas fraquezas. Na prática, funções de distância são freqüentemente selecionadas sem que haja uma justificativa teoricamente fundamentada.

A escolha das funções de distância a serem utilizadas, assim como dos métodos comparativos que as utilizam, devem levar em conta diversas características do respectivo sistema:

- a natureza estatística do contexto;
- os tipos de descritores escolhidos;
- os tipos de dados relativos aos descritores analisados;
- o tipo de pré-processamento que foi aplicado aos dados;
- o inter-relacionamento entre os diversos descritores;
- as peculiaridades semânticas do ambiente de aplicação.

Nas próximas Subseções será apresentada uma revisão das principais funções de distância de uso mais geral, considerando os seus aspectos relativos às características citadas. Analisaremos também alguns dos principais métodos comparativos propostos na literatura.

3.2. A Distância Minkowski (Lp)

O tipo de função de distância mais conhecido e utilizado é a Distância Euclidiana, a qual representa uma das variações da Distância Minkowski, que por sua vez é baseada nas normas Lp.

Consideremos um conjunto de objetos, cada qual possuindo n atributos e uma assinatura que consiste em um vetor de valores dos n atributos. Para dois objetos Q e C, representados pelos vetores $q=(q_1, q_2, \dots, q_n)$ e $c=(c_1, c_2, \dots, c_n)$, a distância Minkowski será dada por:

$$d(Q, C) = \sqrt[p]{\sum_{i=1}^n |c_i - q_i|^p} \quad (3.1)$$

De acordo com o valor atribuído a p , teremos as variações:

$p = 1$: distância **City Block** ou **Manhattan** (L_1)

$p = 2$: distância **Euclidiana** (L_2)

$p = 4$: distância **Infinity** ou **Chebychev** (L_4)

A distância L_2 define o lugar geométrico de todos os pontos equidistantes do ponto que representa o objeto de consulta, ou seja, uma circunferência no espaço 2D (uma casca esférica no espaço 3D) com centro no objeto de consulta (Figura 3.1).

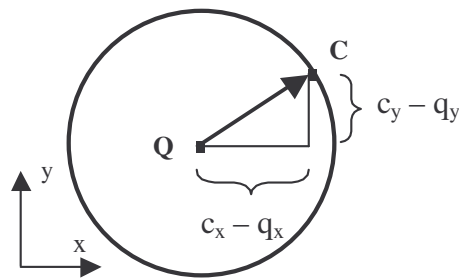


Figura 3.1. Lugar geométrico definido pela distância L_2

A equação da distância L_2 no espaço 2D, considerando atributos x e y , ficará:

$$d(Q, C) = \sqrt{(c_y - q_y)^2 + (c_x - q_x)^2} \quad (3.2)$$

Muitas vezes, a fim de otimizar o tempo de cálculo, a L_2 é usada na sua forma quadrática (sem a extração da raiz quadrada). Esta forma preserva a distância entre os pontos, retornando o mesmo conjunto de objetos no resultado.

A distância L_1 , por sua vez, define o lugar geométrico de todos os pontos que possuem o mesmo valor da soma $|c_y - q_y| + |c_x - q_x|$ em 2D (Figura 3.2). É como se um deslocamento de Q até C somente fosse possível por meio de segmentos lineares paralelos ou perpendiculares entre si, da forma como ocorre nas ruas de uma cidade (por isso, o nome "city block"). A distância será a soma dos segmentos em x e y .

A equação da L_1 no espaço 2D ficará:

$$d(Q, C) = |(c_y - q_y)| + |(c_x - q_x)| \quad (3.3)$$

A distância L_∞ pode ser aproximada por:

$$d(Q, C) = \max_{i=1}^n |c_i - q_i| \quad (3.4)$$

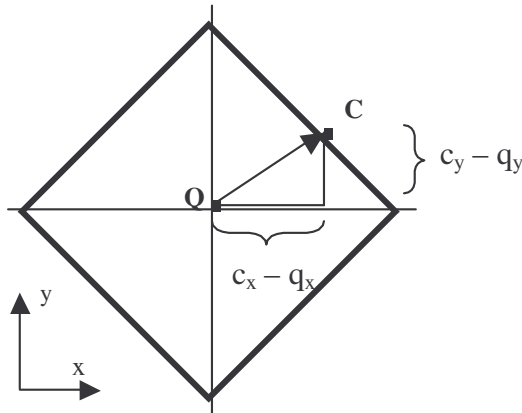


Figura 3.2. Lugar geométrico definido pela distância L_1

O lugar geométrico da L_∞ em 2D é mostrado na Figura 3.3. A distância entre Q e C é determinada unicamente pelo valor representado no eixo no qual a diferença entre as coordenadas dos pontos é a maior.

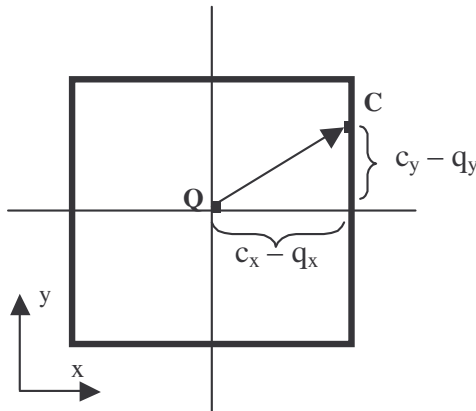


Figura 3.3. Lugar geométrico definido pela distância L_∞

Para outros valores de p , o *locus* de todos os pontos que estão igualmente distanciados do ponto Q produz diversos contornos convexos simétricos, intermediários entre o octaedro ($p=1$), a esfera ($p=2$) e o cubo ($p=4$). A Figura 3.4 apresenta os diferentes contornos definidos pelos diferentes valores de p , no espaço 3D.

Em [Akleman, E. '99] os autores propõem uma versão generalizada da família L_p cujo objetivo é flexibilizar o formato do contorno, podendo-se obter novos polígonos convexos. Essa generalização se dá através da extensão dos operadores de Minkowski (L_p), multiplicando-os por conjuntos de vetores apropriados. Os resultados obtidos, exemplificados na Figura 3.5, mostram que novos formatos de poliedros são obtidos. Esse método poderia ser

utilizado para se proceder a "ajustes finos" de acordo com o inter-relacionamento entre os atributos, em situações específicas.

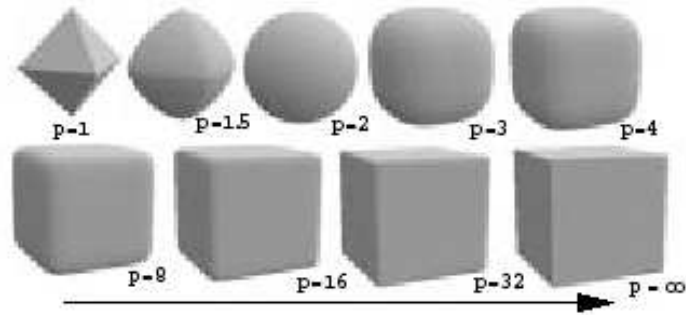


Figura 3.4. Lugar geométrico definido pelos diferentes valores de p , no espaço 3D

Figura 3.5. Outros poliedros gerados pela extensão dos operadores de Minkowski [Akleman, E. '99]



3.2.1. Distância Minkowski Ponderada

Se houver diferentes influências entre os atributos sobre o valor final da similaridade, poderá ser utilizado um vetor de ponderação $W = (w_1, w_2, \dots, w_n)$ no cálculo das distâncias, atribuindo-se, assim, diferentes "pesos" para cada atributo:

$$d(Q, C) = \sqrt[p]{\sum_{i=1}^n w_i |c_i - q_i|^p} \quad (3.5)$$

A figura geométrica do espaço 2D associada a esta distância, considerando-se que os pesos dos dois atributos seja diferente, é a elipse.

Esta flexibilidade de atribuir diferentes "pesos" aos atributos compatibiliza-se com o trabalho de *relevance feedback*: a partir de um conjunto de teste, levantam-se as distâncias através de avaliações de especialistas do domínio e, com esses resultados em mãos, procuram-se os melhores valores para os elementos de W que ponderam os atributos extraídos da imagem de forma a apresentar resultados os mais próximos possíveis das avaliações. Esses

valores de W poderão ser aperfeiçoados à medida que novas avaliações são introduzidas pelos usuários especialistas.

Além disso, diferentes vetores W poderão ser especificados, para diferentes tipos de tarefas que utilizam as consultas (tecidos similares, nódulos, fraturas, regiões anômalas, etc), considerando-se que diferentes atributos poderão ter maior ou menor relevância do que outros de acordo com o tipo de tarefa.

3.3. Outras Distâncias de Uso Consolidado

Algumas distâncias novas têm sido propostas na literatura. Segundo [Keogh, E. '01a]: "Há dezenas de artigos que propõem novas funções de distância... Nenhuma delas (que eu tenha conhecimento) consegue produzir motivação suficiente para testá-la". A despeito desta posição radical, há algumas funções de distância que já foram bastante utilizadas e que têm demonstrado sua eficiência, conforme discutiremos a seguir.

3.3.1. Distância Mahalanobis

Todas as variações da distância Minkowski são aditivas, no sentido de que cada atributo contribui de forma independente para o cálculo da distância, ou seja, considera-se que o valor de um atributo não influencia o valor de outros. Em muitos contextos, porém, poderá ocorrer uma interdependência (implícita ou não) entre atributos. Nesse caso, dever-se-á levar em consideração a **covariância** entre os atributos. A covariância é uma medida de como dois atributos variam conjuntamente. Se valores altos de um atributo tendem a associar-se com valores altos de outro atributo, ou valores baixos com valores baixos, então a covariância entre esses atributos terá um valor alto positivo. Se valores altos de um atributo associam-se com valores baixos de outro, a covariância terá um valor negativo. O efeito da covariância entre atributos no cálculo da distância é considerado quando se incorpora a matriz de covariância na definição da função de distância. Isso leva à **distância Mahalanobis**, definida como [Theodoridis, S. '99]:

$$d_{MH}(Q, C) = \sqrt{(c - q)^T V^{-1} (c - q)} \quad (3.6)$$

onde V é a matriz de covariância entre os vetores de atributos c e q .

3.3.2. Divergência de Kullback-leibler e Divergência de Jeffrey

A divergência de Kullback-Leibler (DKL) é definida como:

$$d_{KL}(Q, C) = \sum_{i=1}^n c_i \log \frac{c_i}{q_i} \quad (3.7)$$

De um ponto de vista teórico, a DKL mede o quanto ineficiente em média seria codificar uma assinatura usando a outra como referência. Entretanto, a DKL não é simétrica e é sensível, por exemplo, à divisão dos *bins* de um histograma.

A divergência de Jeffrey, derivada empiricamente, é uma modificação da DKL, que se mostra numericamente estável, simétrica e robusta com respeito a ruído e ao tamanho dos *bins* [Rubner, Y. '01]. Ela é definida como:

$$d_J(Q, C) = \sum_{i=1}^n c_i \log \frac{c_i}{m_i} + q_i \log \frac{q_i}{m_i} \quad (3.8)$$

onde: $m_i = (c_i + q_i) / 2$

3.3.3. Medida χ^2

A medida χ^2 é definida como:

$$d_{\chi^2}(Q, C) = \sum_{i=1}^n \frac{(c_i - m_i)^2}{m_i} \quad (3.9)$$

onde: $m_i = (c_i + q_i) / 2$

Essa função valoriza as grandes discrepâncias entre as duas assinaturas comparadas, de forma que duas assinaturas cujos "desenhos" são semelhantes são consideradas mais próximas do que quando há altos e baixos de diferença entre elas.

3.3.4. Interseção de Histogramas

Esta função, como o próprio nome diz, é normalmente utilizada para se medir distância entre histogramas. Sua formulação é a seguinte:

$$d_{IH}(Q, C) = 1 - \left(\frac{\sum_{i=1}^n \min(c_i, q_i)}{\sum_{i=1}^n q_i} \right) \quad (3.10)$$

A Interseção de Histogramas é interessante devido ao fato de ser possível tratar intervalos parciais, quando a área de um histograma é menor que a do outro. Por outro lado, quando as áreas dos dois histogramas são iguais, a Interseção de Histogramas transforma-se na distância L_1 .

3.4. Outras Propostas de Distância

Em [Aggarwal, C. C. '03], o autor discute o problema da rigidez imposta pelas funções de distância, diante da grande variedade de aplicações, domínios e contextos de comparações entre dados. Para responder a essa questão, é proposto um esquema para aplicações de mineração de dados, que consiste de um *framework* onde o usuário, interativamente, "constrói" a função de distância a partir de uma estrutura parametrizada, de forma a tornar esta função mais efetiva e adequada às características do domínio onde ocorre o processo de comparação.

Em [Hiransakolwong, N. '05], os autores propõem uma distância que considera uma pré-classificação de cada atributo em grupos (relevante, irrelevante ou neutro), no cálculo da distância entre duas imagens. Conjuntos de teste são usados para avaliar e definir a faixa de pertinência de cada atributo dentro desses grupos, e os valores dessas faixas são usados para ajustar o peso de cada atributo no cálculo da distância.

Em [Qamra, A. '05], os autores propõem e testam uma função de distância perceptual, baseada em operações dinâmicas realizadas sobre a imagem, com o objetivo específico de buscar por imagens muito parecidas, com aplicação na investigação do uso indevido de réplicas.

Já em [Hertz, T. '04], os autores propõem o uso de classificadores, que são treinados para distinguir pares de imagens que pertençam à mesma classe e pares de imagens que não pertençam. Os algoritmos classificadores particionam o espaço de atributos para separar as diferentes classes e a margem de diferenciação definida entre os valores dos atributos é usada posteriormente como função de distância para classificar os objetos.

3.4.1. Considerações sobre Tipos de Dados

Para tipos de dados **binários**, uma estratégia que pode ser utilizada consiste em se computar o número de atributos para os quais dois objetos possuem o mesmo valor e valores distintos, montando-se a matriz de classificação cruzada das variáveis [Hand, D. J. '01], mostrada na Tabela 3.1.

O **coeficiente de similaridade** entre i e j é definido como:

$$sim(i, j) = (n_{1,1} + n_{0,0}) / (n_{1,1} + n_{1,0} + n_{0,1} + n_{0,0}) \quad (3.11)$$

E a distância entre i e j :

$$d(i, j) = 1 - sim(i, j) \quad (3.12)$$

Tabela 3.1. Matriz de classificação cruzada entre dois objetos

\	j = 1	j = 0
i = 1	$n_{1,1}$	$n_{1,0}$
i = 0	$n_{0,1}$	$n_{0,0}$

onde: i e j são dois objetos distintos e

$n_{i,j}$ é a quantidade de atributos para os quais ambos os objetos i e j possuem valores iguais a i .

Em algumas situações, como, por exemplo, quando os valores 0 e 1 indicam a ausência ou a presença de certa propriedade, não se deve incluir o valor $n_{0,0}$ na fórmula de cálculo do coeficiente de similaridade, podendo-se, inclusive, considerar os valores $n_{1,0}$ e $n_{0,1}$ com peso igual a $\frac{1}{2}$.

Quando ocorre uma **heterogeneidade entre tipos de atributos**, mais especificamente com a mistura de atributos numéricos lineares com atributos nominais, faz-se necessário estabelecer um tratamento que permita a integração dos mesmos para se chegar a valores corretos de distância. Em [Wilson, D. R. '97] os autores propõem uma função de distância composta para tratar da integração entre atributos contínuos e nominais (categóricos).

3.4.2. Funções de Distância Métricas

Na Seção 2.3.1 foi definida a **função de distância métrica** como sendo aquela que obedece às 4 propriedades dos espaços métricos: simetria, não negatividade, auto-similaridade e desigualdade triangular.

Particularmente com relação à propriedade de **desigualdade triangular**, o cumprimento da mesma por parte da função de distância é importante devido ao ganho de desempenho propiciado por ela junto às estruturas de indexação. O processo de "poda" das árvores faz uso dessa propriedade, gerando, assim, um ganho considerável no tempo de processamento de busca na árvore. Esse ganho se deve à possibilidade de eliminação de objetos candidatos sem a necessidade de calcular a distância dos mesmos ao objeto de busca.

A despeito desta vantagem, existe um aspecto semântico freqüentemente levantado pelos pesquisadores, com respeito à desigualdade triangular, que chamaremos aqui de "efeito centauro", segundo o qual, em determinadas situações, a desigualdade triangular não condiz com a percepção que se tem da inter-relação entre três objetos. O exemplo clássico, mostrado na Figura 3.6, é o distanciamento entre os objetos "homem", "cavalo" e "centauro".



Figura 3.6. Efeito centauro: incongruência da desigualdade triangular

A falha semântica que caracteriza o efeito centauro consiste na observação de que o centauro está relativamente próximo tanto do cavalo quanto do homem, embora o cavalo e o homem não tenham nenhuma característica similar entre si. Dessa forma não teríamos a formação do triângulo entre os três objetos (lembrando que para existir um triângulo, qualquer das combinações possíveis entre os seus três lados tem que obedecer à desigualdade triangular). Semanticamente, o fato de não existir atributos comuns entre o cavalo e o homem deveria fazer com que a distância entre os mesmos fosse infinitamente grande.

Assim como a desigualdade triangular, a propriedade de **simetria** também é alvo de críticas que contestam a sua validade baseando-se em resultados experimentais [Chang, E. Y. '00]. Por exemplo, suponhamos que Q seja mais "saliente" do que C. Na comparação de Q com C, ao se tomar Q como referência, em testes de percepção, obtém-se um grau de similaridade maior do que ao se tomar C como referência, contrariando, assim, a propriedade da simetria.

Além disso, diversos autores argumentam que as distâncias métricas impõem restrições muito rígidas ao mecanismo natural de comparação entre imagens. Essas restrições muitas vezes acabarão por se opor às necessidades da natureza dos atributos ou da semântica da aplicação. Dependendo do domínio de aplicação ao qual pertencem as imagens a serem avaliadas, esse modelo métrico poderá ou não se mostrar aceitável no que diz respeito à adequação dos resultados à expectativa dos especialistas do domínio. Essa adequação é discutida em [Smeulders, A. W. M. '00], onde os autores definem o chamado "*gap* semântico", como sendo a divergência entre a informação que pode ser extraída dos dados visuais e a interpretação que os mesmos dados recebem de um usuário em um determinado contexto.

Alguns trabalhos [Chang, E. Y. '00; Rosh, E. '75; Tversky, A. '77; Tversky, A. '82] discutem formas de se buscar a redução do *gap* semântico, através de extensões ou reformulações dos modelos matemáticos correntemente utilizados, independentes de obedecerem às propriedades dos espaços métricos.

Por sua vez, [Tversky, A. '77] e [Tversky, A. '82] apresentam uma análise da percepção humana no que diz respeito ao conceito de similaridade. Defendem a idéia de que o espaço métrico é um modelo muito rígido e pouco representativo da percepção humana de similaridade, apresentando uma série de resultados experimentais. Propõem um **modelo de contraste** com propriedades diferentes do modelo métrico e obedecendo a nova formulação.

No trabalho realizado em [Jacobs, D. W. '00] os autores afirmam que funções de distância não métricas são mais eficientes em espaços de busca extensos, principalmente quando há ocorrências freqüentes de exceções. Além disso, sistemas de classificação que buscam modelar a percepção humana ou que utilizam métodos muito robustos de comparação entre imagens tendem a resultar em métodos de julgamento da similaridade que não são métricos.

3.5. Funções de Distância Típicas para Descritores Específicos

Considerando-se o campo das características visuais, os descritores de uma imagem podem ser classificados como genéricos ou específicos do domínio [Rui, Y. '97]. Os descritores genéricos contemplam características tais como cor, textura e forma, enquanto os descritores específicos do domínio são dependentes do contexto, como, por exemplo, faces humanas, impressões digitais ou nódulos de tumor.

Devido à subjetividade da percepção humana, não existe uma única representação ótima para um dado descritor, mas sim múltiplas opções de representação que caracterizam o descritor a partir de diferentes perspectivas. Ainda assim, as pesquisas têm indicado a adequação de certas funções de distância para serem usadas com diferentes descritores. As utilizações mais frequentes das funções de distância associadas a dados específicos, segundo [Hand, D. J. '01], são as seguintes:

- vetores de cores: distância Euclidiana simples e distância *City Block*;
- histogramas de cores: distância Mahalanobis ou interseção de histogramas;
- textura: distância Euclidiana Ponderada, distância Chebychev ou distância Mahalanobis;
- forma: distância Euclidiana simples.

Obviamente, a função de distância a ser usada depende diretamente do método comparativo ao qual ela estará associada, e isto faz com que cada caso de aplicação exija um estudo específico para esta escolha.

3.6. Natureza Estatística das Funções de Distância

Em [Gibbs, A. L. '02] um estudo interessante é conduzido, no sentido de se avaliar diversas funções de distância utilizadas para quantificar a distância entre conjuntos de dados probabilísticos (medidas probabilísticas). Muitas delas não são métricas, porém, como destacam os autores, possuem aplicabilidade bem definida e consolidada. O estudo desenvolvido visa a determinar inter-relacionamentos entre as funções de distância, definindo as operações matemáticas que levam de uma função à outra e determinando relações de limite inferior entre as mesmas. A velocidade de convergência é fator crucial para o bom desempenho de uma função de distância aplicada a conjuntos probabilísticos. As funções de distância avaliadas e comparadas nesse estudo são: Discrepância, Distância de Hellinger, Divergência de Kullback-Leibler, distância de Kolmogorov, distância de Lévy, distância de Prokhorov, distância de Separação, Variância Total, distância de Wasserstein e medida Π^2 .

Em [Vasconcelos, N. '00], os autores traçam uma análise das funções matemáticas mais representativas utilizadas para similaridade entre imagens, correlacionando-as de acordo com as restrições teóricas que permeiam a formulação de cada uma. Diversas funções de distância são analisadas e correlacionadas: classificador Bayesiano, Distância Bhattacharyya, Critério de Probabilidade Máxima, Divergência de Kullback-Leibler, Distância Quadrática, Distância

Mahalanobis e Distância Euclidiana. Em [Vasconcelos, N. '04] algumas destas funções têm seu desempenho comparado no contexto de CBIR.

A importância desses estudos reside no fato de que os mesmos ressaltam a necessidade de, ao se escolher uma função de distância, verificar se o contexto e a base de dados estão compatíveis com as condições que foram assumidas durante a formulação da função, evitando-se, assim, o uso equivocado da mesma.

3.7. Comparações Experimentais entre Funções de Distância

Vários trabalhos têm sido desenvolvidos com o objetivo de comparar o desempenho de diferentes funções de distância, quando aplicadas a contextos, métodos comparativos e descritores específicos.

Em [Zhang, D. S. '03] é feita uma avaliação de um conjunto de funções de distância utilizando descritores de forma sobre uma base de imagens selecionadas com o propósito de análise de contorno: a *MPEG-7 contour shape database*. As funções de distância avaliadas foram: L_1 , L_2 , L_∞ , Cosseno, χ^2 , Interseção de Histogramas, Quadrática e Mahalanobis. Considerando a efetividade de recuperação, a Interseção de Histogramas e a Mahalanobis foram as distâncias que geraram piores resultados. Os melhores resultados foram obtidos com as distâncias L_1 e χ^2 .

Em espaços de alta dimensionalidade, os conceitos de proximidade e distância perdem muito de seu significado qualitativo. Em [Aggarwal, C. C. '01] funções de distância são avaliadas em termos de efetividade de recuperação quando usadas em espaços de alta dimensionalidade. Especificamente, é examinado o comportamento da família L_p , chegando-se à conclusão de que a dimensionalidade é sensível ao valor de p , em outras palavras, quanto maior a dimensionalidade, menor o valor de p que produz resultados mais precisos. Baseando-se nisso, os autores investigaram também a norma L_p para valores de p menores do que 1, concluindo que esses valores produzem resultados ainda melhores.

Também considerando altas dimensões, [Qian, G. '04] comparam, em consultas aos vizinhos mais próximos, a distância Euclidiana com a chamada distância do Cosseno Angular, que trata-se simplesmente do valor do cosseno entre os dois vetores. Mostra-se que os resultados obtidos pela distância do Cosseno Angular são próximos aos resultados da distância Euclidiana, em CBIR com alta dimensionalidade.

Em [Spertus, E. '05], seis funções de distância são avaliadas quanto à sua capacidade em representar a similaridade entre grupos de dados. As funções avaliadas foram: L_1 , L_2 , *Pointwise MutualInformation*, *Positive MutualInformation*, Salton (IDF) e Log-Odds. Os grupos usados na avaliação foram os grupos representados pelas diferentes comunidades da rede social Orkut. Os melhores resultados obtidos foram da L_1 , seguida da L_2 .

3.8. Métodos Comparativos

Nesta tese está sendo considerada uma distinção funcional entre a função de distância propriamente dita - cuja formulação matemática é utilizada no cálculo da similaridade - e o que chamamos de **método comparativo**, que diz respeito a:

- o tratamento inicial aplicado aos dados;
- a forma como as assinaturas são obtidas;
- a forma como as assinaturas são manipuladas no processo de efetivação do cálculo da distância;
- o algoritmo utilizado no processo de cálculo da distância.

Exemplificando, no tratamento de textura de uma imagem, pode-se calcular a matriz de co-ocorrência dos pixels e gerar a assinatura da imagem efetuando cálculos a partir de um descritor de contraste aplicado a esta matriz para diferentes distâncias entre pixels. Depois disso, calcula-se a distância usando a métrica L_2 e efetuando uma rotação entre os elementos que compõem a assinatura. Esse conjunto de procedimentos resume um possível método comparativo.

É importante distinguir método comparativo e função de distância, já que muitas vezes um trabalho que afirma estar propondo uma nova função de distância, na verdade está propondo um método comparativo inédito que incorpora uma função de distância já existente. Na verdade, pode-se constatar facilmente que um método comparativo, pela sua natureza exploratória e empírica, em geral é definido para um contexto ou uma base de dados específica, enquanto uma função de distância, por tratar-se de uma formulação matemática, possui caráter mais amplo e geral.

Outro aspecto importante relacionado a esta distinção diz respeito à uniformidade na apresentação de propostas e de resultados. Em [Keogh, E. '02a], os autores enfatizam a necessidade de haver coerência e clareza quando um pesquisador apresenta uma nova proposta, a fim de que a mesma possa ser testada por outros pesquisadores gerando os

mesmos resultados. Frequentemente verifica-se em propostas de novos métodos comparativos que os mesmos são apresentados como se fossem funções de distância e os autores limitam-se a fornecer uma fórmula matemática, omitindo a forma de implementação, que quase sempre contém a estrutura principal da nova metodologia.

3.8.1. Pré-Processamento dos Dados

Salvo casos excepcionais, um pré-processamento deve ser efetuado nos dados, a fim de torná-los mais adequados, fáceis de processar, ou mesmo compatíveis entre si. Dentre as operações possíveis ou necessárias podemos citar:

- Geração de atributos: pode ser que seja mais prático ou mais eficaz a reformulação de um atributo ou a geração de um novo, a partir de outros atributos cujas medidas foram efetuadas;
- Limpeza dos dados: para eliminar ruído ou *outliers*;
- Normalização: tratamento para compatibilizar atributos de magnitudes numéricas distintas;
- Redução de dimensionalidade: quando são utilizados muitos atributos, a viabilidade de medições eficazes é mais bem efetuada através da aplicação de métodos que reduzem a dimensionalidade com a menor perda possível da qualidade e da precisão das informações.

3.8.2. Normalização

Como os atributos normalmente possuem diferentes unidades, escalas e domínios, poderá surgir uma arbitrariedade na influência de cada um deles sobre o valor final da distância. Para eliminar essa arbitrariedade, uma estratégia usual é realizar uma **normalização** dos dados, a qual pode ser feita pela divisão de cada valor pelo intervalo de variação da amostra, ou então pela divisão de cada valor pelo **desvio-padrão** da amostra. O uso do desvio-padrão é mais adequado, devido ao fato de o mesmo ser pouco sensível a valores de *outliers*, ao contrário do que pode ocorrer com o intervalo de variação. Após a normalização, todos os atributos passam a ter desvio-padrão igual a 1.

3.8.3. Redução da Dimensionalidade

Cada imagem pode ser representada, através de seus n atributos, em um espaço n -dimensional, possibilitando, assim, a indexação da base de imagens por meio de estruturas de indexação multidimensionais ou métricas.

Porém, freqüentemente o número de atributos é elevado o suficiente para gerar problemas de desempenho que podem levar à inviabilidade do sistema. Isso ocorre porque: (a) o método comparativo utilizado pode envolver a manipulação de valores de atributos de forma cruzada, gerando um custo quadrático para o cálculo da função de distância; (b) a maioria dos métodos de indexação multidimensional demandam espaço de memória e/ou tempo que crescem exponencialmente.

Existe uma série de técnicas, propostas e testadas, para efetuar redução na dimensionalidade de um conjunto de atributos. Em [Kom, F. '01] algumas destas técnicas são discutidas.

Redução da dimensionalidade usando a dimensão fractal

Em [Traina-Jr, C. '00b] é proposto um método onde a redução da dimensionalidade se faz através da investigação da inter-dependência entre os atributos, que por sua vez é determinada utilizando-se a teoria de fractais. O método consiste em se inserir os atributos em um conjunto de verificação, um de cada vez e, a cada inserção, calcula-se o valor da "dimensão fractal" do conjunto. Quando um atributo inserido é dependente de outro que já faz parte do conjunto, o valor da dimensão fractal não se altera (ou sofre alteração desprezível), e, dessa forma, esse atributo pode ser desconsiderado na assinatura, reduzindo-se assim a dimensionalidade da mesma. Baseando-se nesse raciocínio, em [Souza, E. P. M. '02] é apresentado o algoritmo que implementa a técnica FD-ASE (*Attribute Significance Estimator based on the Fractal Dimension*), que será descrita a seguir.

A existência de atributos correlatos em conjuntos de dados de alta dimensão é muito comum, o que leva ao uso de abordagens para redução da dimensionalidade baseada na eliminação de atributos dependentes. FD-ASE encontra um subconjunto de atributos independentes para um determinado conjunto de dados, além de identificar grupos de atributos correlatos. A técnica é baseada no conceito de dimensão intrínseca D de um conjunto de dados, isto é, a dimensionalidade do objeto, independente da dimensão do espaço onde o mesmo encontra-se inserido.

A idéia fundamental que suporta a descoberta de atributos correlatos consiste em calcular a dimensão intrínseca de seqüências incrementais (S_i) de atributos, definidas por inclusões incrementais de atributos, e usar a diferença entre valores de D de seqüências consecutivas para identificar a existência ou ausência de correlações entre atributos. Em outras palavras, considerando um conjunto de dados $A = \{a_1, a_2, \dots, a_E\}$ composto de E atributos e uma seqüência de atributos $S_i \subset A$, tem-se que $D(S_i)$ denota a dimensão intrínseca do conjunto de dados, levando-se em conta apenas os atributos em S_i . Um atributo $a_k \in (A - S_i)$ está de alguma forma correlacionado a pelo menos um atributo de S_i se a adição de a_k a S_i não causa uma mudança significativa em $D(S_i)$. Os atributos específicos aos quais a_k está correlacionado são descobertos por meio da comparação dos valores de $D(S_i \cup a_k)$ e $D(S_i \cup a_k - a_i)$, $\forall a_i \in S_i$, de forma que uma alta diferença significa que o atributo a_i não está correlacionado a nenhum atributo em S_i . Um valor limite, por exemplo, de 0.1 é usado para determinar a significância de uma alteração no valor de D , ou seja, mudanças abaixo do limite não são consideradas significativas. Como regra, consideram-se limites baixos para identificar correlações fortes, tais como as lineares, enquanto limites mais altos são usados para identificar correlações fracas, tais como as não polinomiais.

Redução da dimensionalidade usando técnicas de mineração de dados

As técnicas de mineração de dados, ao executar tarefas tais como associação, classificação ou agrupamento, implicitamente acabam por determinar conjuntos de atributos que são relevantes para a discriminação entre os objetos manipulados. Esse processo pode ser usado para se descartar atributos pouco influentes, reduzindo, assim, a dimensionalidade do conjunto. A seguir são descritos dois algoritmos que podem ser utilizados com esse objetivo: um algoritmo clássico de classificação - o gerador de árvores de decisão C4.5, e um novo algoritmo gerador de regras de associação baseado em técnicas estatísticas - o StARMiner.

C4.5. Indução de árvores de decisão [Quinlan, J. R. '86] é um processo de aprendizado de máquina no qual uma classificação pode ser obtida a partir de um conjunto de dados, por meio da geração de um conjunto de regras que pode ser visualizado na forma de uma estrutura de árvore [Russel, S. J. '03].

Dentre os diversos métodos existentes para a construção de classificadores, a indução de árvores de decisão apresenta diversas vantagens, tais como esforço computacional reduzido, simplicidade na interpretação de regras e seleção automática de atributos relevantes, quando

comparada, por exemplo, com redes neurais [Cabena, P. '97]. O processo de indução exige a execução de um número de passos de iteração sobre o conjunto de dados de treinamento.

O algoritmo C4.5 permite que se trabalhe com atributos contínuos ou discretos e apresenta melhor desempenho e precisão do que outros algoritmos similares [Kohavi, R. '97].

StARMiner. Mineração de regras de associação é uma tarefa que tem sido exaustivamente explorada nas últimas décadas. O algoritmo StARMiner (*Statistical Association Rule Miner*), apresentado em [Ribeiro, M. X. '05a], estende as técnicas estatísticas de mineração de regras apresentadas em [Aumann, Y. '99] para descobrir padrões em imagens. Uma regra possui o formato $x \rightarrow A_i$ e é identificada apenas se as seguintes condições são satisfeitas:

- O comportamento do atributo A_i em imagens da categoria x dever ser diferente do seu comportamento em imagens de outras categorias;
- O atributo A_i deve apresentar um comportamento uniforme em imagens da categoria x .

Estas condições são implementadas no StARMiner, incorporando restrições do processo de mineração. Seja T uma base de dados de imagens médicas, x uma categoria de imagens, $T_x \subset T$, o subconjunto de imagens pertencentes à categoria x e A_i um atributo. As restrições de interesse implementadas no StARMiner são as seguintes:

$$1) |AvgA_i(T_x) - AvgA_i(T-T_x)| \geq mindif$$

onde: $AvgA_i(Z)$ é a média aritmética de valores de A_i no subconjunto Z de imagens;

$mindif$ é o parâmetro de entrada que indica a mínima diferença permitida entre a média de A_i nas imagens da categoria x e a média de A_i nas imagens restantes da base de dados.

2) Teste de hipótese. A hipótese H_0 deve ser rejeitada com confiança igual ou maior que $minconf$ em favor da hipótese H_1 .

$$\text{com: } H_0: AvgA_i(T_x) = AvgA_i(T-T_x)$$

$$H_1: AvgA_i(T_x) \neq AvgA_i(T-T_x)$$

onde: $minconf$ é o parâmetro de entrada que indica a confiança mínima para rejeitar a hipótese H_0 .

$$3) \sigma A_i(T_x) \leq maxstd$$

onde: $\sigma A_i(T_x)$ é o desvio-padrão dos valores do atributo A_i no subconjunto de imagens T_x ;

$maxstd$ é o parâmetro de entrada que indica o máximo desvio-padrão de valores de A_i permitidos nas imagens da categoria x .

O uso do algoritmo StARMiner para encontrar regras de associação em um conjunto de imagens permite determinar um conjunto de atributos com alto poder de discriminação, desde que os atributos possuam um padrão de comportamento específico e uniforme nas imagens de uma dada categoria. Os atributos que apresentam comportamento uniforme para todas as imagens da base de dados, independentemente da categoria, não contribuem para a discriminação de categorias e devem ser descartados.

3.8.4. Exemplos de Métodos Comparativos

Nesta Seção serão apresentadas algumas pesquisas que propuseram novos métodos comparativos para a recuperação de imagens por similaridade.

Histograma Métrico

Em [Traina, A. J. M. '03a] é apresentado um novo método comparativo chamado Histograma Métrico, que consiste de modificações sobre os histogramas tradicionais, acompanhado de uma nova forma de cálculo da distância. Os histogramas métricos permitem comparar imagens de tamanhos variados e mapeadas em diferentes faixas de quantização, ou seja, além de serem – como os histogramas normalizados - invariantes a transformações geométricas nas imagens, os histogramas métricos são também invariantes com relação a transformações lineares de brilho.

Para permitir consultas por similaridade a partir de estruturas de índices baseadas em Métodos de Acesso Métricos (MAM), foi necessário desenvolver uma nova função de distância que considera a diferença de área entre pares de histogramas métricos no cálculo da distância.

A aquisição do histograma métrico, a partir de um histograma convencional, é realizada da seguinte forma: o contorno original do histograma convencional é reduzido a uma seqüência de segmentos de reta. O algoritmo utilizado identifica pontos de máximo, pontos de mínimo e pontos intermediários significativos, com base em valores de gradiente. Esses pontos são chamados de pontos de controle.

Cada região definida por um segmento, entre dois pontos de controle, é chamada de *bucket*. Um *bucket* corresponderia, assim a um *bin* de um histograma normalizado, possuindo uma largura e valores de frequência para os seus pontos inicial e final. A diferença é que a divisão de *bins* é pré-determinada e igual para todas as imagens, enquanto a quantidade e a

divisão dos *buckets* depende de cada imagem. Um *bucket* normalmente engloba vários *bins* e, dessa forma, consegue-se manter o formato do histograma dentro de limites confiáveis, e ao mesmo tempo reduzir drasticamente sua dimensionalidade. A Figura 3.7 mostra um histograma métrico.

Uma consequência importante desse processo é que não existe uma dimensão pré-definida para o domínio do problema, já que cada imagem pode ter um número diferente de *buckets*. Assim, as imagens não podem ser representadas como pontos em um espaço dimensional e isso conduz à necessidade de se definir também uma nova forma de cálculo da distância entre duas imagens, uma vez que tanto o número de *buckets*, quanto a sua distribuição em diferentes histogramas são variáveis.

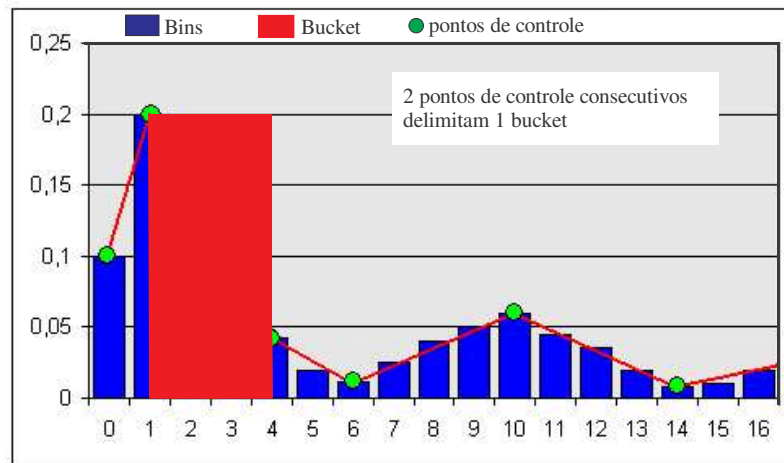


Figura 3.7. Exemplo de histograma métrico [Bueno, J. M. '02b]

A distância proposta – **Distância Métrica** - para comparar histogramas métricos efetua a comparação entre as áreas dos mesmos. Partindo do princípio de que a similaridade entre dois histogramas pode ser definida pelo grau de sobreposição das áreas de suas distribuições de frequência, a distância entre duas imagens é determinada pela diferença entre as suas áreas:

$$d_M(Q, C) = \int_0^{bm} |M_H(C, x) - M_H(Q, x)| dx \quad (3.13)$$

onde: bm é a largura máxima entre os dois histogramas;

$M_H(\text{imagem}, x)$ é a função contínua que representa o histograma métrico da imagem x

Experimentos realizados com bases de dados de imagens de tomografias mostraram que, para a maioria das imagens, o número de *buckets* ficou na faixa de 10-62, contrastando com os

256 *bins* que normalmente são utilizados em histogramas normalizados. Em consultas por similaridade efetuadas, foram obtidos resultados equivalentes na recuperação de conjuntos de imagens usando histogramas métricos e histogramas convencionais associados à distância L_1 . Porém, o histograma métrico apresenta ganhos significativos de desempenho no processo de indexação e recuperação de imagens, chegando a ser 4 vezes mais rápido do que o histograma convencional e mantendo o mesmo poder de recuperação de informação.

Warping Dinâmico

Programação Dinâmica [Gusfield, D. '99] é uma técnica para comparação entre vetores, que consiste na avaliação de todas as possíveis combinações de resultados de distâncias entre pares de atributos adjacentes correspondentes, geralmente buscando-se minimizar um parâmetro que represente algum tipo de "custo". Dentre as várias aplicações desta técnica, destacam-se as análises comparativas entre seqüências de caracteres, seqüências de DNA e séries temporais.

Dentre as técnicas mais difundidas usadas para se comparar duas séries temporais, destaca-se a chamada *Dynamic Time Warping* [Keogh, E. '01a] [Faloutsos, C. '05], ou simplesmente *Dynamic Warping*, que consiste na aplicação da Programação Dinâmica com o objetivo de avaliar a aproximação de séries, minimizando a distância final entre séries que possuem desenvolvimentos semelhantes, porém defasados ou "deformados" um em relação ao outro, ao longo do eixo temporal.

A sensibilidade à deformação apresentada por esta técnica deve-se ao fato de que as comparações entre os pontos não são feitas rigidamente entre os pares localizados nas mesmas coordenadas, mas com a flexibilidade de que um ponto de uma série possa ser comparado com pontos adjacentes da outra série, conforme mostrado na Figura 3.8.

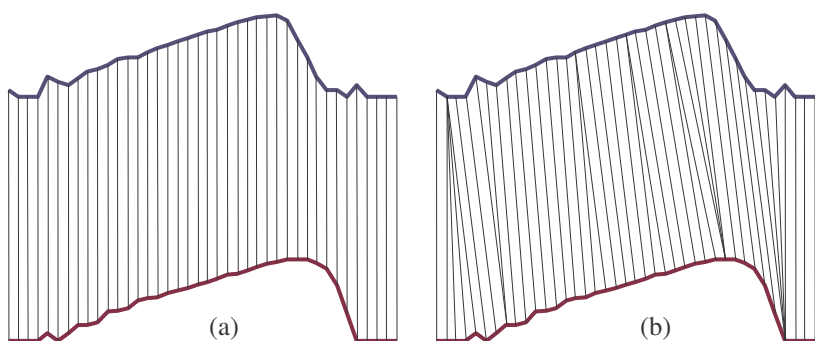


Figura 3.8. Comparação entre séries: a) convencional; b) com *Dynamic Time Warping*

Suponhamos que temos duas séries $Q=(q_1, q_2, \dots, q_n)$ e $C=(c_1, c_2, \dots, c_m)$, de tamanhos n e m respectivamente. Para compará-las usando *Dynamic Warping*, construímos uma matriz $n \times m$ cujo elemento (i,j) contém a distância $d(q_i, c_j)$. Tipicamente é usada a distância Euclidiana. Cada elemento da matriz corresponde ao alinhamento entre os pontos nele representados.

Um percurso de ajuste $W=(w_1, w_2, \dots, w_k)$ é um conjunto de elementos contíguos da matriz que define um mapeamento entre Q e C . O percurso de ajuste deverá obedecer às seguintes regras [Keogh, E. '02b]:

- Iniciar e terminar em células diagonalmente opostas da matriz, ou seja, $w_1=(1,1)$ e $w_k=(n,m)$;
- A seqüência entre os elementos do percurso deverá conter elementos adjacentes da matriz (incluindo elementos diagonalmente posicionados);
- A seqüência não poderá "voltar" no caminho dos elementos da matriz.

Há muitos possíveis percursos de ajuste, porém o percurso a ser escolhido é aquele que minimiza o custo da deformação, ou seja, cuja distância acumulada ao longo do mesmo é mínima. Esta distância – *Warp Distance* (WD) – é definida na equação (3.14):

$$WD(Q, C) = \min\left(\sum_{k=1}^K w_k / K\right) \quad (3.14)$$

onde:

w_k é o $k^{\text{ésimo}}$ elemento do percurso de ajuste

K é o número de elementos do percurso de ajuste

A WD tem sido usada com sucesso em aplicações tais como: *data mining*, reconhecimento de gestos, robótica e processamento da fala [Keogh, E. '01b].

Detecção de Alterações em Sinais de Eletrocardiogramas

Em [Figueiredo, J. C. B. '01a], os autores propõem um método estatístico para detectar alterações anormais em sinais de eletrocardiogramas (ECG). O método consiste em se gerar reconstruções de porções do sinal no espaço de fase e comparar a representação de um sinal normal com janelas temporais obtidas ao longo do sinal, comparando essas representações através da distância de Kullback-Leibler (ver Seção 3.3.2). Cada representação no espaço de fase é considerada como sendo uma distribuição estatística. Quando são comparadas as

representações de trechos do sinal com o trecho normal de referência, verifica-se que, para porções normais do sinal, o valor da distância permanece praticamente constante e abaixo de um valor de corte. Em casos em que ocorrem transientes anormais, em partes anormais do ECG, a distância aumenta de forma consistente, permitindo, assim, a detecção de alterações anormais nos exames.

Earth Mover's Distance

Em [Rubner, Y. '01] é apresentado um novo método comparativo, a *Earth Mover's Distance* (EMD), proposto para ser utilizado para buscas por similaridade baseadas em cor e em textura. O método proposto baseia-se no antigo e conhecido **problema dos transportes**, bastante estudado na área de pesquisa operacional e programação linear. Parte-se do princípio de que temos um conjunto de produtores, com suas respectivas quantidades de produto e um conjunto de consumidores, com suas respectivas demandas. Dada a localização de cada produtor e cada consumidor, o problema dos transportes irá gerar a configuração de transporte do produto que atende a todos os envolvidos, com os custos de transporte reduzidos ao mínimo.

Para cor foi utilizado o padrão *Lab*, que utiliza três parâmetros: L representando a luminância, a e b representando a cromaticidade (ver Seção A.2.1). Os valores de L , a e b possuem correspondência com padrões mais conhecidos, como o RGB.

Para textura, as imagens foram pré-processadas usando filtros de Gabor (ver Seção A.4.3), os quais têm apresentado os melhores resultados, comparados com outras representações de textura. A função de distância utilizada foi a L_2 .

Em testes nos quais a EMD foi comparada com aplicações diretas de funções de distância aos valores das assinaturas, a EMD sempre gerou melhores resultados. A EMD será métrica sempre que a função de distância utilizada como base para os cálculos o for. Porém seu cálculo é relativamente pesado. Para melhorar o desempenho, [Rubner, Y. '01] propõem uma função mais barata, que serve como limite inferior para a mesma.

Extratores Baseados em *Wavelets*

Em [Castanón, C. A. B. '03] é apresentado um método comparativo de imagens médicas que utiliza a transformada de *Wavelets* para a geração das assinaturas. Foi desenvolvida uma aplicação que executa a extração de atributos e permite a consulta por similaridade a partir de uma imagem de consulta, comparando-a com outras imagens de uma base de dados. O processo de extração de características baseia-se na aplicação de transformadas *Wavelets*

(Daubechies⁴) sobre os valores de luminância dos pixels da imagem. As assinaturas foram geradas a partir da aplicação de três níveis de decomposição *Wavelet*, gerando dez sub-bandas, sendo um de baixa frequência e nove de alta frequência. A Figura 3.9, extraída de [Castanón, C. A. B. '03], mostra uma imagem original juntamente com seus dois primeiros níveis de decomposição. Para cada sub-espaco foram calculados valores de média, energia e entropia, gerando, assim um total de trinta atributos em cada assinatura.

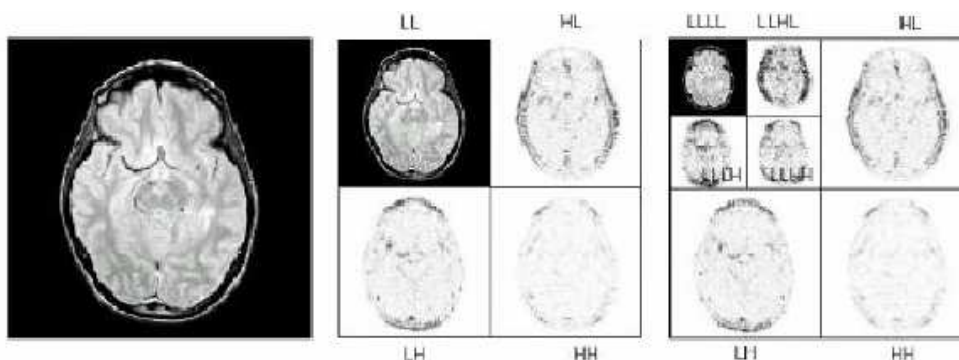


Figura 3.9. Imagem original e seus dois níveis de decomposição *Wavelet*

A função de distância utilizada foi a Euclidiana. Experimentos realizados com imagens médicas formando grupos de imagens semelhantes geraram resultados muito promissores com respeito a consultas por similaridade.

Em outro experimento, a recuperação de imagens é realizada a partir da análise de textura das mesmas, obtida através do seu tratamento por atributos das *Wavelets* de Gabor. Dada uma imagem, é aplicada a transformada de Gabor e, assumindo que as regiões locais de textura são homogêneas no espaço, a média μ_{mn} e o desvio-padrão Φ_{mn} da magnitude dos coeficientes da transformada são usados para representar as regiões a serem comparadas. A assinatura é então construída usando-se μ_{mn} e Φ_{mn} como sendo os atributos básicos.

Nos experimentos foram usados como parâmetros quatro escalas ($S=4$) e seis orientações ($K=6$), resultando em uma assinatura com 24 pares de atributos:

$$f = \{ \mu_{00}, \Phi_{00}, \mu_{01}, \Phi_{01}, \dots, \mu_{23}, \Phi_{23} \}$$

Segundo [Manjunath, B. S. '96] a distância entre duas imagens no espaço de atributos pode ser calculada como:

$$d(Q,C) = \sum_m \sum_n d_{mn}(Q,C) \quad (3.15)$$

$$e \quad d_{mn}(Q,C) = \left| \frac{\mu_{mn}(C) - \mu_{mn}(Q)}{\alpha(\mu_{mn})} \right| + \left| \frac{\sigma_{mn}(C) - \sigma_{mn}(Q)}{\alpha(\sigma_{mn})} \right| \quad (3.16)$$

onde: $\forall(\cdot)_{mn}$ e $\forall(\Phi)_{mn}$ são os desvios-padrão dos respectivos atributos sobre a totalidade da base de dados, usados para normalizar os atributos individuais.

CBIR Usando Atributos Baseados em Fractais

Em [Balan, A. G. R. '05] as imagens são segmentadas em regiões que representam classes de textura da mesma, dependendo de propriedades estatísticas de cada pixel com sua vizinhança. O algoritmo utilizado para isto é baseado nos campos estatísticos aleatórios de Markov (*Markov Random Field – MRF*). A Figura 3.10 apresenta um exemplo de imagem segmentada em 4 classes.

Uma vez determinadas as regiões da imagem, para cada uma delas é calculada a respectiva dimensão fractal, a qual deverá funcionar como um descritor de textura. Juntamente com outros atributos, é então composto o vetor de características da imagem. O uso da análise fractal permite compactar a dimensão do vetor de características, ao representar com apenas um valor numérico uma série de atributos.

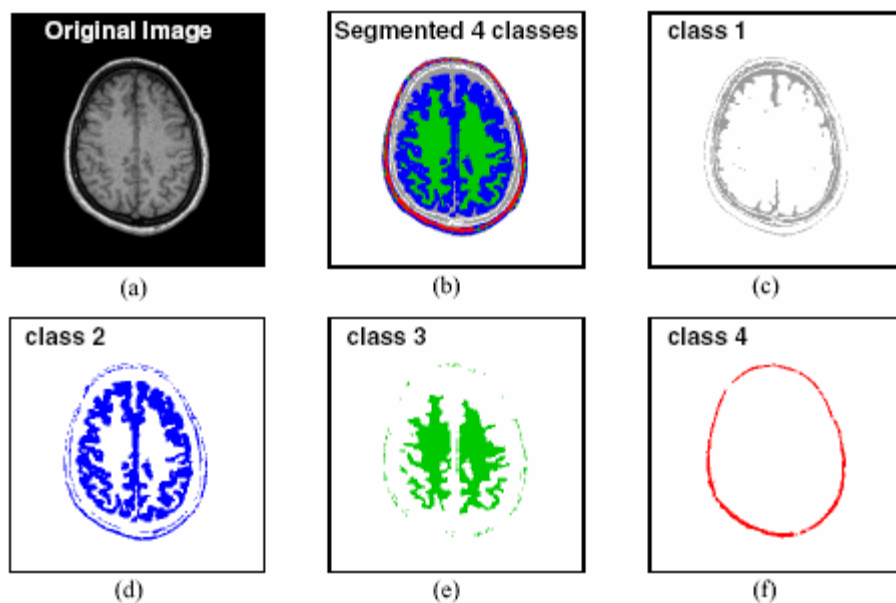


Figura 3.10. Imagem original (a), segmentada em 4 classes (b) e cada uma das classes separadas (c - f) [Balan, A. G. R. '05]

Os atributos utilizados para caracterizar a imagem, extraídos de cada uma das regiões, foram: a dimensão fractal, as coordenadas x e y do centro de massa, a massa, a média de níveis de cinza e o coeficiente linear de ajuste da curva fractal. Com esses 6 atributos, uma imagem de 5 regiões terá 30 atributos.

Algumas funções de distância foram testadas para a comparação entre vetores, sendo que a que gerou melhores resultados foi a distância L_1 ponderada pela variância do conjunto.

Experimentos realizados em bases de imagens médicas mostraram a eficácia do método para a discriminação de imagens oriundas de diferentes regiões do corpo humano.

3.9. Considerações Finais

Neste Capítulo, foi apresentado um levantamento de funções de distância e métodos comparativos utilizados para comparação entre imagens. Aplicações específicas e aspectos de implementação foram discutidos.

Capítulo 4

4. MÉTODO PARA COMPARAÇÃO OTIMIZADA ENTRE HISTOGRAMAS

Neste Capítulo apresentaremos um novo método para comparação entre histogramas, com o qual se busca a redução da imprecisão inerente ao uso dos histogramas como vetores de características associados ao descritor "cor" (tons de cinza) de imagens, com o uso de procedimentos onde características locais e globais são consideradas, no sentido de se atingir uma melhor aproximação de resultados perceptivos. Resultados experimentais com imagens médicas são apresentados e discutidos.

4.1. Introdução

Apesar de apresentar limitações conhecidas – que serão discutidas nas Seções subseqüentes - em relação à sua capacidade de representação, histogramas têm sido largamente utilizados nos mais diversos campos de atuação da Computação e das Engenharias. Isto pode ser constatado numa simples análise de trabalhos científicos publicados recentemente abordando histogramas como tema. Somente na biblioteca digital DBLP (*Database Systems and Logic Programming*) [DBLP '05] constam, entre os anos de 2002 e 2005, 163 publicações envolvendo histogramas, em áreas tais como: representação de imagens, recuperação da informação, modelagem visual, detecção de objetos em vídeo, reconhecimento de faces e impressões digitais, estimativa de seletividade, otimização de consultas, análise fractal, radiologia digital e roteamento em sistemas distribuídos. Logo, o estudo de métodos comparativos que possam minimizar efeitos negativos e gerar maior precisão nos resultados de comparações mostra-se bastante pertinente e útil como abordagem geral de manipulação e análise de situações envolvendo histogramas.

4.2. Definição do Problema

Histogramas, quando utilizados para se realizar comparações e buscas por similaridade, apresentam uma importante limitação, que consiste na geração de uma quantidade freqüentemente significativa de resultados falsos positivos, devido ao fato de objetos pouco similares poderem apresentar histogramas muito semelhantes, ou até mesmo iguais. Além

disso, os métodos tradicionais de comparação entre histogramas tendem a aumentar esse volume de falsos positivos, devido à sua rigidez de tratamento dos diferentes *bins* dos histogramas, e também por não levarem em consideração características globais dos mesmos.

Em CBIR, o uso de histogramas como passo inicial na seleção de imagens, justifica-se ainda mais devido ao fato de algoritmos que lidam com outras características intrínsecas – textura e forma – apresentarem freqüentemente um custo computacional alto, além de serem dependentes do domínio da aplicação. Histogramas, por outro lado, são obtidos com custo linear em relação ao número de pixels da imagem. Portanto, um procedimento eficiente para recuperação de imagens seria o uso de histogramas para filtrar e reduzir o conjunto de imagens candidatas, como um passo preliminar do processo, seguindo-se de métodos mais precisos e custosos [Alferez, R. '05].

Métricas tradicionalmente usadas para comparar histogramas apresentam duas restrições principais. A primeira deve-se à rigidez da comparação *bin a bin*, a qual leva ao descarte de imagens que, apesar de serem similares, apresentem diferenças de brilho e contraste, diferenças estas que se refletem diretamente como "distorções" no traçado gráfico dos histogramas. A Figura 4.1 apresenta duas imagens e seus respectivos histogramas. A imagem (b) é resultado de alterações de brilho e contraste sobre a imagem (a). Pode-se ver que seus histogramas são parecidos, mas o histograma da imagem (b) apresenta defasagens e distorções em relação ao histograma da imagem (a). Por causa disso, uma comparação *bin a bin* resultará em um valor de dissimilaridade relativamente alto. Esse tipo de problema é bastante comum em bases de imagens gerais e heterogêneas, levando à geração de resultados falsos negativos em consultas por similaridade.

A segunda restrição está ligada a uma característica mais semântica: imagens de mesma categoria costumam apresentar desenvolvimentos semelhantes dos respectivos histogramas e isto deveria, de alguma forma, ser considerado pelo método comparativo. A Figura 4.2 apresenta exemplos de histogramas típicos de imagens oriundas de regiões específicas do corpo humano. De acordo com a técnica empregada e a categoria de imagens, seus histogramas adquirem especificidades quanto a alguns parâmetros, tais como quantidade de picos, "achatamento" (curtose) do gráfico e "balanceamento" do gráfico em relação à mediana (simetria). Por exemplo, um histograma típico de cérebro com dilatação ventricular apresenta um pico extra na região central da curva, quando comparado com um histograma típico de cérebro normal, como pode ser visto na Figura 4.2. Se estas características forem consideradas pela função de distância, ocorrências de resultados falsos positivos poderão ser minimizadas, melhorando, dessa forma, a precisão das comparações.

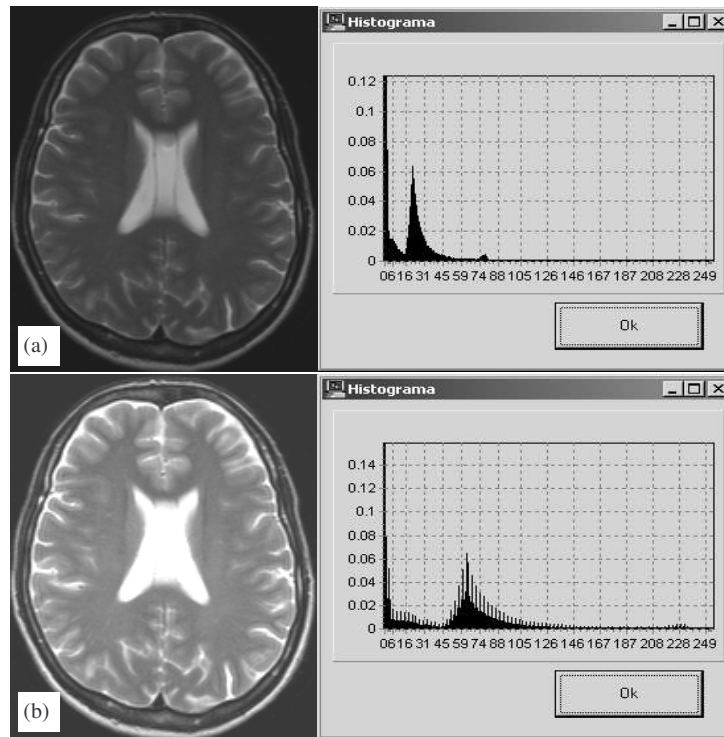


Figura 4.1. (a) Imagem original e (b) com alterações de brilho e contraste e respectivos histogramas

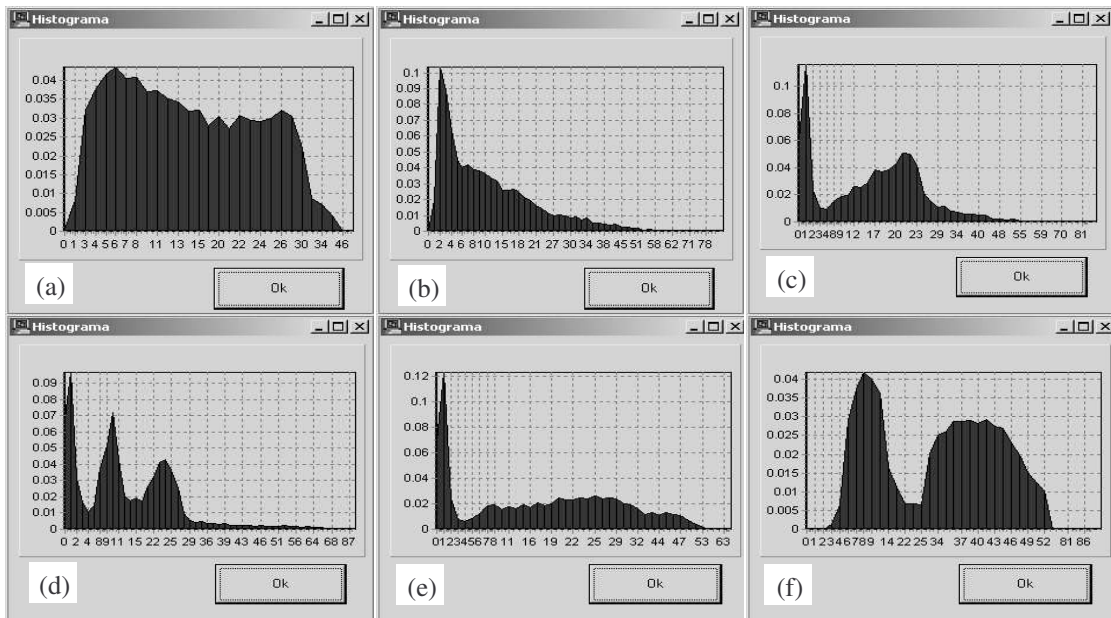


Figura 4.2. Histogramas típicos de regiões específicas do corpo humano: a) abdômen; b) pescoço; c) cérebro; d) cérebro com dilatação ventricular; e) tecido ósseo; f) pélvis

A primeira restrição leva a resultados falsos negativos, enquanto a segunda propicia resultados falsos positivos. Nosso objetivo é obter um novo método de comparação entre histogramas que leve à redução desses efeitos indesejáveis. O método aqui apresentado consiste da definição de uma medida de comparação local flexível, combinada com um conjunto de características globais dos histogramas.

4.3. O Método Desenvolvido

O desafio crucial deste método consiste em determinar comparações efetivas de histogramas objetivando minimizar tanto os resultados falsos positivos quanto os falsos negativos, no processo de comparação entre duas imagens. Para resolver esse problema, este método aplica comparações do tipo *warp* (definidas na Seção 3.8.4) aos histogramas métricos (também definidos na Seção 3.8.4), ao mesmo tempo em que incorpora características globais dos histogramas, a fim de compor uma nova métrica: a *Global Warp Distance*.

4.3.1. Características Globais de Histogramas

Analisamos o desenvolvimento gráfico apresentado por diferentes histogramas típicos (como aqueles mostrados na Figura 4.2), com o objetivo de identificar algumas características globais significativas que pudessem ser usadas pela função de distância na comparação das respectivas imagens. As características relevantes escolhidas foram: **simetria**, **curtose** e **número de picos**. Elas podem ser definidas a partir de medidas estatísticas sobre as distribuições de frequências dos histogramas. As medidas usadas neste trabalho são: média, mediana, quartil e decil. A média é definida pela equação (4.1). A mediana é o valor de tom de cinza x que divide o histograma em duas partes que possuem a mesma área. Quartis (Q) são os valores de tom de cinza x que dividem o histograma em quatro partes iguais, enquanto os decis (D) são os valores de tom de cinza que dividem o histograma em dez partes iguais [Spiegel, M. R. '76]. A Figura 4.3 exemplifica esses conceitos.

$$m\acute{e}dia = \frac{\sum_{x=1}^n f(x)x}{\sum_{x=1}^n f(x)} \quad (4.1)$$

onde:

x é um valor de tom de cinza

n é a quantidade de valores de tom de cinza

$f(x)$ é a frequência do nível de cinza x

Simetria é um parâmetro que quantifica o grau de balanceamento do gráfico em relação à sua mediana. O valor de assimetria pode ser obtido pela diferença entre os valores da média e da mediana. Curvas simétricas possuem média igual à mediana. Curvas de assimetria positiva possuem a média maior do que a mediana e curvas de assimetria negativa, o oposto.

$$\text{assimetria} = \text{média} - \text{mediana} \quad (4.2)$$

Curtose é um parâmetro que mede o achatamento da curva. Neste trabalho, o seu valor é definido como sendo a proporção dada pela diferença entre Q3 e Q1 com a diferença entre D9 e D1, como mostra a equação (4.3) [Spiegel, M. R. '76]. Curvas achatadas apresentam valores de curtose maiores do que curvas delgadas.

$$\text{curtose} = (Q3 - Q1)/(D9 - D1) \quad (4.3)$$

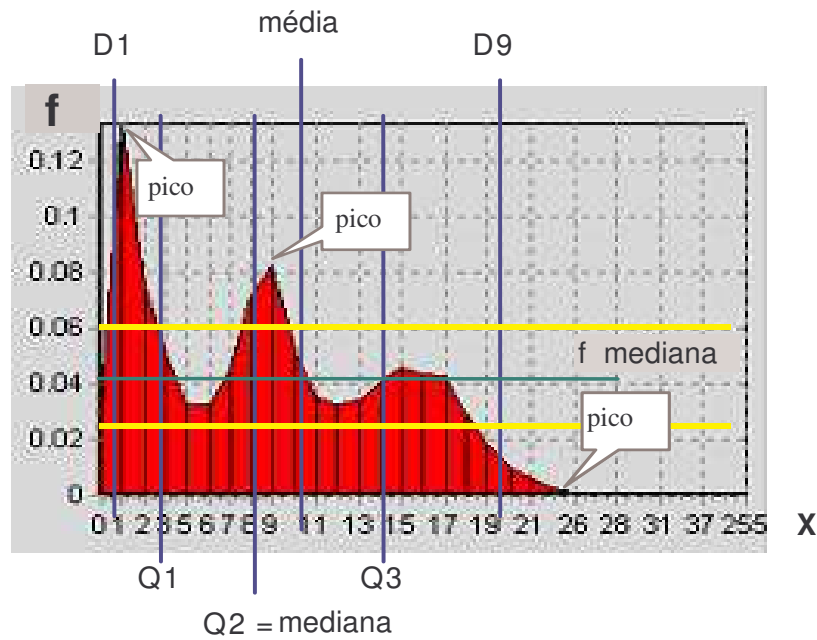


Figura 4.3. Características estatísticas e globais dos histogramas

O **número de picos** positivos e negativos, neste trabalho, é determinado percorrendo a curva e identificando os pontos que se encontram de fora (acima ou abaixo) de uma faixa de valores centrada no valor de frequência mediana do histograma. Na Figura 4.3, esta faixa é representada pelas linhas amarelas, as quais são definidas como uma porcentagem acima e abaixo do valor da mediana de f . Pontos que estejam localizados fora da faixa e que representem inversões de inclinação da curva são marcados como candidatos. O valor de

porcentagem é estabelecido como 20% do máximo valor de frequência do histograma. Esse valor foi definido a partir da observação de diversos histogramas típicos. Além disso, para ser computado como pico, o ponto candidato deverá apresentar uma diferença de valor de frequência com relação ao pico anterior maior do que 20% do máximo valor de frequência do histograma (limiar definido empiricamente).

Em suma, para ser considerado um pico um ponto tem que satisfazer simultaneamente as seguintes condições:

- Estar situado externamente à faixa [-20%, +20%] da mediana f ;
- Representar inversão na inclinação da curva ou ser o primeiro ou o último ponto da curva;
- Apresentar uma diferença de frequência relativa ao pico anterior maior do que 20% do máximo valor de frequência.

4.3.2. Global Warp Distance

Como dito anteriormente, nosso objetivo é definir um método comparativo entre histogramas que produza resultados mais precisos. Variações não lineares de brilho e/ou contraste apresentadas por uma imagem produzem deslocamentos locais dos *bins* do histograma, de forma que uma função de distância *bin a bin* produzirá um valor de dissimilaridade que pode descartar uma imagem candidata bastante similar à imagem de referência, gerando um falso negativo. Por outro lado, a despeito de diferentes categorias de imagens apresentarem características globais específicas nos seus histogramas, funções de distância *bin a bin* podem gerar valores de similaridade que consideram como próximos, histogramas bastante distintos em suas características globais (com alta probabilidade de pertencerem a categorias distintas), gerando, assim, falsos positivos.

O método desenvolvido neste trabalho lida de forma flexível com discrepâncias locais, ao mesmo tempo em que considera características globais dos histogramas a fim de definir uma métrica – a *Global Warp Distance* (GWD) – para comparar duas imagens, de acordo com o seguinte procedimento:

1. Extrair o histograma métrico de ambas as imagens, gerando os seus *buckets* e pontos de controle. A seqüência de valores de frequência dos pontos de controle de cada imagem forma o seu vetor de características.
2. Considerando os vetores de características de cada imagem, aplicar a distância *warp* entre os mesmos. A Figura 4.4 exemplifica esse processo aplicado a dois histogramas métricos hipotéticos, através da matriz de ajuste dos mesmos (ver

Seção 3.8.4), com o percurso de ajuste em destaque. Neste trabalho, chamaremos a *Warp Distance* (equação (3.14)) aplicada a histogramas métricos de *Metric Warp Distance* (MWD).

3. Calcular características globais – assimetria, curtose e número de picos – de cada histograma.

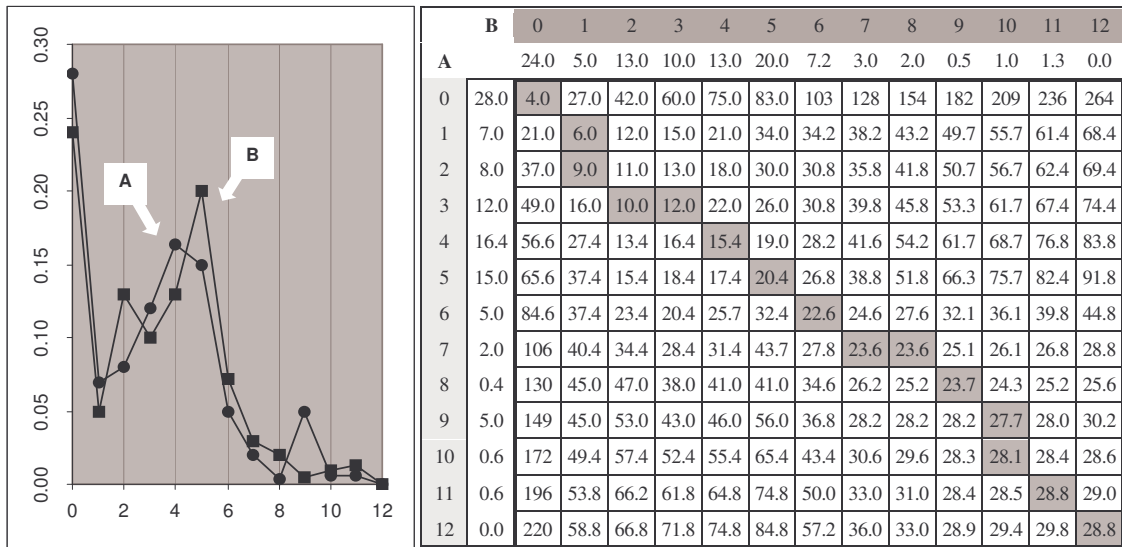


Figura 4.4. Dois histogramas e suas respectivas matrizes de ajuste, com percurso de ajuste em destaque

4. Calcular um conjunto de coeficientes, um para cada característica global. Para duas imagens Q e C:

- a. Coeficiente de simetria:

$$c_s = 1 + |simetria(Q) - simetria(C)| \quad (4.4)$$

- b. Coeficiente de curtose:

$$c_c = 1 + |curtose(Q) - curtose(C)| \quad (4.5)$$

- c. Coeficiente de picos:

$$c_p = 1 + |pi\ cos(Q) - pi\ cos(C)| * 0.1 \quad (4.6)$$

O valor 0.1 usado no coeficiente de pico foi escolhido experimentalmente, de forma a manter a influência deste coeficiente similar à dos outros.

5. Calcular a *Global Warp Distance* (GWD) por meio da ponderação da MWD gerada no passo 2 com os coeficientes gerados no passo 4:

$$GWD(Q,C) = c_s * c_c * c_p * MWD(Q,C) \quad (4.7)$$

4.4. Resultados Experimentais

Para avaliar o método proposto, foi utilizada uma ferramenta pré-existente (*Dicom Viewer*), desenvolvida pelo GBDI, a qual permite a visualização de imagens nos formatos DICOM ou *bitmap* e executa consultas aos vizinhos mais próximos. Esta ferramenta foi estendida com a implementação de um novo método para a comparação de histogramas métricos utilizando a GWD.

Para verificar a precisão da GWD, três tipos de experimentos foram realizados sobre uma base de dados contendo 4200 imagens de diversas regiões do corpo humano, obtidas a partir de exames de Tomografia Computadorizada de Ressonância Magnética. As imagens foram adquiridas em 256 níveis de cinza, apresentando resolução espacial variável (de 256x256 até 1024x1024 pixels). Uma vez que os histogramas de intensidade que dão origem aos histogramas métricos são normalizados pelo número de pixels da imagem, a resolução espacial não afeta os resultados do método. As imagens que compõem essa base são oriundas do HCRP e foram obtidas junto ao CCIFM da FMRP-USP.

4.4.1. Experimento 1

Um subconjunto composto por 1100 imagens selecionadas da base foi montado formando grupos de imagens de mesmas regiões do corpo humano, da seguinte forma:

- 49 imagens de cortes axiais de pélvis;
- 24 imagens de cortes sagitais de pélvis;
- 46 imagens de cortes axiais de cérebro;
- 59 imagens de cortes sagitais e coronais de cérebro;
- 50 imagens de cortes sagitais e coronais de cérebro apresentando uma lesão (dilatação ventricular);
- 84 imagens de cortes sagitais de espinha;
- 23 imagens de cortes axiais da região do fígado;
- 112 imagens de cortes axiais de abdômen inferior;
- 45 imagens de cortes axiais de perna;
- 608 imagens de outras regiões do corpo, excluindo as regiões definidas acima.

Imagens dos diferentes grupos descritos acima foram selecionadas como referência para a execução de consultas aos k vizinhos mais próximos, com k variando de 1 até 40, usando 3 distâncias diferentes: a Distância de Histogramas convencional (HD), a Distância Métrica (MD)

(equação (3.13)) e a *Global Warp Distance* (GWD) (equação (4.7)). Nesse caso, a Distância de Histogramas convencional é calculada pela norma L_1 aplicada aos pares de *bins* dos histogramas:

$$HD(Q, C) = \sum_{x=1}^n |f_Q(x) - f_C(x)| \quad (4.8)$$

onde:

n é a quantidade de níveis de cinza

$f_{imagem}(x)$ é a frequência do nível de cinza x em *imagem*

A fim de analisar os resultados e avaliar a eficácia das funções de distância, os conhecidos conceitos de "precisão" e "revocação" (*precision e recall*) foram aplicados. Para um conjunto resultante de uma consulta a uma base de imagens, precisão e revocação são definidos como:

$$precisão = \frac{TRO}{TO}, \quad revocação = \frac{TRO}{TR} \quad (4.9)$$

onde:

TR = total de imagens relevantes disponíveis na base de dados

TRO = total de imagens relevantes obtidas no resultado

TO = total de imagens obtidas no resultado

A construção de um gráfico apresentando diferentes valores de revocação versus precisão produz uma curva onde, como regra de análise, o resultado é melhor à medida que a curva se aproxima do topo, ou seja de valores de precisão iguais a 1.

Resultados

Para avaliar numericamente os resultados, curvas de revocação versus precisão foram construídas para cada função de distância (HD, MD e GWD), considerando-se os valores médios encontrados nas consultas executadas. Os gráficos resultantes são mostrados na Figura 4.5.

Analisando as curvas, podemos ver que o comportamento de MD é praticamente equivalente ao de HD, o que corrobora os resultados apresentados em [Traina, A. J. M. '03a], comprovando que a MD, a despeito de ser mais eficiente em termos de tempo, apresenta a mesma habilidade de recuperação da HD.

Por outro lado, a GWD apresenta uma capacidade de recuperação muito maior do que a HD e a MD, gerando valores mais elevados de precisão ao longo de todo o espectro de

valores de revocação. Além disso, o comportamento apresentado por cada curva revela que a precisão da GWD é alta para um largo intervalo de revocação, enquanto as outras curvas perdem precisão rapidamente, à medida que os valores de revocação crescem: GWD mantém os valores de precisão acima de 80% para valores de revocação até 70%, enquanto com as outras distâncias 80% de precisão é mantida apenas até aproximadamente 30% de revocação. Isto significa que é possível trabalhar com valores mais elevados de k em consultas aos k vizinhos mais próximos, conseguindo manter os respectivos valores de precisão em patamares aceitáveis.

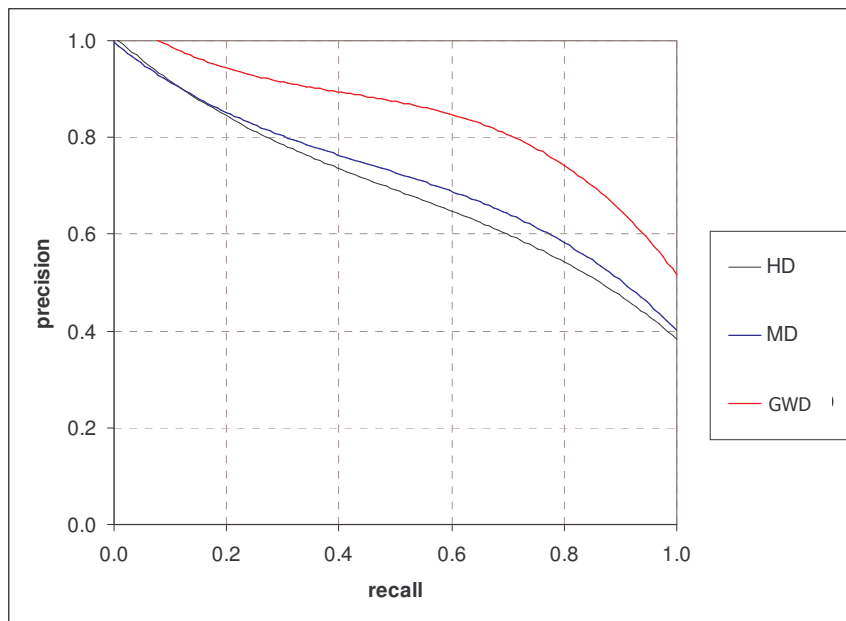


Figura 4.5. Curvas de revocação versus precisão de consultas aos k vizinhos mais próximos com HD, MD e GWD

4.4.2. Experimento 2

Usando a totalidade da base de imagens (4200 imagens), tomamos imagens de diferentes regiões do corpo humano (abdômen, coluna, pélvis e cérebro), escolhidas de forma aleatória, para servirem de referência a consultas aos 20 vizinhos mais próximos. Estas consultas foram repetidas usando as diferentes distâncias a fim de comparar o poder de recuperação das mesmas.

A Figura 4.6 apresenta os resultados obtidos tomando como referência uma imagem de pélvis e aplicando a MD e a MWD. A imagem de referência é mostrada no canto superior esquerdo e a seqüência dos resultados é da esquerda para a direita e de cima para baixo. Visualmente é possível observar que o conjunto resultante da MWD, se comparado ao

conjunto resultante da MD, contém uma maior quantidade de imagens com alto grau de similaridade em relação à imagem de referência. As imagens bastante similares à imagem de referência que aparecem no conjunto resultante da MWD e não aparecem no conjunto resultante da MD são imagens que apresentavam diferenças de brilho e contraste (bem mais escuras do que as outras imagens do grupo de imagens de pélvis). Essas diferenças não estão visíveis na figura porque as imagens sofreram um auto-nivelamento a fim de viabilizar a sua visualização neste documento. Este resultado comprova a maior flexibilidade da MWD em manipular defasagens e deformações dos histogramas, ao recuperar imagens que foram indevidamente "negativadas" pela MD.



Figura 4.6. Resultados de consulta aos 20 vizinhos mais próximos para: (a) MD e (b) MWD

A Figura 4.7 mostra os resultados obtidos tomando como referência uma imagem patológica. Trata-se de um cérebro com dilatação ventricular. Foram obtidos resultados para as distâncias MD, MWD e GWD. Analisando a Figura 4.7, podemos verificar que:

- 1) No conjunto (b) todas as imagens recuperadas são de cérebro, contrastando com o conjunto (a), onde 9 das 20 imagens são muito dissimilares à imagem de referência. Isto mostra que de (a) para (b) – ou seja, substituindo a distância MD pela MWD – o número de resultados falsos negativos reduziu significativamente;
- 2) Em ambos os conjuntos resultantes (b) e (c) todas as imagens são de cérebro, porém em (c) 18 imagens apresentam dilatação ventricular, enquanto em (b) apenas 13. Isto mostra que substituindo a distância MWD pela GWD o número de resultados falsos positivos foi reduzido.

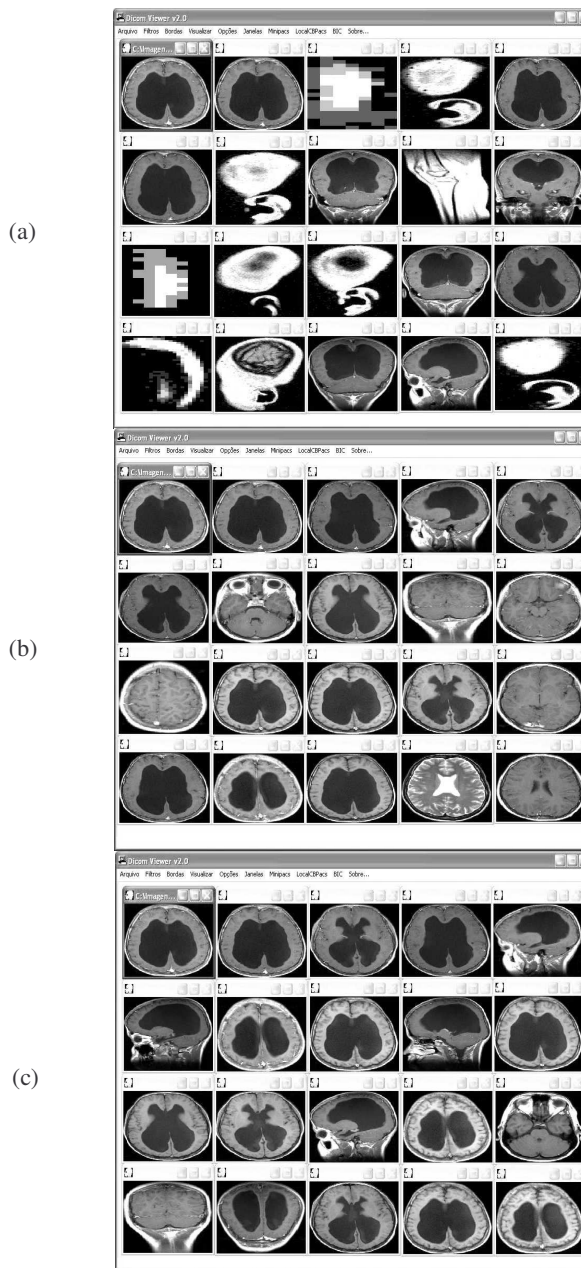


Figura 4.7. Resultados de consultas aos 20 vizinhos mais próximos para: (a) MD, (b) MWD e (c) GWD

Outras consultas realizadas dentro deste experimento revelaram que a GWD e a MWD não apresentam resultados melhores nas consultas em que a imagem de referência é muito escura (com quantidade de *buckets* abaixo de 10). Nesta situação, os poucos valores de níveis de cinza presentes fazem com que dois níveis vizinhos representem regiões distintas e específicas da imagem e, por isso, o efeito *warp* acaba por "confundir" essas regiões. Nesse caso, é melhor optar pelo uso da MD nos cálculos de similaridade.

4.4.3. Experimento 3

Um conjunto de 15 imagens de ressonância magnética contendo cortes sagitais de cérebro humano foram apresentadas a 5 radiologistas e eles foram convidados a classificá-las em ordem de similaridade com uma imagem referência dada. Com esse mesmo conjunto, efetuamos a classificação por similaridade das imagens usando as distâncias MD e GWD. O objetivo foi avaliar qual das distâncias mais se aproxima do critério utilizado pelo especialista. A Figura 4.8 mostra as seqüências obtidas respectivamente pelo radiologista 1 (R1), pela distância GWD e pela distância MD.

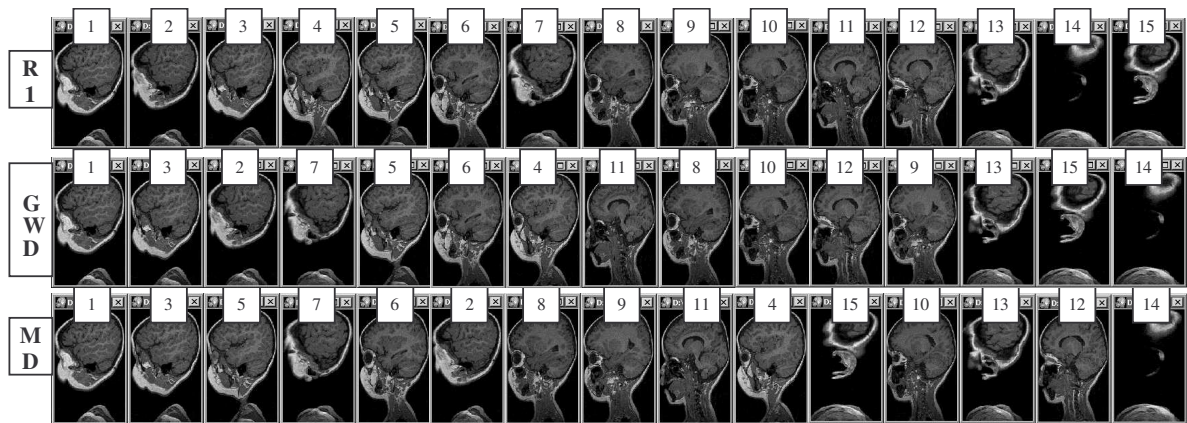


Figura 4.8. Seqüências obtidas com o radiologista 1 (R1), a distância GWD e a distância MD

As seqüências obtidas com os 5 radiologistas foram todas diferentes entre si. A fim de quantificar os resultados, calculamos valores de precisão para diferentes consultas aos k vizinhos mais próximos, com k variando de 2 a 12. Precisão é definida aqui como sendo o grau de coincidência entre os dois resultados, ou seja, o número de imagens que coincidem nas seqüências avaliadas, dividido por k . A Tabela 4.1 contém os resultados obtidos.

Embora a média geral da GWD tenha sido um pouco superior à da MD (5%), esses resultados não são suficientes para que possamos concluir pela superioridade da GWD numa avaliação semântica, pois a quantidade de radiologistas que participaram do experimento foi muito baixa e a quantidade de imagens usadas também foi pequena. Esta última é limitada pela capacidade do analista humano de lidar com várias imagens simultaneamente. O que podemos concluir desse experimento é que a precisão média alcançada pelas funções de distância em relação às seqüências dos radiologistas é razoavelmente boa, o que leva a crer que esses usuários sentir-se-iam satisfeitos com as seqüências geradas automaticamente com o uso das funções de distância. Isto foi confirmado posteriormente junto aos mesmos.

Tabela 4.1. Valores de precisão das seqüências dos radiologistas em relação às seqüências das funções de distância MD e GWD

k	MD						GWD					
	R1	R2	R3	R4	R5	média	R1	R2	R3	R4	R5	média
2	0.5	1	1	1	0.5	0.80	0.5	1	1	1	0.5	0.80
3	0.67	0.67	1	0.67	0.67	0.74	1	1	0.67	1	0.33	0.80
4	0.5	0.75	0.75	0.75	0.5	0.65	0.75	0.75	0.5	0.75	0.25	0.60
5	0.6	0.8	0.6	0.6	0.8	0.68	0.8	1	0.8	0.8	0.6	0.80
6	0.83	0.83	0.83	0.67	0.83	0.80	0.83	0.83	0.83	0.67	0.83	0.80
7	0.86	0.71	0.86	0.71	0.86	0.80	1	0.86	1	0.71	1	0.91
8	0.88	0.75	0.88	0.88	0.75	0.83	0.88	0.88	0.88	0.75	0.88	0.85
9	0.89	0.78	0.89	0.78	0.67	0.80	0.89	0.89	0.89	0.78	0.78	0.85
10	0.9	0.8	0.9	0.8	0.8	0.84	0.9	0.9	0.9	0.8	0.8	0.86
11	0.91	0.82	0.82	0.82	0.91	0.86	0.91	0.82	0.82	0.91	0.82	0.86
12	0.92	0.83	0.92	0.83	0.83	0.87	1	0.92	0.92	1	0.92	0.95
					média	0.79					média	0.83

Uma observação importante sobre este experimento é o fato de que cada radiologista definiu uma seqüência de similaridade diferente, uma vez que cada um deles utilizou um critério diferente para a escolha das imagens na sua seqüência. Argüidos sobre esse critério, as justificativas foram as seguintes: o radiologista 1 valorizou a região principal das imagens, o radiologista 2 valorizou os detalhes das imagens, o radiologista 3 valorizou a posição de corte das imagens, o radiologista 4 valorizou a seqüência de regiões anatômicas presentes nas imagens e o radiologista 5 valorizou o nível de contraste das imagens. Isto ilustra plenamente o alto grau de dificuldade enfrentado pelos sistemas automáticos quando tentam aproximar seus mecanismos de avaliação dos conceitos utilizados pelo analista humano.

4.5. Conclusões

Partindo da avaliação dos resultados obtidos e das discussões conduzidas na Seção 4.4, podemos concluir que a *Global Warp Distance* apresenta, em consultas por similaridade envolvendo imagens médicas, um poder de seletividade superior ao de outras funções de distância específicas para comparação entre histogramas. Os resultados demonstram que a função de distância proposta aumenta a precisão de recuperação em consultas aos vizinhos mais próximos, promovendo a redução de resultados falsos positivos e também de resultados

falsos negativos. Esse maior poder de discriminação se deve ao tratamento aplicado pela função de distância aos histogramas, no sentido de atenuar discrepâncias de brilho e contraste apresentadas pelas imagens e também por considerar características globais dos histogramas no processo de comparação.

Capítulo 5

5. RECUPERAÇÃO POR SIMILARIDADE E CLASSIFICAÇÃO DE IMAGENS MÉDICAS USANDO DESCRITORES DE FORMA E APLICANDO REDUÇÃO DE DIMENSIONALIDADE

Neste Capítulo, apresentamos um método comparativo que combina um extrator de características baseado em forma com um algoritmo para identificação de atributos relevantes para classificação das imagens e outro algoritmo para redução da dimensionalidade. O resultado é um método de apoio ao diagnóstico, computacionalmente eficiente, que permite ao usuário realizar, na mesma operação, recuperação de imagens por similaridade e classificação das mesmas.

5.1. Introdução

Características de forma e textura das imagens podem ser usadas para separá-las em classes distintas predefinidas e essas classes, em se tratando de imagens médicas, podem ser usadas pelos radiologistas para auxílio ao diagnóstico. Um exemplo é a avaliação de nódulos tumorais detectados em mamogramas. Estas imagens podem ser classificadas pelo radiologista, em uma avaliação inicial, como benignas ou malignas, de acordo com a forma do contorno apresentada por lesões detectadas. Nódulos malignos, em geral, infiltram o tecido adjacente, resultando em um contorno irregular ou de difícil distinção, enquanto nódulos benignos apresentam um contorno bem definido e contínuo, como mostrado na Figura 5.1.

Um padrão para classificação de lesões da mama internacionalmente conhecido é o BI-RADS (*Breast Imaging Reporting and Data System*) [ACR '05], que foi desenvolvido em conjunto por vários comitês internacionais, liderados pelo *American College of Radiology* (ACR). O BI-RADS classifica as lesões da mama em seis categorias:

- Categoria 0: necessidade de avaliações adicionais de imagens;
- Categoria 1: tecido normal;
- Categoria 2: lesões benignas;
- Categoria 3: alta probabilidade de lesões benignas;
- Categoria 4: alguma probabilidade de malignidade, recomenda-se biópsia;
- Categoria 5: alta probabilidade de malignidade.

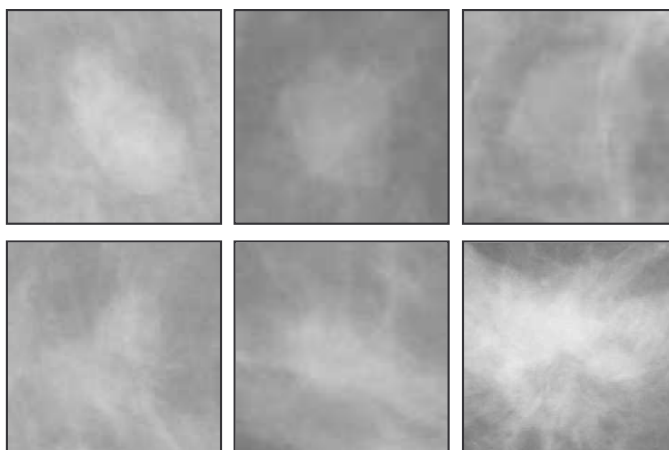


Figura 5.1. Nódulos tumorais de mama típicos: benignos (linha de cima) e malignos (linha de baixo)

Nos últimos anos, os sistemas de Diagnóstico Auxiliado por Computador (*Computer-aided Diagnosis – CAD*) [Yin, T.-K. '04] [Joo, S. '04] [Coppini, G. '03] [Andre, T. C. S. S. '02] [Verma, B. '01] [Giger, M. L. '01] têm recebido especial atenção por parte de pesquisadores de áreas tais como processamento de imagens, inteligência artificial e informática médica. Estes sistemas almejam auxiliar o trabalho de diagnóstico realizado pelos médicos, fornecendo facilidades para acesso a imagens e a informações textuais relacionadas a patologias e diagnósticos.

5.2. Definição do Problema

A representação e análise de imagens médicas por meio de atributos de forma é um campo amplo e complexo. Há diversos trabalhos dedicados a esse assunto [Antani, S. K. '03] [Sclaroff, S. '94] [Korn, P. '98] [Alvarenga, A. V. '02] [Li, Q. '03] [Mlsna, P. A. '04], quase sempre abordando domínios específicos.

Um algoritmo eficiente para caracterização de forma de imagens médicas deverá apresentar os seguintes recursos:

- Ser invariantes a rotação, translação e escala, devido ao fato de não haver uniformidade desses parâmetros no processo de aquisição da imagem;
- Ser capaz de representar a imagem em baixas dimensões, por questões de desempenho, uma vez que os sistemas PACS são complexos e manipulam grande volume de dados;

- Ser capaz de reter informações relevantes sobre patologias, ou seja, permitir que valores numéricos resultantes possam ser associados a parâmetros de análise médica.

Para atingir esses objetivos, o método que será apresentado na Seção 5.3 faz uso de quatro recursos principais: pré-processamento para preparar as imagens; extração de atributos de forma; identificação dos atributos relevantes para classificação e redução da dimensionalidade.

Tentativas iniciais de segmentação de regiões de interesse mostraram-se ineficazes no sentido de discriminar o contorno dos nódulos benignos do contorno dos nódulos malignos. A partir disso, optamos por um método de extração de características de forma que não necessita de segmentação prévia da imagem.

5.3. O Método Desenvolvido

O método proposto tem por objetivo determinar um conjunto de atributos (momentos de Zernike) que possam ser usados como discriminantes para classificar imagens e, ao mesmo tempo, como um conjunto mínimo para definir uma medida de similaridade para recuperar imagens baseando-se em características de forma das mesmas. Como mostrado na Figura 5.2, o método é aplicado em três etapas:

5.3.1. Etapa 1: Pré-processamento e extração de momentos

Um conjunto de imagens de treinamento, consistindo de imagens que representam cada classe de imagens da base de dados, é submetida ao extrator de atributos, o qual gera um vetor de atributos para cada imagem, contendo os seus momentos de Zernike (ver Seção A.5.2).

Antes da extração, cada imagem passa por um pré-processamento que objetiva realçar e diferenciar a região patológica (nódulos tumorais, por exemplo) do restante de tecido. Esse processo depende da natureza das imagens que estão sendo utilizadas e será detalhado na Seção 5.4.1 para a categoria de imagens utilizadas nos experimentos.

Para obter invariância de translação, rotação e escala, o algoritmo que calcula os momentos determina o centro de massa da imagem e define um raio ao seu redor, com um valor necessário e suficiente para conter toda a região de interesse. Esse raio é então normalizado no intervalo $[0, 1]$.

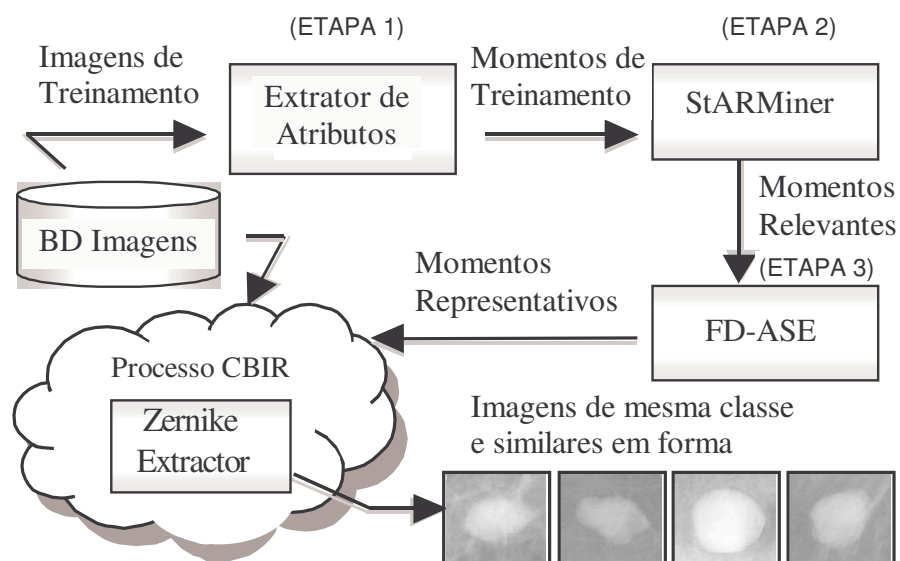


Figura 5.2. Etapas para definição de atributos e CBIR

5.3.2. Etapa 2: Determinação dos momentos relevantes

Para ordem 30, um conjunto de 256 momentos de Zernike são gerados para cada imagem na Etapa 1. Quantos dentre eles são relevantes para discriminar as imagens nas classes originais? Para responder a esta questão, é necessário verificar a relevância de cada momento para a classificação das imagens.

O conjunto de vetores de características, juntamente com as classes conhecidas a priori de cada imagem de treinamento, são usados pelo algoritmo StARMiner (ver Seção 3.8.3) para produzir um conjunto de regras estatísticas de associação, na quais os momentos mais relevantes para discriminar as classes participam nas regras com maior valor de confiança no conjunto resultante de regras. Os momentos são, assim, identificados e o vetor de características de cada imagem é construído usando esses valores.

O StARMiner foi escolhido como gerador de atributos relevantes após termos realizado testes (Seção 5.4.2) para comparar a precisão do mesmo com o algoritmo C4.5 (ver Seção 3.8.3).

5.3.3. Etapa 3: Determinação dos momentos representativos

Mesmo sendo relevantes para a classificação das imagens, os momentos selecionados na Etapa 2 podem apresentar (e provavelmente apresentarão) dependências uns em relação aos

outros. Assim sendo, nesta etapa pretende-se determinar um conjunto de momentos dependentes, juntamente com a quantização da dependência inerentes aos mesmos.

Os momentos relevantes do conjunto de imagens gerado na Etapa 2 são submetidos ao algoritmo FD-ASE (ver Seção 3.8.3). Este retorna um conjunto de atributos que são representativos do conjunto, baseando-se na contribuição de cada atributo na formação da dimensão intrínseca do conjunto de dados. Atributos que não provocam alterações significativas na dimensão intrínseca são considerados dependentes de outros, sendo então descartados.

Cabe aqui ressaltarmos que as Etapas 2 e 3 são ambas necessárias para que possamos obter o conjunto mínimo de atributos representativos. Se aplicarmos apenas a Etapa 2, obteremos o conjunto de momentos relevantes para classificação, mas alguns deles podem ser dependentes entre si e, portanto, desnecessários para a comparação entre imagens. Por outro lado, se aplicarmos apenas a Etapa 3, obteremos um conjunto de momentos independentes, mas não necessariamente todos eles serão relevantes para a classificação das imagens.

Ao final desse processo, a quantidade de momentos é reduzida em cerca de uma ordem de magnitude (de centenas para dezenas). Somente esses momentos selecionados irão compor os vetores de características das imagens que serão usados para indexação e recuperação das mesmas no ambiente de CBIR. Essa abordagem, além de reduzir os custos computacionais para a extração dos momentos, torna o tempo de pesquisa viável para um ambiente operacional de consultas a imagens, onde a base de imagens pode ser bastante robusta e o tempo gasto com o processo de recuperação pode tornar-se crítico.

5.4. Abordagem Experimental

Inicialmente, uma ferramenta de extração de atributos – *Zernike Extractor* – foi implementada para extrair os momentos de Zernike das imagens e para executar consultas aos vizinhos mais próximos. A Figura 5.3 apresenta uma tela da interface na qual a imagem mostrada mais à esquerda é a imagem de referência.

Uma base de dados contendo 250 imagens foi utilizada para testar e validar o método proposto. Estas imagens foram obtidas da biblioteca *Digital Database for Screening Mammography* (DDSM) [IARL '05] [M. Heath K. Bowyer, D. K. '00] disponibilizada pela Universidade do Sul da Flórida. Estas imagens consistem em regiões de interesse (*Regions of Interest* - ROIs) contendo nódulos tumorais, obtidos a partir de exames de mamografia. Todos

os casos foram previamente analisados e classificados por médicos radiologistas e eventualmente confirmados por exames complementares.

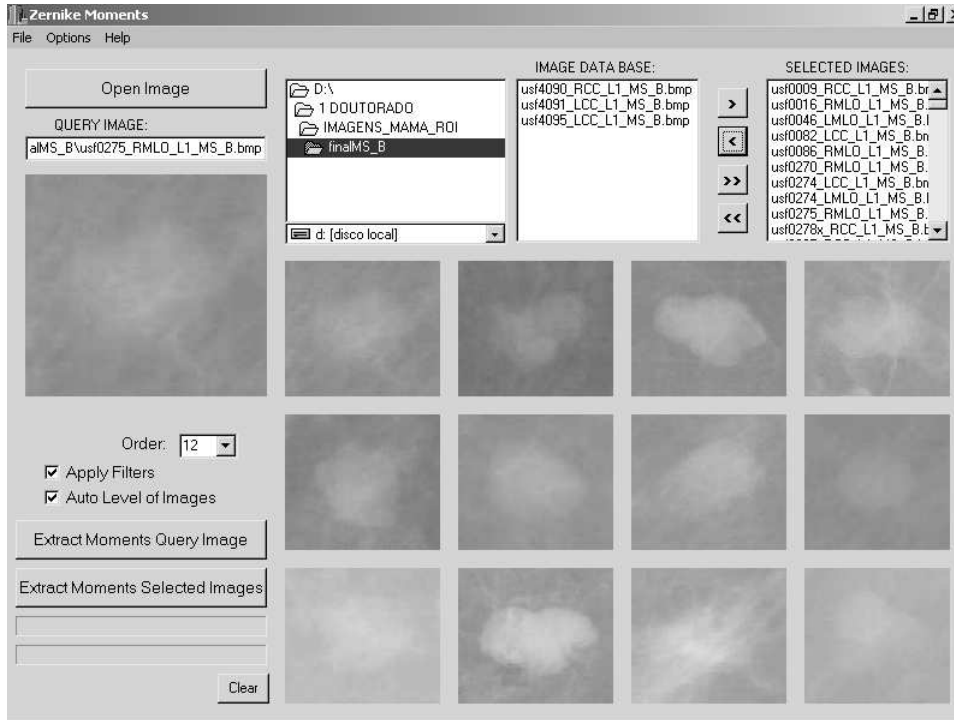


Figura 5.3. Zernike Extractor executando uma consulta aos k vizinhos mais próximos

Apesar dessa quantidade de imagens ser bastante menor do que o tamanho real de bases de dados de imagens médicas, podemos considerar um ambiente em que se usa um banco de imagens-exemplo para se trabalhar com avaliação e diagnóstico. Dessa forma, 250 imagens são suficientes para cobrir um conjunto de imagens típicas de grupos com variações patológicas, e também para serem usadas na escolha de momentos relevantes e para testar a precisão alcançada pelo método.

5.4.1. Testes com Segmentação

Inicialmente, realizamos testes com segmentação das imagens, objetivando destacar os contornos dos nódulos. Dois métodos bastante conhecidos foram empregados: o algoritmo de Linha Divisora de Águas (LDA) (ou *Watershed*) (ver Seção A.3.5) e um aperfeiçoamento do algoritmo EM/MPM, o qual baseia-se em Cadeias Randômicas de Markov (ver Seção 3.8.4).

Dez imagens representativas de cada classe foram segmentadas utilizando cada um desses métodos, pretendendo-se usar o contorno resultante para posteriormente calcular

características de forma (grau de circularidade e grau de irregularidade, dentre outros). A Figura 5.4 mostra alguns desses resultados.

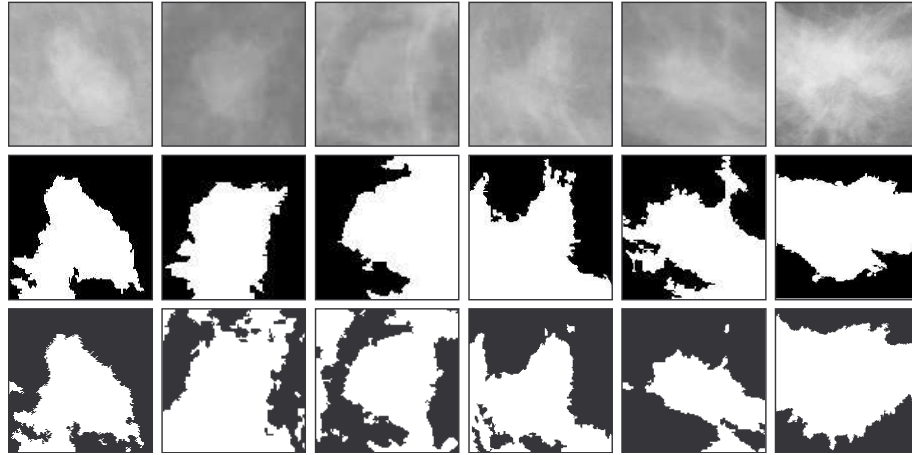


Figura 5.4. 1ª linha: imagens originais (3 nódulos benignos seguidos de 3 malignos); 2ª linha imagens originais segmentadas pelo LDA; 3ª linha: imagens originais segmentadas pelo EM/MPM

A partir desses resultados, é fácil ver que com esses métodos de segmentação não obtemos contornos de nódulos suficientemente próximos do contorno real para permitir a discriminação entre as duas classes. Isto ocorre porque esse tipo de imagem apresenta alta ocorrência de ruído, representado pela superposição do parênquima sobre os nódulos, o que faz com que os algoritmos se confundam no momento da segmentação.

Com essa constatação, partimos então para buscar uma solução que não dependesse da segmentação e de posterior definição do contorno dos nódulos. Avaliando possíveis conceitos e métodos descritos na literatura, chegamos à conclusão de os momentos de Zernike poderiam cumprir satisfatoriamente (ver estudos da literatura apresentados na Seção A.5.2) a função de representar a forma das imagens, sem a necessidade de segmentação prévia.

5.4.2. Pré-processamento das Imagens

Para viabilizar a extração de momentos sem a necessidade de segmentação prévia das imagens, a ferramenta *Zernike Extractor* executa alguns procedimentos básicos de pré-processamento, a fim de destacar regiões de interesse das imagens. No caso deste conjunto de imagens de nódulos tumorais de mama, o procedimento foi o seguinte:

- a) Auto nivelamento: realce de contraste (ver Seção A.3.1) com extensão dos valores máximo e mínimo de nível de cinza da imagem para o intervalo máximo (0-255). Isto é

importante porque a maioria das ROIs extraídas de imagens de mamografia são escuras, apresentando pouco contraste.

- b) Redução no número de níveis de cinza: os níveis de cinza são reduzidos para 12. O intervalo completo (0-255) é particionado em 12 intervalos e os níveis de cinza que se encontram no mesmo intervalo recebem o mesmo valor:

$$\text{novo_nível} = \lambda \text{nível_antigo} * 12 / 256\mu$$

- c) Filtro da mediana (ver Seção A.3.2): este filtro reduz o ruído branco sem afetar muito os contornos do objeto, tornando, assim, as regiões da imagem mais homogêneas.

Uma vez obtido o intervalo de níveis de cinza da imagem, esses três processos podem ser aplicados em uma varredura única sobre a grade de pixels da imagem.

5.4.3. Momentos Relevantes e Momentos Representativos

O método de 3 etapas descrito na Seção 5.3 foi executado sobre um conjunto de imagens de treinamento consistindo de 90 imagens selecionadas da base de dados inicial.

Na Etapa 1, momentos de ordem até 30 foram extraídos, gerando um total de 256 momentos para cada imagem.

Para a Etapa 2, usamos dois algoritmos diferentes e comparamos seus resultados: o gerador de árvores de decisão C4.5 e o gerador de regras de associação StARMiner, descritos na Seção 3.8.3. Para isso, cada vetor de momentos de treinamento foi acrescido da classe (benigna ou maligna) à qual pertence a respectiva imagem.

Para aplicação do C4.5, os valores dos momentos foram discretizados em 4 faixas de variações. C4.5 gerou uma árvore de decisão em cujos nós figuram os momentos mais relevantes, ou seja, com maior poder de discriminação para a classificação das imagens. Desta árvore – mostrada parcialmente na Figura 5.5 – um conjunto de 38 momentos foram identificados como relevantes.

StARMiner teve seus parâmetros calibrados para gerar um conjunto de regras de associação onde outros 38 momentos foram identificados como os mais relevantes.

Na Etapa 3, os 38 momentos relevantes oriundos de ambos os algoritmos C4.5 e StARMiner foram processados pelo algoritmo FD-ASE. FD-ASE utiliza um parâmetro de calibração (*threshold*) que determina o limite de influência que faz com que um atributo seja ou não dependente de outro. Para esse conjunto de dados, o valor de *threshold* foi de 0.5. A Figura 5.6 apresenta o gráfico gerado pelo FD-ASE mostrando, para os 38 atributos obtidos

com o StARMiner, a contribuição de cada um deles para o aumento no valor da dimensão fractal do conjunto.

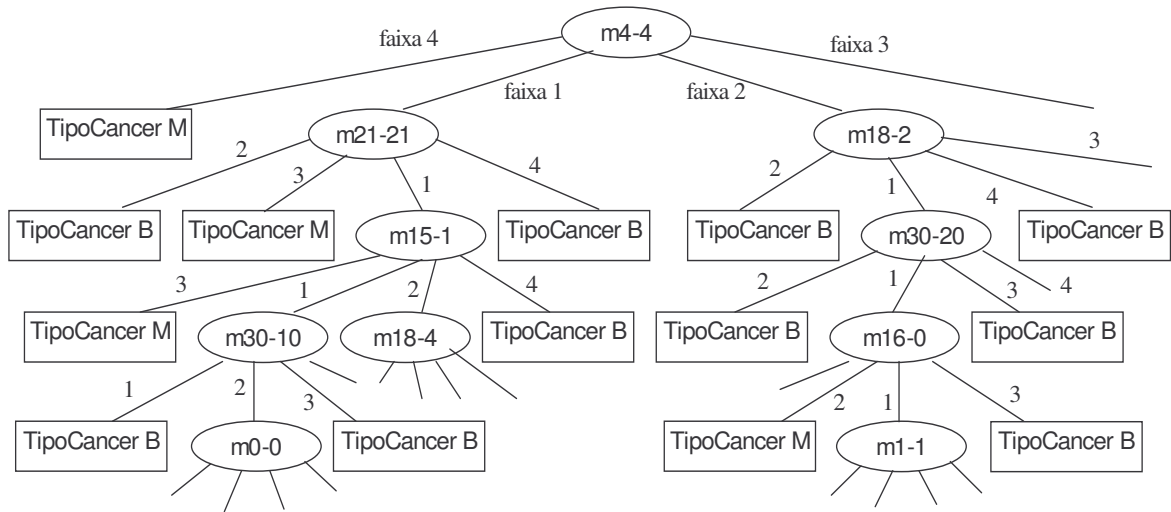


Figura 5.5. Apresentação parcial da árvore de decisão gerada pelo algoritmo C4.5. Os nós são momentos mostrados com ordem-repetição

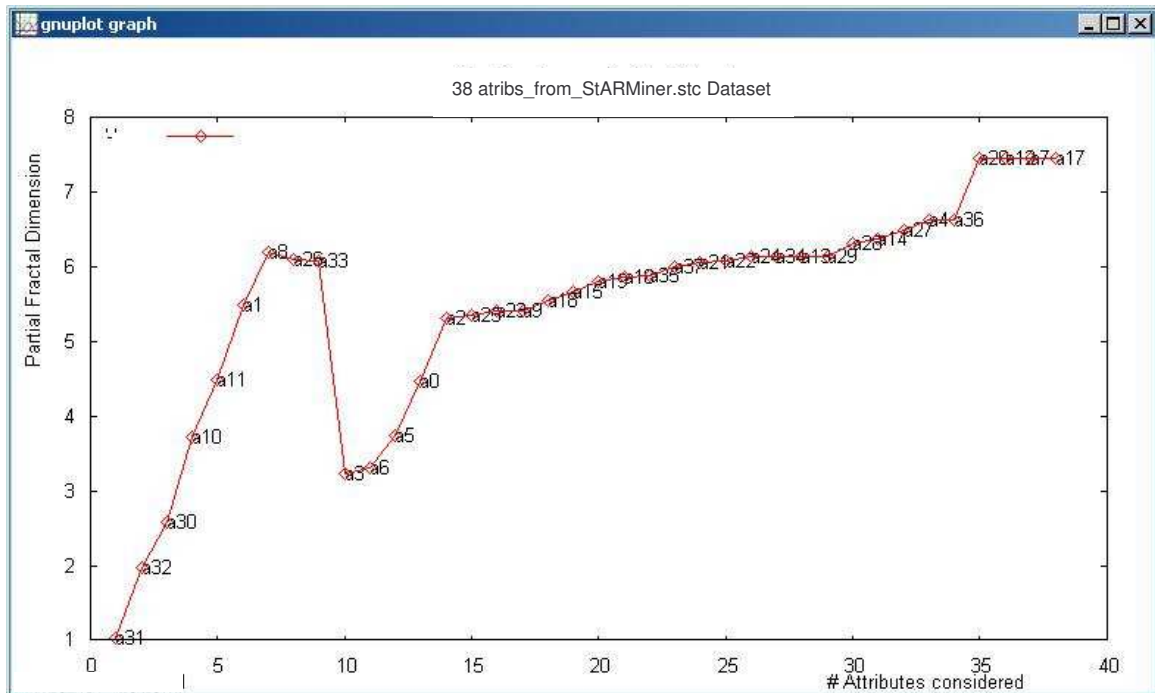


Figura 5.6. Composição da dimensão fractal parcial, considerando a influência da adição de cada um dos atributos ao conjunto

A análise resultante do FD-ASE filtrou um conjunto de 18 momentos representativos a partir do conjunto de relevantes originado pelo C4.5. Para o conjunto de relevantes originados

pelo StARMiner, o FD-ASE filtrou um conjunto de 16 momentos representativos. Esses dois conjuntos finais foram então analisados a fim de determinarmos o mais preciso.

5.4.4. Resultados

Executamos consultas aos k vizinhos mais próximos sobre as 250 imagens da base, tomando aleatoriamente 10 imagens de cada grupo como imagem de referência e variando os valores de k de 1 a 250. Repetimos esse experimento para todos os conjuntos de atributos gerados nas Etapas 1, 2 e 3, a fim de avaliar a precisão obtida com cada um deles.

A Figura 5.7 apresenta as curvas de revocação *versus* precisão resultantes de: 256 momentos obtidos na Etapa 1; 38 momentos obtidos na Etapa 2 com o C4.5; 38 momentos obtidos na Etapa 2 com o StARMiner; 18 momentos obtidos com FD-ASE sobre C4.5 na Etapa 3; 16 momentos obtidos com FD-ASE sobre StARMiner na Etapa 3.

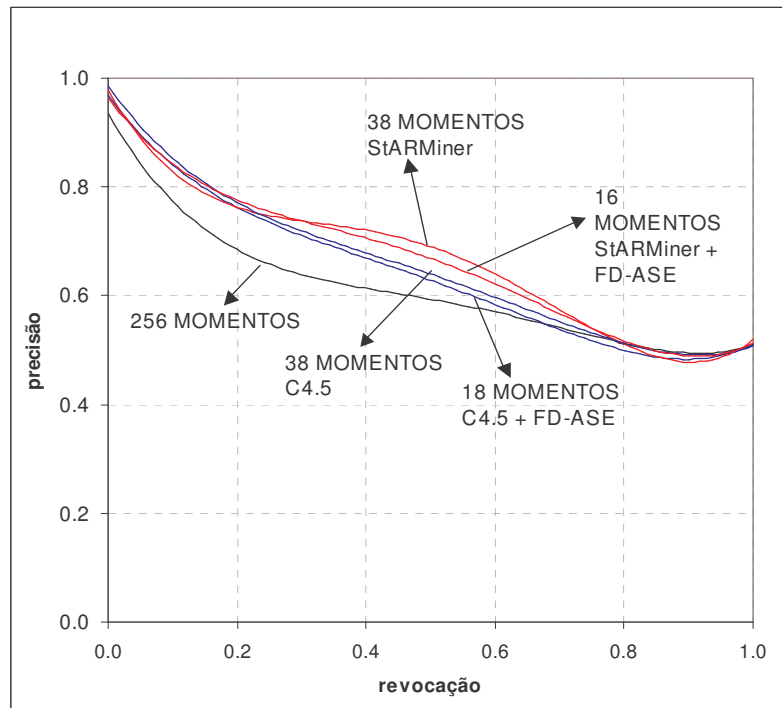


Figura 5.7. Precisão versus revocação: 256, 38, 18 e 16 momentos

Pela análise do gráfico, podemos concluir que:

- Os resultados obtidos com 38 momentos são melhores do que com 256 momentos para ambos StARMiner e C4.5. A redução da dimensionalidade realizada na Etapa 2 forneceu um ganho expressivo de precisão: na região de revocação 0.2, a precisão aumenta de aproximadamente 70% para aproximadamente 80%. As regiões de baixos

valores de revocação são mais importantes em ambientes PACS, devido ao fato de consultas aos k vizinhos usualmente não buscarem por valores elevados de k . Esses resultados comprovam que a maldição da alta dimensionalidade realmente degenera os resultados: os atributos irrelevantes "atrapalham" a influência dos relevantes.

- O algoritmo StARMiner atinge melhor desempenho do que o C4.5. Ele mantém valores elevados de precisão (acima de 0.7) para todos os valores de revocação inferiores a 0.5, enquanto para o C4.5, o valor de revocação que retém esse patamar de precisão está em torno de 0.3.
- A redução na dimensionalidade obtida na Etapa 3 não afeta a precisão dos resultados. A precisão atingida com 16 momentos é praticamente a mesma que com 38 momentos para o StARMiner, ocorrendo o mesmo com os resultados de C4.5.

Com o objetivo de verificar se o conjunto final de atributos obtido consiste no conjunto mínimo que mantém a precisão, removemos um dos momentos de cada conjunto final, resultando em 15 momentos para o StARMiner e 17 momentos para o C4.5. A Figura 5.8 mostra que a precisão sofre um decréscimo significativo para ambos os conjuntos. Isto demonstra que os conjuntos de atributos gerados na Etapa 3 são os mínimos representativos.

Com esses resultados, podemos agora configurar o *Zernike Extractor* para, quando for extrair as características das imagens, extrair apenas os 16 momentos representativos obtidos com o StARMiner e, com esses valores, montar os vetores de características da imagem.

5.5. Conclusões

O método proposto combina em uma única operação os processos de classificação e de recuperação de imagens baseada no seu formato. As características descritivas de forma são representadas pelos momentos de Zernike. A extração de atributos é realizada sobre as imagens pré-processadas, sem a necessidade de segmentação. Além disso, o método cuida para que apenas os atributos relevantes para a classificação e independentes entre si sejam utilizados, obtendo uma redução drástica na dimensionalidade dos vetores de características. A ferramenta desenvolvida para aplicar o método permite ao usuário executar consultas aos vizinhos mais próximos, em tarefas de auxílio ao diagnóstico.

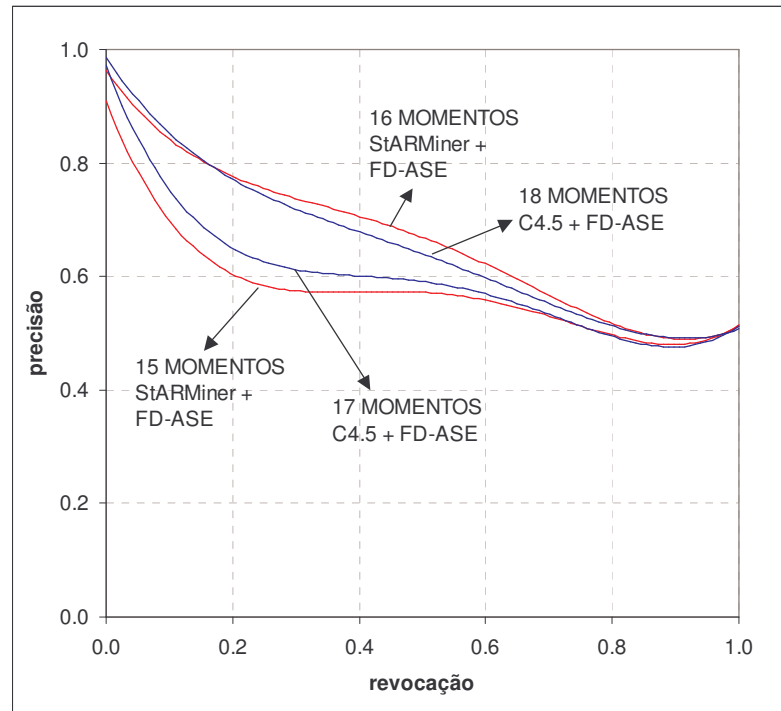


Figure 5.8. Precisão versus revocação: 18, 17, 16 e 15 momentos

Os resultados dos experimentos realizados sobre uma base de imagens composta de ROIs de mamografias apresentando nódulos tumorais benignos e malignos permitiram a escolha de um algoritmo para seleção de atributos relevantes para classificação, dentre dois algoritmos testados. A aplicação do método permitiu uma redução no número de atributos de 256 para 16, levando a um aumento na precisão do método – valores de precisão se mantêm em torno de 80% para valores de revocação inferiores a 20%. Além disso, os resultados experimentais demonstram que o conjunto final de 16 atributos é o conjunto mínimo com o qual é possível manter a precisão para tarefas de recuperação e classificação.

Capítulo 6

6. DESCRITORES DE TEXTURA PARA CARACTERIZAÇÃO DE TECIDOS

Este capítulo apresenta um aprimoramento dos métodos de obtenção e comparação de descritores tradicionais de textura de imagens (descritores de Haralick). O aprimoramento consiste no tratamento da imprecisão gerada por diferenças de posições angulares das imagens comparadas. Além disso, é também proposto um novo descritor – Histograma de Contraste – que, segundo os experimentos apresentados, revela-se mais eficiente do que os descritores tradicionais. Os experimentos realizados avaliam o desempenho de vários dos descritores, assim como uma combinação dos mesmos, para a comparação de imagens médicas.

6.1. Introdução

Dentre as características intrínsecas utilizadas para representar imagens médicas em ambientes de CBIR, **textura** destaca-se como um descritor eficaz para a discriminação de tecidos. E essa tarefa de discriminação é, normalmente, a etapa preliminar exigida pelos outros procedimentos de comparação. Por exemplo, se um radiologista tem em mãos uma imagem que apresenta nódulo de pulmão, ele não desejará compará-la com uma imagem que apresente nódulo de mama, mas sim com alguma outra imagem de pulmão.

Os descritores de Haralick, baseados na abordagem estatística e obtidos a partir das matrizes de co-ocorrência da imagem, têm sido largamente utilizados para a representação baseada em textura da imagem [Smeulders, A. W. M. '00] [Müller, H. '04]. Outra forma eficaz de representação de textura são os coeficientes obtidos a partir da transformada de Wavelets [Lado, M. J. '01] [Semler, L. '05] [Do, M. N. '02] [Castanón, C. A. B. '03] [Ding, S. '04] [Emine Krichen, L. A. '05]. Devido ao fato de necessitar de várias etapas de processamento, a transformada de Wavelets caracteriza um processo computacionalmente caro, quando comparado com o processo de obtenção da matriz de co-ocorrência. Sendo assim, o uso dos descritores de Haralick justifica-se em tarefas onde o esforço computacional apresenta-se como fator crítico para a viabilidade do processo de comparação por similaridade. A

conceituação da abordagem estatística e o processo de obtenção das matrizes de co-ocorrência, assim como os principais atributos de Haralick são descritos na Seção A.2.2.

Vários trabalhos [Aksoy, S. '98] [Celebi, E. '00] [Lehmann, T. M. '05] [Inoue, M. '05] [Joo, S. '04] [Zhou, X. S. '02] [Celebi, E. '00] [Kurani, A. '04] [Kalinin, M. '05] têm usado a abordagem estatística com resultados satisfatórios para imagens gerais. Especificamente para imagens médicas, aplicações dessa abordagem podem ser encontradas em: [Boukerroui, D. '99] para segmentação de imagens de ecocardiografia; [Iakovidis, D. K. '05] para discriminação de elementos em imagens de vídeo de endoscopia; [Vince, D. '00] para discriminação de elementos em imagens de ultra-som de coronárias; [Tsang, W. '05] para auxílio ao diagnóstico de lesões em imagens de tomografia de pulmões; [Takahashi, N. '00] para reconhecimento de textura e forma. Várias pesquisas com imagens médicas considerando a textura têm se concentrado no estudo de mamografias: [Li, H. '04] [Verma, B. '01] [Mudigonda, N. '00] [Honda, M. O. '01] [El-Naqa, I. '02] [Zhang, P. '05] [Panchal, R. '04].

Este Capítulo apresenta um estudo sobre a representação e a comparação de imagens médicas baseando-se em descritores de Haralick. Regiões de imagens representando diferentes tecidos do corpo humano serão utilizadas nos experimentos. O objetivo é caracterizar uma imagem por meio de valores numéricos obtidos a partir de cálculos sobre a sua distribuição de níveis de cinza, que possam ser associados a diferentes nuances de textura, de acordo com o inter-relacionamento entre esses valores. Desenvolvemos um método para comparação de descritores de Haralick que elimina a principal restrição detectada no método usual de comparação, que é a sua sensibilidade em relação à rotação da imagem.

Implementamos uma ferramenta que extrai valores dos descritores, definindo vetores de características, e recupera e classifica imagens usando o novo método. O usuário escolhe uma imagem de referência e executa consultas aos vizinhos mais próximos. O objetivo é verificar a capacidade dos descritores em discriminar imagens de acordo com o tecido representado pelas mesmas.

Uma outra contribuição desta tese é a definição de um novo descritor, que será apresentado na Seção 6.3.

6.2. O Método Desenvolvido

O método que propomos para a comparação de imagens faz uso dos passos básicos de extração de atributos e composição dos vetores de características (ou assinaturas). A distância entre duas imagens é obtida comparando-se as respectivas assinaturas. Especificamente nesse método, foram consideradas 4 direções (θ) e 5 distâncias (d), gerando 20 matrizes de co-ocorrência para cada imagem. Uma vez escolhido um descritor, teremos um valor de descritor para cada par (θ, d). Poderemos então gerar uma assinatura com 4 elementos, calculada a partir da média de valores de d para cada θ , ou então uma assinatura com 5 elementos, calculada a partir da média de valores de θ para cada d . No caso de se utilizar as assinaturas com médias para cada θ , a comparação entre as assinaturas de duas imagens será então realizada por meio de um processo de rodízio, o qual elimina uma possível discrepância no valor da distância devido a incompatibilidade entre posições angulares (rotação) das duas imagens. Os quatro passos do método são mostrados na Figura 6.1 e descritos a seguir.

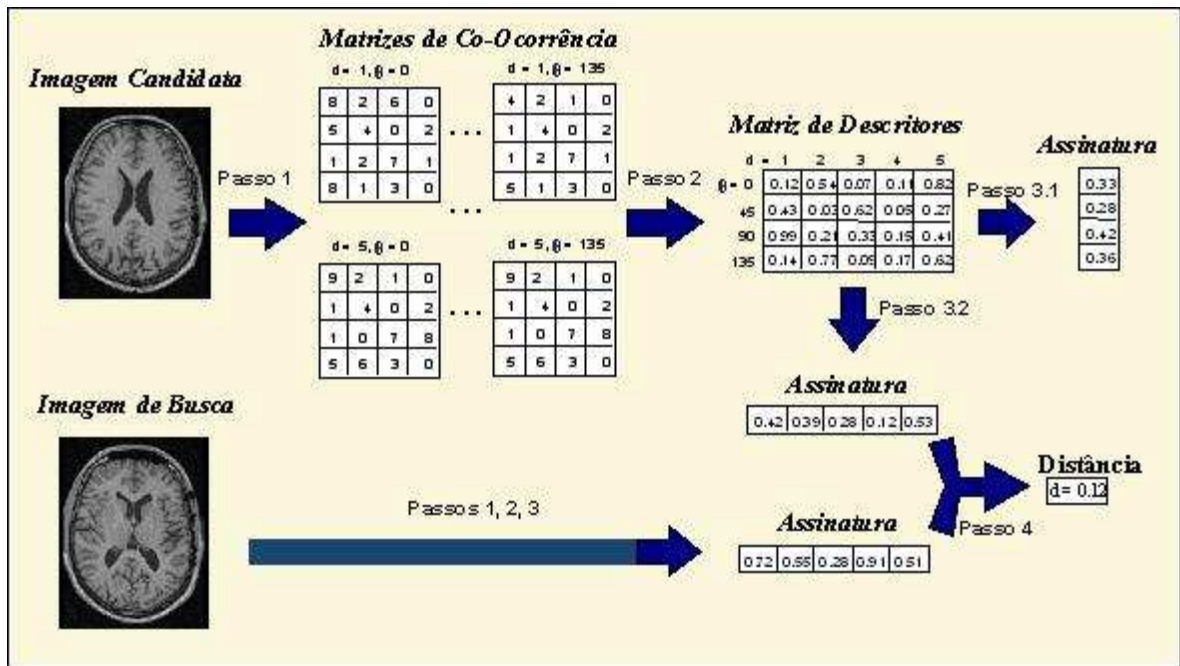


Figura 6.1. Passos do método de comparação entre imagens

Passo 1: Obtenção das matrizes de co-ocorrência de cada imagem. Os níveis de cinza da imagem são reduzidos de 256 para 16. Isto permite uma redução na ordem das matrizes. São calculadas as matrizes de co-ocorrência para as direções de 0° , 45° , 90° e 135° e para as distâncias 1, 2, 3, 4 e 5. Esse processo produz 20 matrizes de 16x16 elementos para cada imagem.

Passo 2: Obtenção dos valores para os descritores escolhidos. Para cada matriz de co-ocorrência, o valor do descritor (ou descritores) é calculado. Para cada imagem, os valores resultantes de descritores são armazenados em uma matriz na qual as linhas representam cada uma das 4 direções e as colunas representam as 5 distâncias.

Passo 3: Geração das assinaturas das imagens. As assinaturas são compostas a partir da matriz de descritores, calculando-se as médias dos valores obtidos com diferentes distâncias, para cada direção.

Passo 4: Comparação das imagens. As assinaturas de duas imagens são comparadas e a distância entre elas é calculada, usando-se a função de distância Euclidiana entre os elementos das assinaturas. Para eliminar a dependência com relação à rotação das imagens, são calculadas 4 diferentes distâncias, executando-se um rodízio entre os elementos referentes aos valores de descritor referentes a cada uma das possíveis direções da imagem. A menor destas distâncias será obtida na posição angular relativa que melhor acopla as duas imagens. Esta será, então, adotada como sendo a distância entre as imagens. Esse processo é mostrado na Figura 6.2. No caso de se combinar mais de um descritor, utiliza-se a função de distância Euclidiana ponderada, a fim de normalizar os intervalos de variação de cada descritor.

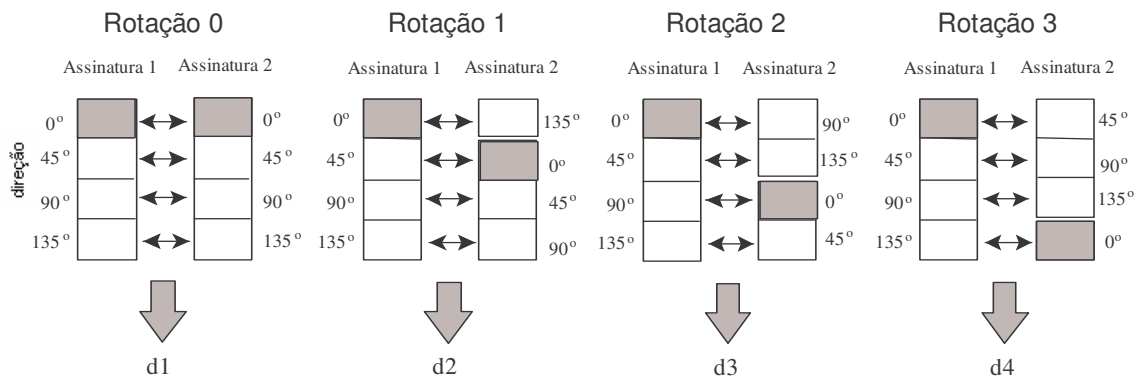


Figura 6.2. Quatro possíveis distâncias, obtidas a partir do rodízio entre os elementos das assinaturas

6.3. O Descritor Proposto

Um refinamento do descritor de Variância (ver Tabela A.1) foi implementado. Chamamos esse descritor de Histograma de Contraste. Ele consiste na geração, a partir da matriz de co-ocorrência, de um vetor de variâncias cujos índices representam valores de $|i-j|$, contendo a soma da quantidade de ocorrências nos elementos com mesmo valor de $|i-j|$ da

matriz de co-ocorrência (Figura 6.3), onde i e j representam respectivamente os índices de linha e coluna da matriz de co-ocorrência.

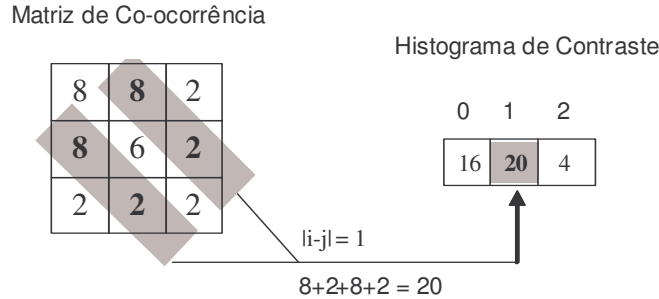


Figura 6.3. Composição do Histograma de Contraste

Este descritor retém ocorrências dos diversos possíveis níveis de contraste da imagem. No exemplo mostrado na Figura 6.3, o elemento 0 do histograma indica que há na imagem 16 ocorrências de vizinhanças de pixels possuindo mesmo valor de nível de cinza, assim como o elemento 1 indica que há 20 ocorrências de vizinhanças de pixels com diferença igual a 1 nos níveis de cinza, e assim, por diante.

Esses vetores são posteriormente utilizados ponderando-se seus valores com o fator $(i-j)^2$ no cálculo das distâncias. Esse fator potencializa a influência do valor de $|i-j|$ no valor final da distância, da mesma forma como ocorre com o descritor de Variância.

A sua utilização permite que se trabalhe com um nível maior de precisão sobre as características de textura da imagem, uma vez que, ao invés de se trabalhar com um único valor de descritor para cada matriz de co-ocorrência, poderemos trabalhar com um conjunto de valores representativos do mesmo descritor.

6.4. Resultados Experimentais

Para os experimentos realizados, utilizamos uma base de 100 imagens contendo regiões de interesse (ROIs) de imagens originais de tomografia computadorizada, ressonância magnética e mamografia, sobre segmentos de diferentes tecidos, formando 9 grupos: cérebro, coluna, coração, pulmões, mama, tecido adiposo, músculo, fígado e osso. Alguns exemplos destas ROIs são mostrados na Figura 6.4.

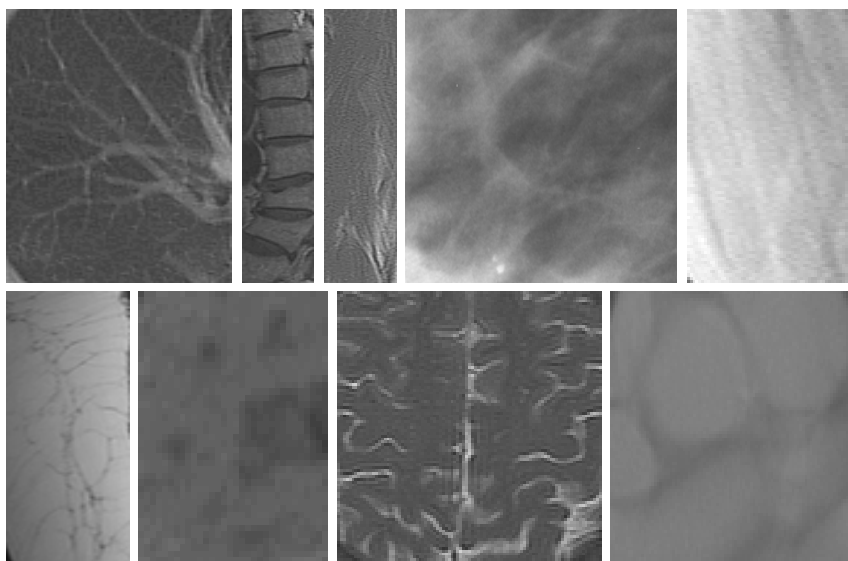


Figura 6.4. Exemplos de ROIs utilizadas nos experimentos. Da esquerda para a direita e de cima para baixo: pulmões, coluna, músculo, mama, osso, tecido adiposo, fígado, cérebro e coração

Uma ferramenta para extração e consulta aos vizinhos mais próximos foi implementada a fim de procedermos aos testes experimentais. Dois tipos de experimentos foram realizados: um para avaliar comparativamente o desempenho dos diferentes descritores e outro para avaliar a eficácia do método em discriminar diferentes tecidos.

6.4.1. Ferramenta para Extração de Características e Consulta aos Vizinhos mais Próximos

A ferramenta desenvolvida para extração de textura (*Texture Extractor*) disponibiliza uma interface onde o usuário escolhe uma imagem de referência e um conjunto de imagens candidatas armazenadas em uma base de imagens previamente construída. O usuário escolhe um descritor (ou uma combinação de descritores) de textura, assim como um método de cálculo da assinatura e executa a extração, seguida da classificação por proximidade das imagens selecionadas com a imagem de referência (Figura 6.5). Todos os seis descritores apresentados na Tabela A.1 foram implementados, além do Histograma de Contraste, proposto na Seção 6.3.

A ferramenta foi implementada em C++, utilizando-se o programa Borland C++Builder,

versão 5. A manipulação de imagens utilizou recursos da biblioteca DicomLib³, implementada pelo Grupo de Bases de Dados e Imagens (GBDI) do Instituto de Ciências Matemáticas e de Computação (ICMC-USP).

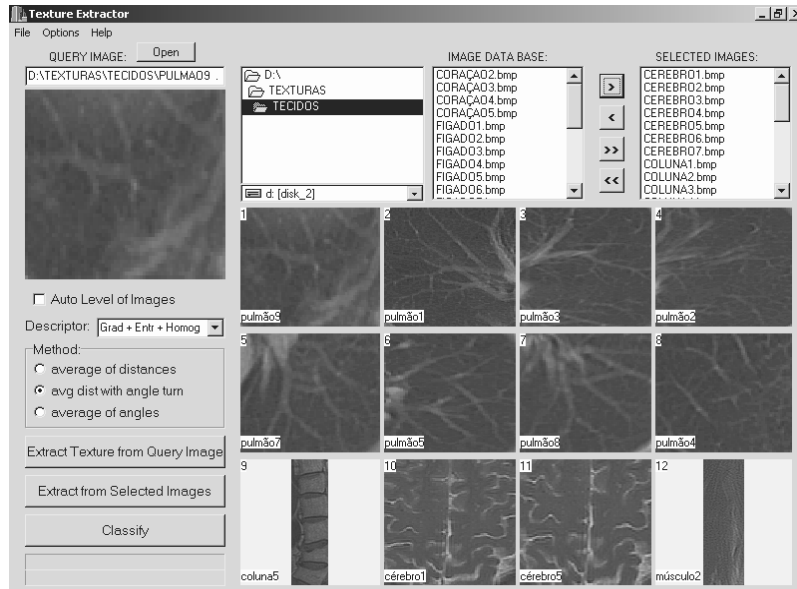


Figura 6.5. Interface da ferramenta *Texture Extractor*

6.4.2. Experimento 1

O objetivo desse experimento foi comparar os descritores implementados na *Texture Extractor*: Variância, Energia, Entropia, Homogeneidade, Momento de Terceira Ordem, Variância Inversa, Histograma de Contraste e também uma combinação (soma das distâncias) de três deles (Histograma de Contraste, Entropia e Homogeneidade). Desejamos analisar o poder de discriminação de cada um deles.

Executamos consultas aos vizinhos mais próximos para cada um dos grupos de imagens e calculamos valores de precisão e revocação.

Uma forma de obter valores numéricos para comparar duas curvas de revocação versus precisão é calcular o valor da área sob as curvas. Sabemos que quanto melhor a curva, maior o valor da área, sendo que área igual a 1 é o melhor valor possível. O valor da área nem sempre é uma medida indicativa de eficácia, uma vez que a região de baixos valores de revocação normalmente é mais importante do que o restante do gráfico. Porém, no caso desse experimento, verificamos que o formato das curvas obtidas é bastante parecido e, por isso, a área sob a curva estará indicando de forma apropriada o ganho de desempenho das mesmas.

³ <http://gbdi.icmc.usp.br/projetos.php>

Os valores de áreas obtidos são também mostrados na Tabela 6.1, onde podemos ver, para cada descritor, o valor médio de área obtido com os diferentes grupos de imagens, assim como o valor do desvio-padrão considerando as curvas de cada grupo, e também o valor de ganho (ou perda) em porcentagem de precisão de cada descritor em relação à área média.

Tabela 6.1. Resultados de revocação versus precisão (áreas) para cada descritor

Descritor	Área Média	Desvio-padrão	ganho em relação ao valor médio (%)
Combinação	0,90	0,06	+32,3
Hist. de Contraste	0,74	0,09	+8,8
Homogeneidade	0,72	0,16	+5,9
Variância	0,69	0,11	+1,4
Entropia	0,64	0,23	-5,9
Momento 3 ^a Ordem	0,62	0,11	-8,8
Energia	0,61	0,24	-10,3
Variância Inversa	0,53	0,11	-22,1
Valor Médio	0,68		

Com os resultados apresentados na Tabela 6.1, podemos ver que a combinação de descritores (Histograma de Contraste, Entropia e Homogeneidade) mostrou-se superior aos descritores individuais, sendo seguida pelos descritores Histograma de Contraste, Homogeneidade e Variância. Considerando apenas os descritores individuais, o descritor Histograma de Contraste, proposto neste trabalho, gerou os melhores resultados, tanto considerando-se valores de área, quanto valores de desvio-padrão. No caso do desvio-padrão, valores baixos indicam maior uniformidade de resultados com relação aos diferentes grupos de imagens testados, ou seja, os descritores que geraram valores mais baixos de desvio-padrão tendem a ser eficazes para todos os grupos de imagens.

O critério adotado para se escolher quais descritores iríamos utilizar para compor a combinação de descritores foi o significado físico de cada um, tendo sido escolhidos descritores relacionados a aspectos distintos de textura. O desempenho ótimo obtido com a combinação confirma a adequação do critério adotado.

6.4.3. Experimento 2

Neste experimento, valores de precisão e revocação foram calculados para consultas aos vizinhos mais próximos, tomando como referência todas as imagens da base e usando a combinação de descritores apresentada na Seção 6.4.2. Os resultados, considerando-se as imagens recuperadas pertencendo ao mesmo grupo da imagem de referência, foram calculados e as médias foram consideradas no traçado dos gráficos para cada grupo de tecidos. Baseando-nos nessas curvas, mostradas na Figura 6.6, podemos analisar a capacidade da combinação de descritores para identificação de diferentes tecidos.

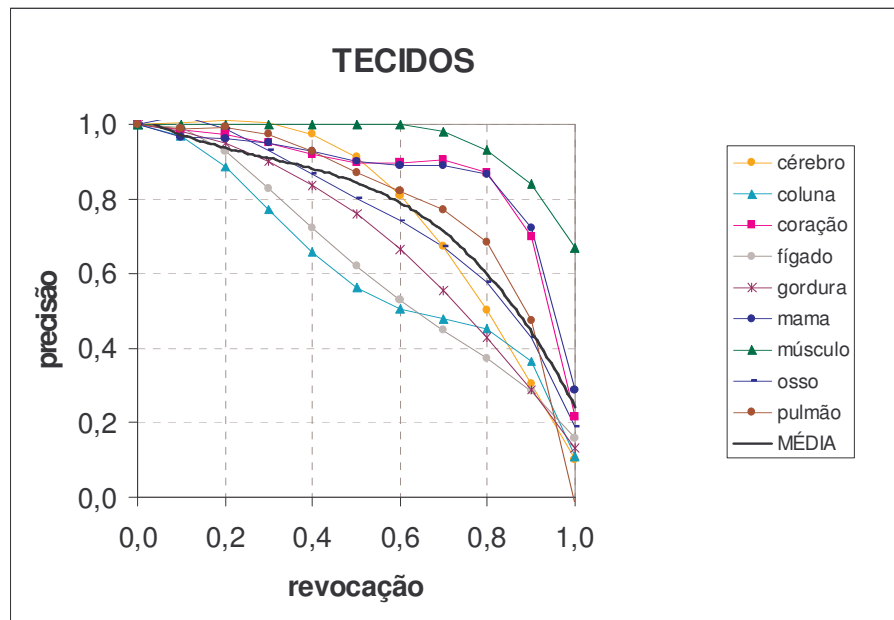


Figura 6.6. Revocação versus precisão da combinação de descritores, para discriminação de diferentes tecidos

Na Figura 6.6, podemos ver que a maioria dos tecidos foi bem discriminada nas consultas por similaridade usando a combinação de descritores proposta. Os resultados inferiores para imagens de coluna provavelmente devem-se ao fato desse tecido apresentar elementos de textura maiores. Logo, o maior valor de distância utilizado na matriz de co-ocorrência ($d=5$) não foi suficiente para capturar as informações de textura desses elementos. Um aumento no valor de d deve produzir resultados melhores para este tecido.

Os tecidos de pulmões, adiposidade e osso possuem texturas parecidas e, por isso, ocorre uma mistura no processo de classificação, degenerando, os resultados. Uma possível solução para esta situação poderia ser o uso de outras propriedades das imagens, tais como distribuições de cores, individualmente ou combinadas com descritores de textura.

6.4.4. Experimento 3

A fim de comparar o método e os descritores estudados neste Capítulo, escolhemos as Wavelets de Haar (ver Seção A.4.3). Utilizando o programa MatLab, versão 7, aplicamos sobre a mesma base de imagens dos experimentos anteriores a transformada de Wavelets de Haar em 3 níveis, extraindo a Energia de cada sub-banda, gerando um vetor de características contendo 12 atributos. Executamos então consultas aos vizinhos mais próximos, tomando, aleatoriamente, como referência 25% do total de imagens da base. Construímos gráficos de revocação versus precisão para poder comparar os métodos e descritores. A Figura 6.7 apresenta a curva obtida com as Wavelets de Haar e a curva média obtida utilizando a combinação de descritores.

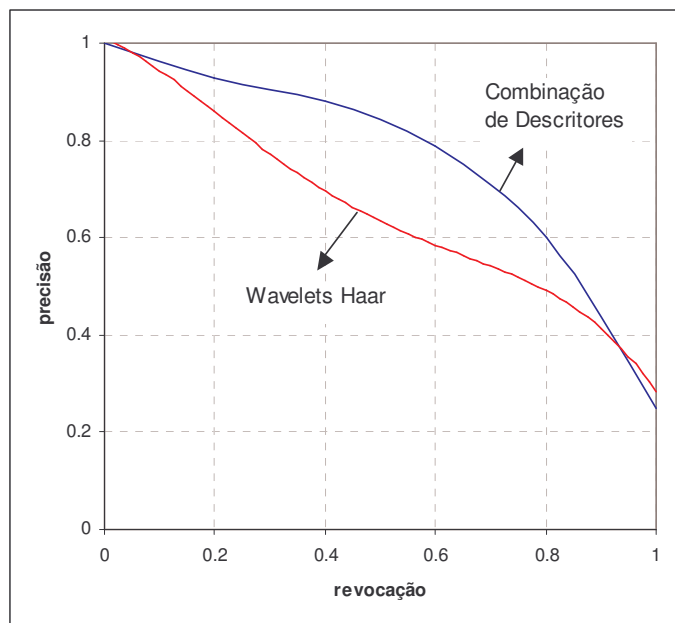


Figura 6.7. Curvas de revocação versus precisão para a combinação de descritores e para as Wavelets Haar

Pelas curvas obtidas, vemos que, para esse conjunto de imagens, a combinação de descritores apresentou melhores resultados. Sabemos, porém, que o método de extratores baseados em Wavelets apresentado no trabalho descrito na Seção 3.8.4 – que utiliza Wavelets de Gabor com uma função de distância especial, baseada em valores de média e desvio-padrão – gerou resultados muito melhores do que com as Wavelets de Haar. Portanto, a exploração das diversas possibilidades dessa transformada deverão levar à obtenção de resultados melhores.

6.5. Conclusões

Neste Capítulo, apresentamos um aprimoramento da técnica clássica de comparação de imagens baseada em descritores de Haralick. Um rodízio entre os elementos do vetor de características, no momento do cálculo da distância, permite que posicionamentos angulares distintos sejam compensados, ajustando-se as direções das duas imagens comparadas.

Um novo descritor – Histograma de Contraste – foi definido como sendo um vetor que contém as ocorrências dos diversos possíveis níveis de contraste da imagem.

Uma ferramenta que calcula os diversos extratores estudados e executa consultas aos vizinhos mais próximos foi implementada para avaliar o comportamento desses extratores em aplicações com imagens médicas.

Três estudos distintos foram realizados usando-se esta ferramenta. O primeiro experimento comparou o desempenho dos diversos descritores e nos permitiu verificar a superioridade do descritor Histograma de Contraste sobre os descritores clássicos. Pudemos também determinar uma combinação de extratores que apresenta desempenho melhor do que qualquer dos extratores individuais. Com o segundo experimento, verificamos a capacidade da combinação de descritores em discriminar diferentes tecidos do corpo humano. Os resultados mostraram que a precisão em recuperação de imagens fica sempre acima dos 90% para consultas que buscam por imagens similares dentro de até 20% de revocação da base de imagens, que caracterizam a região de busca mais usual no ambiente de análise de imagens médicas (a quantidade de imagens que se deseja recuperar em uma consulta é baixa, se comparada com a quantidade de imagens relevantes armazenadas). Além disso, o terceiro experimento revela que a combinação de descritores gera melhores resultados do que as Wavelets de Haar.

Com esses resultados, podemos considerar o método proposto adequado para ser usado em processos que envolvam contextos mais amplos de uso de imagens médicas (integrando-se a métodos de segmentação, por exemplo).

Capítulo 7

7. UMA NOVA FAMÍLIA DE DISTÂNCIAS PERCEPTUAIS

Neste Capítulo, apresentamos a definição e a formalização matemática de uma nova família de funções de distância que têm como base conceitual a interação entre os diversos atributos utilizados na composição dos vetores de características que são comparados. Serão apresentados, também, alguns experimentos preliminares, cujos resultados demonstram a sua eficiência.

7.1. Introdução

Para executar comparações por similaridade entre imagens de modo efetivo, é crucial a escolha de uma função de distância apropriada. Como visto no Capítulo 3, diversas funções de distância têm sido propostas na literatura, usualmente com eficácia limitada a aplicações e contextos específicos.

Ambiente, contexto, aplicação, propósito e natureza dos dados são algumas das variáveis que determinam qual característica será escolhida para representar imagens nas tarefas de recuperação por similaridade. Diferentes descritores, tais como cor, forma e textura podem ser usados para compor os vetores de características. Esses vetores usualmente contêm diversos atributos relacionados com um ou mais descritores. Em um contexto específico, por exemplo, poderíamos decidir usar alguns descritores de cor – como, por exemplo, valores no espaço de cores RGB – juntamente com alguns descritores de textura – por exemplo, homogeneidade e entropia – para serem os atributos que representarão as imagens.

Considerando a forma correta como esses diferentes atributos atuarão no processo de comparação entre duas imagens, visando obter resultados os mais perceptuais possíveis, este Capítulo investiga funções de distância que levam em consideração os efeitos da interação entre os diversos atributos (características) extraídos das imagens.

7.2. Definição do Problema

Ao comparar dois vetores de características, as funções de distância normalmente calculam as diferenças entre os valores apresentados para cada um dos atributos

(*features*) que compõem os vetores de características e calculam a distância considerando todas essas diferenças por meio de um critério uniforme. O máximo que se considera em relação à individualização ou à interação entre os atributos é uma ponderação de pesos entre os mesmos, como faz a distância Minkowski ponderada (Seção 3.2.1), ou a correlação entre eles, como faz a distância Mahalanobis (Seção 3.3.1). Em ambos os casos, os pesos ou a matriz de correlação têm que ser determinados de modo não automático.

Mesmo quando se parte do princípio de que os atributos dos vetores de características são independentes um do outro, há uma pergunta importante que deve ser considerada no momento da comparação entre imagens: como esses atributos afetam a percepção humana de similaridade quando seus valores variam em conjunto? Em outras palavras, se as imagens *A* e *B* apresentam uma grande diferença em apenas um atributo – homogeneidade, por exemplo – e as imagens *A* e *C* apresentam diferenças menores, porém em mais de um atributo – homogeneidade e cor, por exemplo – qual imagem, *B* ou *C*, será considerada mais similar à imagem *A*? Qual é o efeito de variações conjuntas dos atributos no julgamento humano de similaridade?

Nosso objetivo é responder a essa questão para contextos ou propósitos específicos, por meio da definição de famílias de funções de distância (*Attribute Interaction Distances - AID*) que permitem ao usuário estabelecer parâmetros e ajustar a influência da interação entre atributos e, dessa forma, aproximar-se da percepção humana, levando a uma redução do *gap* semântico.

Para simplificar a exposição de idéias, assume-se aqui que há vetores de características de duas dimensões: atributos *a* e *b*. Mais tarde, o espaço de atributos será estendido para qualquer dimensão.

A referência numérica para medir a "interação entre atributos" é a proporção entre as variações de valores desses atributos, apresentada por duas imagens que estejam sendo comparadas. Isto pode ser visto na Figura 7.1, para dois atributos: as imagens *Q* e *C1* apresentam os mesmos valores de diferenças para os atributos *a* e *b* (alta proporção), enquanto as imagens *Q* e *C2* apresentam um valor alto de diferença para o atributo *a* e um valor pequeno de diferença para o atributo *b* (baixa proporção), e as imagens *Q* e *C3* apresentam diferença apenas para o atributo *a* (proporção zero).

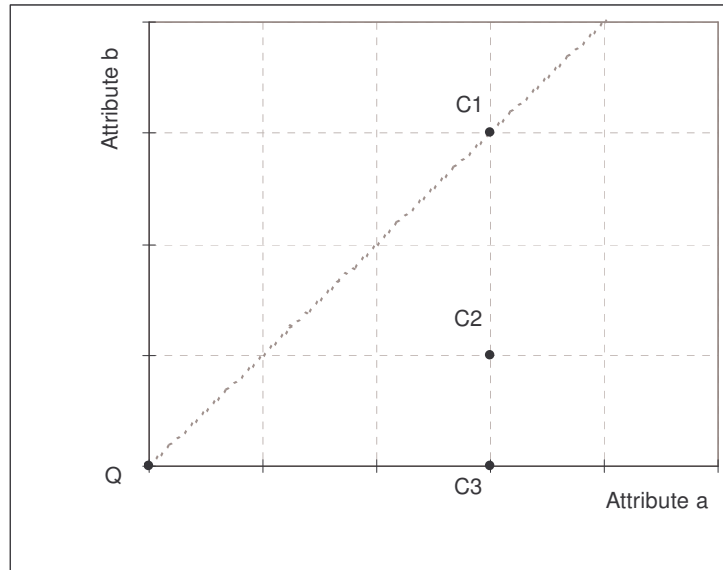


Figura 7.1. Exemplos de interações entre atributos

O que acontece quando duas imagens apresentam uma alta proporção entre atributos? Em alguns contextos, esta alta proporção leva a um efeito perceptual de dissimilaridade mais alta. Quando isto acontece, dizemos tratar-se de um contexto onde a interação entre atributos é **forte**. Por outro lado, uma alta proporção pode levar a um efeito perceptual de dissimilaridade mais baixo. Esse tipo de contexto é dito possuir interação entre atributos **fraca**.

Vejamos um exemplo desse efeito, comparando-se imagens de cores uniformes. Consideremos os valores de Vermelho, Verde e Azul (RGB) da composição de cor, como sendo os atributos que representarão as imagens, com 8 bits de quantização (256 níveis). A imagem *Q* possui valores (130, 30, 30), a imagem *C1* possui valores (130,110,110) e a imagem *C2* possui valores (130,130,30), como mostrado na Figura 7.2. O par (*Q*, *C1*) apresenta alta proporção de diferença, considerando os atributos G e B, enquanto o par (*Q*, *C2*) apresenta proporção zero de diferença (diferença apenas no atributo G). Como veremos nos resultados experimentais da Seção 7.4.1, esse tipo de comparação de cores é um contexto de interação fraca de atributos, o que significa que o ser humano avalia as imagens *Q* e *C1* (alta proporção de diferença) como sendo mais similares do que as imagens *Q* e *C2*. Em outras palavras, variações em apenas um atributo distanciam mais as imagens do que variações em dois atributos, por isso dizemos que a interação é fraca. Para esse contexto de comparação, uma função de distância que considere a interação fraca entre atributos produzirá melhores resultados do que uma função de distância que considere esta interação forte, e

também produzirá resultados melhores do que uma função de distância que não considere a proporção entre os atributos.

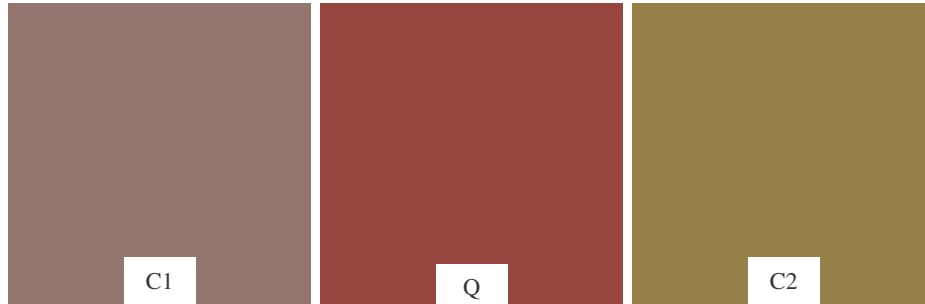


Figura 7.2. Exemplo de alta e baixa proporções de diferença

7.3. Distâncias Propostas: AID (*Attribute Interaction Distances*)

Para lidar com as duas situações identificadas na Seção 7.2 – interação fraca e forte – propomos duas famílias de funções de distância: *Weak Attribute Interaction Distances* (WAID) e *Strong Attribute Interaction Distances* (SAID). Ambas são definidas como polinômios de grau 2, os quais determinam o lugar geométrico de objetos que possuem a mesma distância a um objeto de referência. A família WAID "aproxima" objetos que apresentam alta proporção de diferença, enquanto a SAID "afasta" esses objetos, como pode ser visto na Figura 7.3. Nesse gráfico, todos os pontos sobre a curva WAID estão à distância $d1$ do ponto Q , enquanto todos os pontos sobre a curva SAID estão à distância $d2$ de Q . Tomando o ponto $C2$, por exemplo, se escolhermos trabalhar com a WAID, a sua distância até Q (distância $d1$) será menor do que se escolhermos trabalhar com a SAID (distância $d2$). Comparando as distâncias de $C2$ e $C3$ para Q , podemos ver que se escolhermos a WAID, a distância $(Q, C2) = d1$ será menor do que a distância $(Q, C3) = d3$; porém, se escolhermos a SAID, a distância $(Q, C2) = d2$ será maior do que a distância $(Q, C3) = d3$.

Nesses exemplos, estamos trabalhando com o primeiro octante das curvas, o que significa que a diferença dos valores do atributo a é maior do que a diferença dos valores do atributo b . Se a diferença do atributo b for maior, as mesmas curvas serão aplicadas, porém no segundo octante, espelhadas pela linha de identidade.

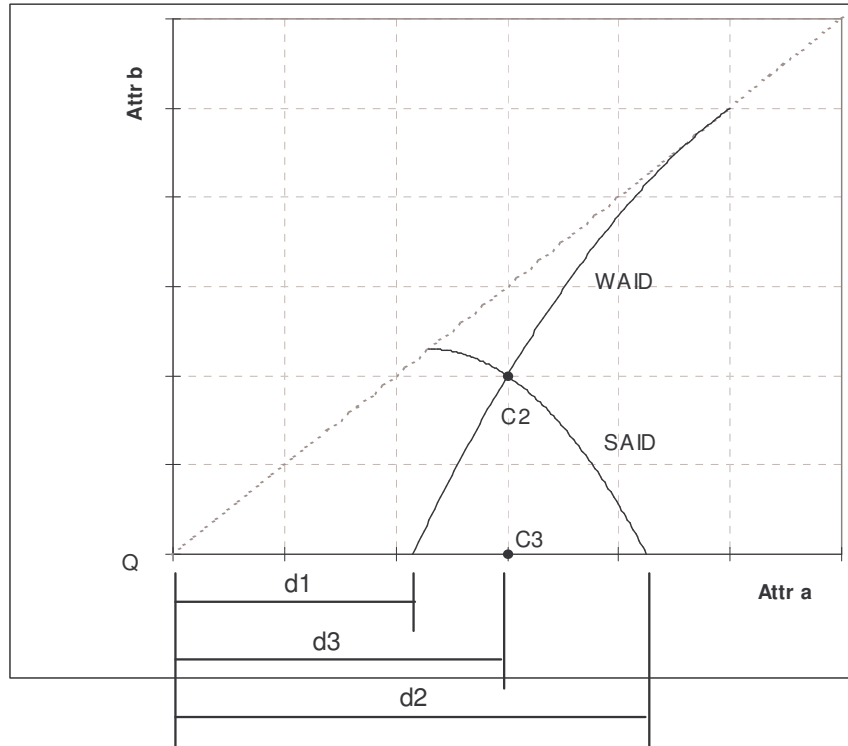


Figura 7.3. Curvas de WAID e SAID. Todos os objetos localizados sobre a curva da WAID estão à distância d_1 do objeto Q. Do mesmo modo, todos os objetos localizados sobre a curva da SAID estão à distância d_2 do objeto Q

7.3.1. Grau de Interação

Para poder quantificar e controlar o efeito da interação de atributos, definimos um parâmetro n que representa o grau dessa interação e determina a alongação das curvas. Com os diferentes possíveis valores de n , teremos uma família de curvas para a WAID e outra família para a SAID. A Figura 7.4 mostra algumas das curvas de cada família. Podemos ver que quanto maior o valor de n , maior será o afastamento dos pontos na família SAID e maior será a aproximação dos pontos na família WAID.

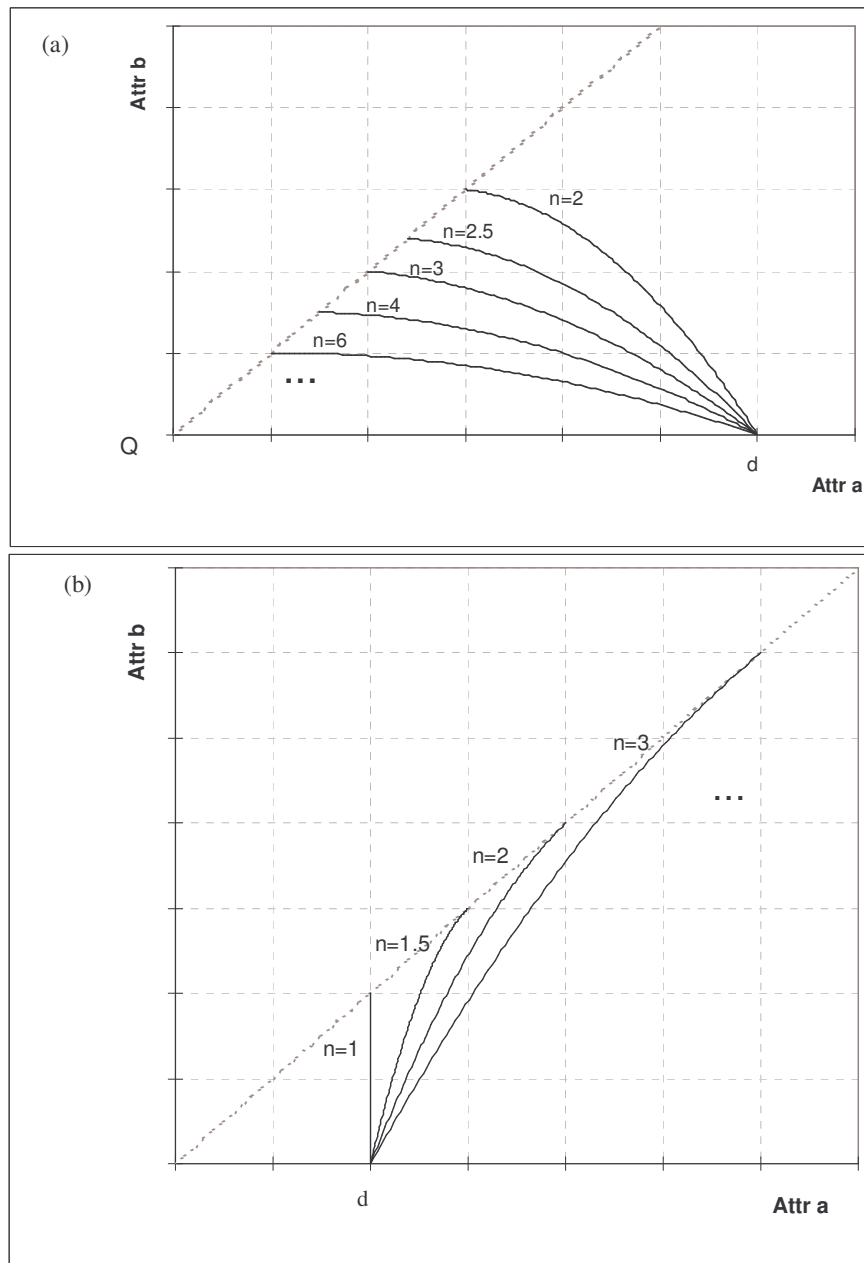


Figura 7.4. Lugar geométrico dos pontos localizados à distância d , para diferentes valores de n : a) família SAID; b) família WAID

7.3.2. Expressões Gerais das Famílias AID

A partir de algumas condições pré-estabelecidas, que caracterizam o desenvolvimento das curvas, as equações que representam as famílias SAID e WAID são determinadas a seguir.

Família SAID

Para determinar a expressão quadrática geral da família SAID, consideramos as seguintes condições de contorno (mostradas na Figura 7.5):

- $f(d) = 0$
- $f(d/n) = d/n$
- $f'(d/n) = 0$ (f máximo em d/n)

onde d é o valor da distância definida pela curva e n é o parâmetro de interação entre atributos.

Considerando

$$y = f(x) = ax^2 + bx + c \quad (7.1)$$

onde x é a maior diferença dos atributos e y é a menor (curva no primeiro octante)

Com estas condições, encontraremos os seguintes valores para os coeficientes a , b e c :

$$a = -\frac{n}{d(n-1)^2}, \quad b = \frac{2}{(n-1)^2}, \quad c = \frac{d}{n} \left(1 - \frac{1}{(n-1)^2} \right) \quad (7.2)$$

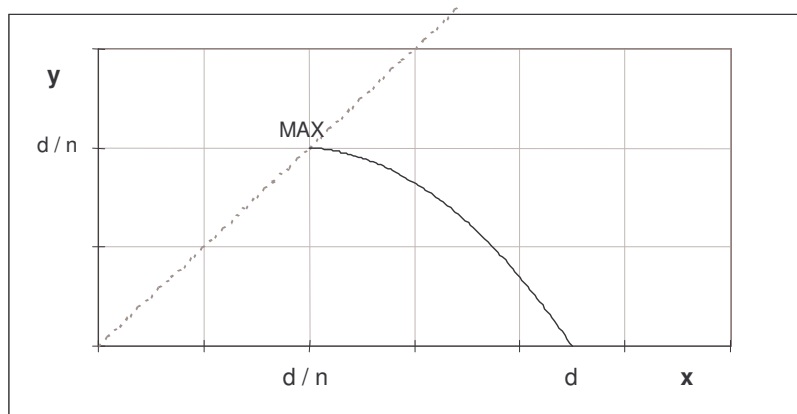


Figura 7.5. Condições para determinar a expressão geral da família SAID

Substituindo os valores de a , b e c em (7.1), obtemos uma expressão quadrática de d em função de x , y e n . Calculando as suas raízes, determinamos o valor de d como:

$$d = \frac{B \pm (n-1)\sqrt{\Delta}}{2(n-2)} \quad (7.3)$$

$$\text{com } n \neq 2 \text{ e } \Delta = 4x^2 - 4xy + y^2(n-1)^2 \text{ e } B = y(n-1)^2 - 2x \quad (7.4)$$

Se $n=2$, a equação de d deixa de ser quadrática e torna-se linear. Nesse caso, o valor de d será dado por:

$$d = \frac{2x^2}{2x - y} \quad (7.5)$$

Da expressão (7.3), vemos que há dois possíveis valores de distância. O valor que desejamos é aquele que é sempre positivo:

$$d = \frac{B + (n-1)\sqrt{\Delta}}{2(n-2)} \quad (7.6)$$

Família WAID

Para determinar a expressão quadrática geral da família WAID, consideramos as seguintes condições (mostradas na Figura 7.6):

- $f(d) = 0$
- $f(nd) = nd$
- $f'(nd) = 1$ (f é tangente à identidade em nd)

onde d é o valor da distância definida pela curva e n é o parâmetro de interação entre atributos.

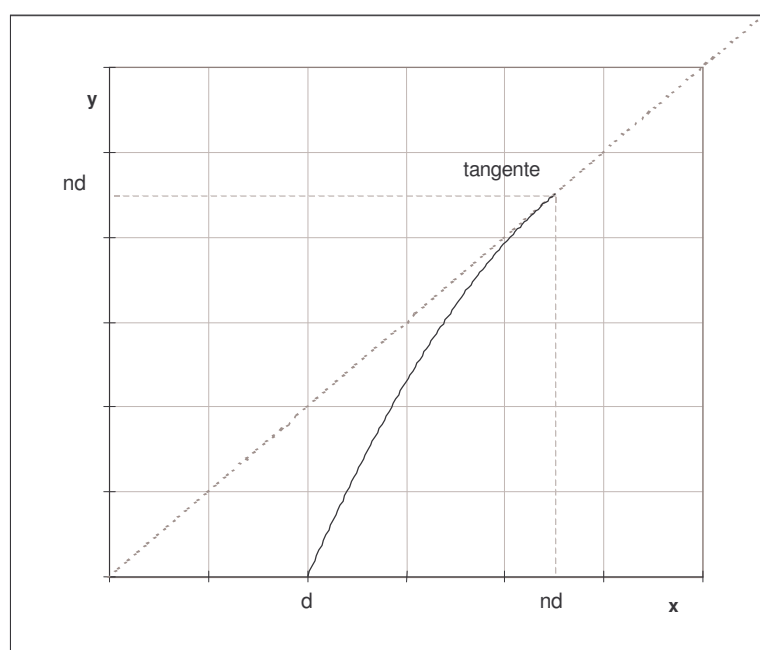


Figura 7.6. Condições para determinar a expressão geral da família WAID

Novamente considerando:

$$y = f(x) = ax^2 + bx + c$$

relembrando que x é a maior diferença dos atributos e y é a menor (curva no primeiro octante).

Com estas condições, encontraremos os seguintes valores para os coeficientes a , b e c :

$$a = -\frac{1}{d(n-1)^2}, \quad b = \frac{(n-1)^2 + 2n}{(n-1)^2}, \quad c = \frac{-dn^2}{(n-1)^2} \quad (7.7)$$

Substituindo os valores de a , b e c em (7.1), determinamos o valor de d como:

$$d = \frac{B \pm \sqrt{\Delta}}{2n^2} \quad (7.8)$$

$$\text{com } \Delta = B^2 - 4n^2x^2 \quad \text{e} \quad B = (n-1)^2(x-y) + 2nx \quad (7.9)$$

Da expressão (7.8), vemos que há dois possíveis valores para a distância. O valor que desejamos é aquele cuja curva apresenta a maior curvatura. Para isso, o valor de d será:

$$d = \frac{B + \sqrt{\Delta}}{2n^2} \quad (7.10)$$

7.3.3. Estendendo as Expressões das Famílias AID para Dimensões Superiores

As expressões de d que foram determinadas na Seção 7.3.2 para as famílias SAID e WAID consideram que apenas dois atributos estão compondo o vetor de características. Nesta Seção, vamos estender aquelas expressões para qualquer dimensionalidade, além de simplificá-las usando as distâncias L_1 e L_∞ .

As expressões (7.4), (7.5), (7.6), (7.9) e (7.10) fornecem o valor da distância entre duas imagens, considerando-se dois atributos. As variáveis x e y representam as diferenças entre os valores dos respectivos atributos nas duas imagens. Além disso, x representa a maior diferença.

Quando houver mais do que dois atributos, iremos considerar a análise da interação entre x (o maior valor de diferença) e cada um dos outros atributos em um plano separado, e então compô-los para formar uma expressão final. Para isso, as distâncias L_1 e L_∞ podem ser bastante úteis, no sentido de simplificar as expressões finais. Pela definição destas distâncias (ver Seção 3.2), teremos:

$$L_\infty = \max \text{ atrib} = x \quad (7.11)$$

$$L1 = \sum \text{todos_atrib} \quad (7.12)$$

onde *atrib* representa o módulo da diferença de valores de um dado atributo para as duas imagens que estão sendo comparadas.

Estendendo a Família WAID

Na análise da relação entre *x* e cada um dos outros atributos, iremos tomar o valor da média para determinar a distância final. Para isto, *y* será substituído pela média dos valores de todos os atributos, excluindo-se *x*:

$$y = \text{média_outros_atrib} = \left(\sum \text{todos_atrib} - \max \text{atrib} \right) / (\text{dim} - 1) \quad (7.13)$$

onde *dim* é o número total de atributos

$$\text{Dessa forma, } y = (L1 - L\infty) / (\text{dim} - 1) \quad (7.14)$$

Agora, substituindo os valores de *x* (expressão (7.11)) e *y* (expressão (7.14)) nas expressões (7.9):

$$\Delta = B^2 - 4n^2 L\infty^2 \quad \text{e} \quad B = (n - 1)^2 (L\infty \text{dim} - L1) / (\text{dim} - 1) + 2nL\infty \quad (7.15)$$

Com Δ e *B* das expressões (7.15), usamos a expressão (7.10) para determinar a distância WAID entre dois objetos num espaço multidimensional.

Estendendo a Família SAID

As funções SAID, para mais de duas dimensões, serão compostas de forma similar à distância L1, ou seja, as diferenças serão somadas para compor a distância final. Desse modo, para determinar a distância SAID, *y* será substituído pela soma dos outros valores de diferenças:

$$y = \text{soma_outros_atrib} = \sum \text{todos_atrib} - \max \text{atrib} \quad (7.16)$$

$$\text{Dessa forma, } y = L1 - L\infty \quad (7.17)$$

Agora, substituindo os valores de *x* (expressão (7.11)) e *y* (expressão (7.17)) na expressão (7.4), temos:

$$\Delta = 8L\infty^2 - 4L1L\infty + (L1 - L\infty)^2 (n - 1)^2 \quad \text{e} \quad B = (L1 - L\infty)(n - 1)^2 - 2L\infty \quad (7.18)$$

Com Δ e B das expressões (7.18), usamos a expressão (7.6) para determinar a distância SAID entre dois objetos num espaço multidimensional.

7.3.4. Composição das Famílias AID com a Família L_p

Se analisarmos as diferenças entre as distâncias L_1 , L_2 e L_∞ , podemos observar que elas apresentam uma diferenciação que considera a interação de atributos. Analisando graficamente, na Figura 7.7, o ponto C_1 representa forte interação entre atributos, enquanto os pontos C_2 e C_3 , fraca interação. Na Figura 7.7(a), o ponto C_2 está mais distante do ponto Q do que o ponto C_3 , qualquer que seja a distância escolhida. Se escolhermos a distância L_1 , o ponto C_1 estará à mesma distância (r_2) de Q do que o ponto C_2 , enquanto se escolhermos a L_2 , o ponto C_1 estará à mesma distância (r_1) de Q do que o ponto C_3 , ou seja, como r_1 é menor que r_2 , o uso da distância L_1 , comparada com a L_2 , "afasta" o ponto C_1 . Na Figura 7.7(b), o ponto C_2 está mais próximo do ponto Q do que o ponto C_3 , independentemente da distância escolhida. Se escolhermos a distância L_∞ , o ponto C_1 estará à mesma distância de Q do que o ponto C_2 , enquanto se escolhermos a distância L_2 , o ponto C_1 estará à mesma distância de Q do que o ponto C_3 , ou seja a distância L_∞ "aproxima" o ponto C_1 .

A maior discrepância entre L_1 , L_2 e L_∞ ocorre na linha de identidade, decrescendo à medida que elas se aproximam dos eixos, até se sobreporem, sobre os eixos. Desse modo, se tomarmos como referência a distância L_2 , podemos dizer que a distância L_1 é uma função para contextos onde a interação entre atributos é forte, enquanto a distância L_∞ é para contextos onde a interação entre atributos é fraca.

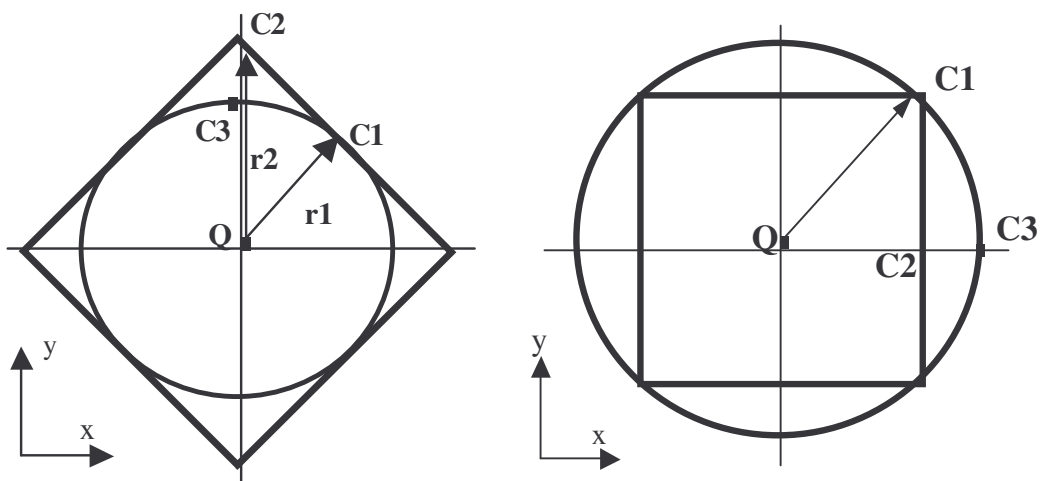


Figura 7.7. Comparações entre L_1 e L_2 (a) e entre L_∞ e L_2 (b)

Porém, estas funções de distância apresentam algumas restrições:

1. Elas são limitadas a uma região específica do espaço: a região triangular delimitada pela reta da L_1 e as duas retas da L_∞ , no espaço bidimensional. Isto significa que um contexto com interação entre atributos mais fraca do que aquela associada à L_∞ não teria como ser representada;
2. L_1 e L_∞ são lineares com a variação de proporção entre atributos. Como veremos nos resultados experimentais, uma variação quadrática se ajusta melhor à percepção humana de interação;
3. L_∞ não é uma boa medida para a tarefa de comparação de imagens por similaridade, uma vez que ela considera como iguais objetos que na verdade podem ser bastante dissimilares.

As famílias SAID e WAID, se associadas à família L_p , permitem estender a cobertura de possibilidades para todo o espaço de atributos, como pode ser visto na Figura 7.8. O limite superior da SAID é a L_1 (SAID com $n=2$) e o limite inferior da WAID é a L_∞ (WAID com $n=1$). Aumentando os valores de n , a SAID se expande para baixo e a WAID se expande para cima, cobrindo todo o espaço, conforme mostrado na Figura 7.4.

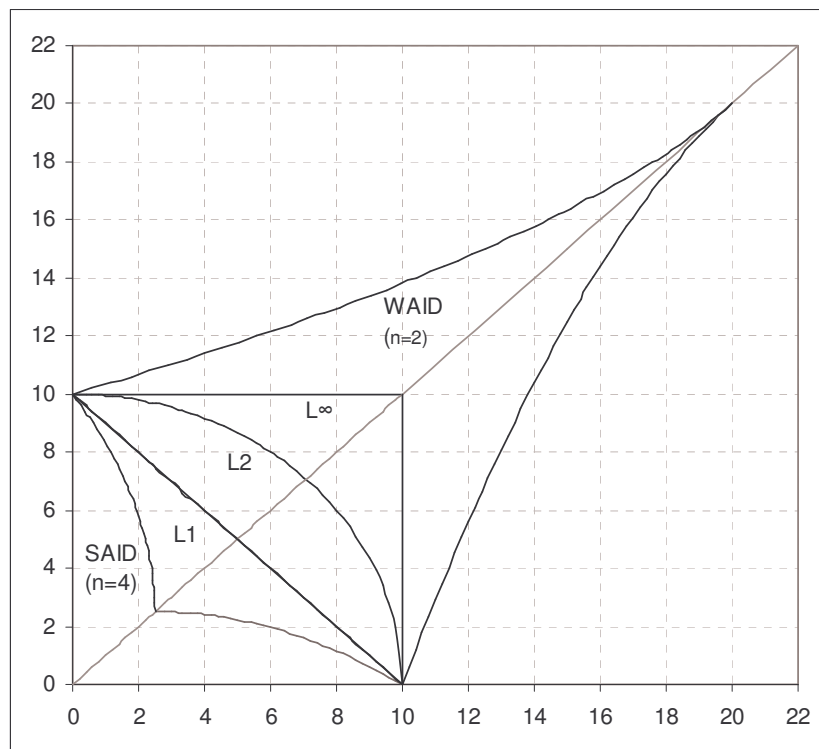


Figura 7.8. Composição da família L_p com as famílias AID

Na verdade, a família L_p pode ser estendida inferiormente, se forem utilizados valores de p menores do que 1. Porém essas funções não refletem a interação entre atributos, uma vez

que as variações apresentadas pelas suas curvas não são consistentes com taxas crescentes do valor da distância em relação ao valor da interação entre os atributos.

Consideremos dois pontos Q e C, representando duas imagens no espaço de atributos. A composição das famílias L_p e AID fornece um escopo de opções de distâncias (que podem ser estendidas variando-se o parâmetro n) que permitem que a interação entre atributos seja avaliada e quantificada com precisão em um contexto (ver Figura 7.9). A partir dessa avaliação, podemos, então, determinar a função de distância mais adequada e eficaz para ser utilizada em buscas por similaridade naquele contexto.

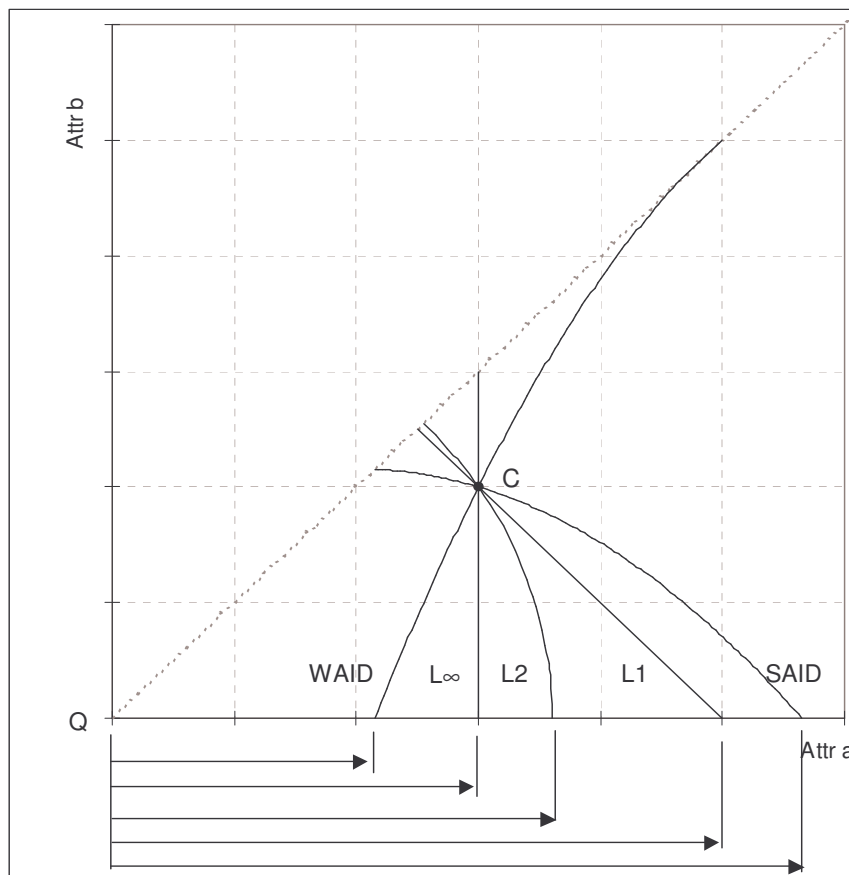


Figura 7.9. Variações da distância de C a Q, com as funções de distância das famílias L_p e AID, considerando o primeiro octante do espaço 2D.

7.3.5. Procedimento para Determinar a Função de Distância mais Adequada

Dado o espectro de funções de distância resultante da composição entre as famílias AID e L_p , desenvolvemos um procedimento de duas etapas para determinar a melhor função de distância dentro do espectro, para um contexto específico.

Etapa 1: Determinar a natureza de interação entre atributos. Com um conjunto de treinamento de imagens, executar testes de similaridade com as distâncias SAID ($n=3$), L1,

L_2 , L_∞ e WAID ($n=3$) ($n=3$ é um valor inicial, que, se necessário, será refinado na Etapa 2). Analisar os resultados de cada distância e compará-las com resultados obtidos de avaliações humanas. Identificar a melhor função de distância para o contexto e, por meio dela, determinar a natureza de interação entre atributos do contexto: se a SAID for a mais apropriada, significa que estamos em um contexto de interação forte entre atributos; se a WAID for a melhor, então temos um contexto de interação fraca entre atributos; se uma das distâncias L_p é a melhor, então temos um contexto de média interação entre atributos.

Etapa 2: Determinar o grau de interação entre atributos e a respectiva função de distância. Se a interação entre atributos é média, concluímos que a melhor função de distância para o contexto é a L_p que alcançou a melhor precisão na Etapa 1. Se a interação entre atributos é forte, então executamos novos testes com a SAID, variando os valores de n , e determinamos qual valor alcançou a melhor precisão. Esta será a função de distância mais adequada para o contexto. Se a interação entre atributos é fraca, então executamos com a WAID esse mesmo procedimento e determinamos a função de distância mais adequada.

Dessa forma, conseguimos determinar a melhor função de distância perceptual para um contexto específico, como poderemos ver nos experimentos descritos na Seção 7.4.

7.4. Resultados Experimentais

A fim de verificar a aplicabilidade das famílias AID, alguns experimentos preliminares foram realizados, em diferentes contextos e com diferentes atributos. O procedimento proposto na Seção 7.3.5 foi executado e funções de distância ideais foram identificadas para três situações específicas de busca por similaridade: imagens compostas por cores uniformes, imagens médicas caracterizadas por atributos de textura e imagens médicas segmentadas em regiões, caracterizadas por atributos gerais, dentre eles, a dimensão fractal.

7.4.1. Similaridade Perceptual de Cores

Avaliar a interação entre atributos de imagens e a sua influência sobre a percepção humana não é uma tarefa trivial. Isto porque em uma imagem muitos atributos diferentes estão presentes simultaneamente, independentemente de quais deles tenham sido escolhidos para representar as imagens. Ou seja, muitas vezes não temos controle sobre os atributos que estão envolvidos no processo de comparação. Então, quando um usuário demonstra a sua

percepção de similaridade entre duas imagens, nada nos garante que ele não tenha sido influenciado por algum atributo alheio ao conjunto que escolhemos como relevantes.

Neste primeiro experimento, o nosso objetivo foi trabalhar com imagens muito simples, onde houvesse a predominância evidente de alguns atributos, os quais foram utilizados por nós para a avaliação das funções de distância. Para isso, escolhemos avaliar a influência da interação entre atributos relativos à cor da imagem e construímos um conjunto sintético de imagens de cores uniformes. Os valores de RGB de cada cor foram usados para representar a respectiva imagem, por meio de vetores de características de três atributos. Doze lâminas foram montadas, cada uma contendo três imagens, sendo uma de referência e as outras duas variações específicas desta, em seus valores de R, G e/ou B. A Figura 7.10 mostra quatro exemplos destas lâminas, onde a imagem Q é a imagem de referência e as imagens C1 e C2 são as suas variações.

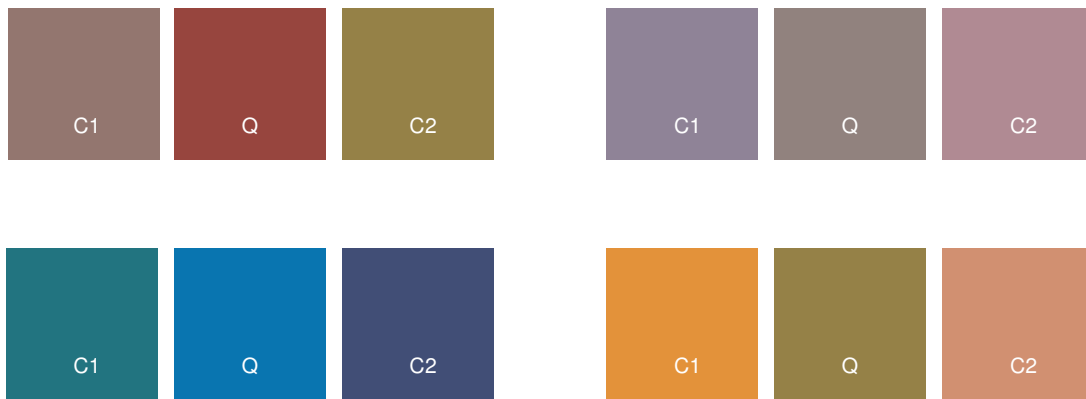


Figura 7.10. Lâminas de cores. A cor central (Q) é a referência e as cores laterais (C1 e C2) são variações específicas da mesma

Em cada lâmina, a imagem C2 é resultado da variação da imagem Q em apenas um dos atributos (R, G ou B), enquanto a imagem C1 é resultado de variações iguais em dois dos atributos. Por exemplo, na primeira lâmina, a qual foi apresentada também na Figura 7.2, como vimos anteriormente, a imagem Q possui valores de RGB (130, 30, 30), enquanto a imagem C1 possui valores (130, 110, 110) e a imagem C2 possui valores (130, 130, 30), ou seja, C1 varia 80 unidades nos atributos G e B e C2 varia 100 unidades apenas no atributo G. Em todas as outras lâminas, essas mesmas variações foram aplicadas para se obter as cores C1 e C2. Assim, vemos que entre Q e C1 existe alta interação entre atributos, enquanto entre Q e C2 a interação entre atributos é nula.

As 12 lâminas foram apresentadas a 50 indivíduos de diferentes idades, sexos e graus de escolaridade e lhes foi solicitado que escolhessem, em cada lâmina, entre C1 e C2, a imagem mais similar à imagem Q.

Calculamos as distâncias da família AID e as da família L_p entre Q e C1 e entre Q e C2. Esses valores são mostrados na Tabela 7.1

Tabela 7.1. Distâncias das imagens C1 e C2 à imagem Q

Imagem	SAID (n=3)	L1	L2	L^∞	WAID (n=3)
C1	240.0	160.0	113.1	80.0	59.1
C2	100	100	100	100	100

Comparando os valores das distâncias obtidas para C1 e C2 com cada uma das diferentes funções de distância, vemos que se usarmos umas das distâncias SAID, L1 ou L2, a imagem C2 será considerada mais similar à imagem Q, uma vez que os valores dessas distâncias são menores para C2 do que para C1. Por outro lado, se usarmos L^∞ ou WAID, a imagem C1 será considerada mais similar à imagem Q. Essa situação se repete em todas as 12 lâminas utilizadas no experimento.

Os resultados obtidos com os 50 indivíduos foram os seguintes: das 600 escolhas (12 lâminas para 50 pessoa), 385 (64.2%) elegeram as imagens C1 como mais similares à imagem Q. Esses resultados revelam que trata-se de um contexto de fraca interação de atributos, ou seja, variações em apenas uma das cores tendem a afastar mais as imagens, do que variações menores em duas das cores.

Das distâncias avaliadas, temos duas que se mostram adequadas aos resultados dos seres humanos: a L^∞ e a WAID.

Entretanto, a L^∞ não é considerada uma boa função de distância para comparar imagens, uma vez que esta função gera o mesmo valor de distância (a máxima diferença entre os atributos) para imagens que podem ser bastante dissimilares entre si. Para ilustrar isto, a Figura 7.11 mostra a imagem Q da Figura 7.2 e outras 4 imagens bem diferentes, porém cuja distância L^∞ à imagem Q é a mesma.

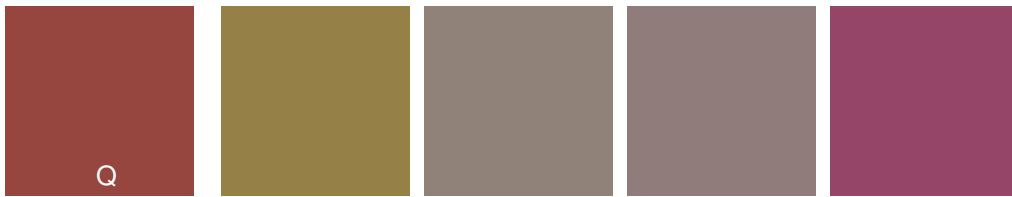


Figure 7.11. Imagens que apresentam o mesmo valor de distância L^∞ em relação à imagem Q

Com esses resultados, podemos então considerar que a percepção humana, no que concerne à interação entre valores de primitivas de cor, tende à fraca interação e, dessa forma, a melhor função de distância para comparar imagens nesse contexto é a WAID.

7.4.2. Similaridade Perceptual de Textura para Imagens Médicas

Para realizar um experimento com imagens reais e com usuários especializados, formamos um conjunto de 30 imagens médicas, consistindo de regiões de interesse (ROIs) extraídas de imagens de ressonância magnética, tomografia computadorizada e mamografia. Algumas delas são mostradas na Figura 7.12. Estas imagens foram obtidas junto ao CCIFM da FMRP-USP.

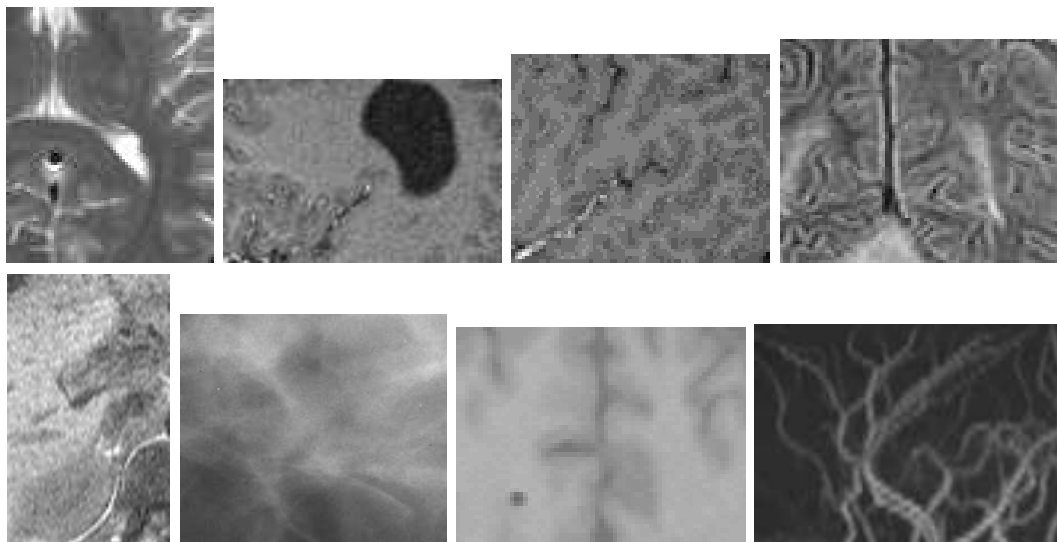


Figura 7.12. Exemplos de imagens usadas no conjunto de testes com atributos de textura

Cinco radiologistas (R1, R2, R3, R4 e R5) foram convidados a organizar as imagens em ordem de similaridade em relação a uma imagem escolhida como referência, tomando como critério de análise a sua percepção em relação à composição de textura de cada imagem.

Esse conjunto de imagens de teste foi processado pela ferramenta *Texture Extractor*, descrita na Seção 6.4.1. Os descritores de textura escolhidos para representar as imagens no experimento foram Uniformidade e Homogeneidade, obtidos a partir das matrizes de co-ocorrência (ver Seção A.2.2) das imagens. Consultas aos k -vizinhos mais próximos foram executadas nesse conjunto de teste, usando cada uma das funções de distância das famílias AID e L_p , de acordo com a Etapa 1 do procedimento proposto na Seção 7.3.5.

A seqüência de imagens obtida com cada radiologista foi então comparada com os resultados de cada função de distância. Para poder analisar os resultados e avaliar o grau de concordância entre as seqüências dos radiologistas e as das funções de distância, calculamos a precisão de cada seqüência, da seguinte forma: dividimos as 2 seqüências que estão sendo comparadas em subseqüências de k imagens; contamos o número num de imagens presentes em ambas as subseqüências e calculamos $precisão(k) = num/k$; variamos k de 2 a 29 e calculamos, para cada k , $precisão(k)$; calculamos a precisão final entre as 2 seqüências como a média de $precisão(k)$. A Figura 7.13 apresenta os valores resultantes de precisão obtidos entre cada função de distância e cada radiologista.

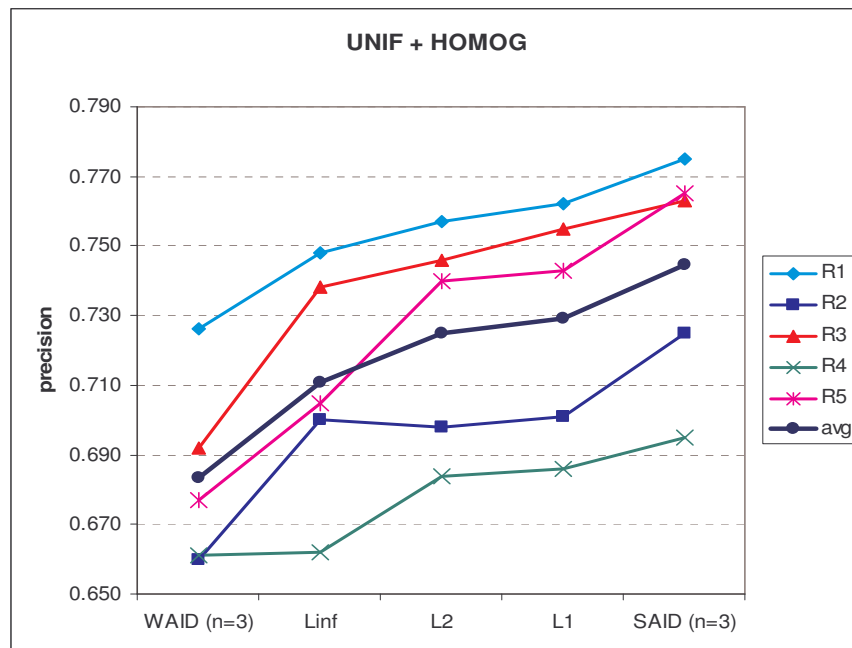


Figura 7.13. Valores de precisão obtidos para avaliar o desempenho de cada função de distância. Cada curva corresponde à comparação com a seqüência de um radiologista, além da curva média (avg)

Analisando as curvas, pode-se ver que o melhor resultado é obtido pela SAID. Isto indica que este contexto é caracterizado por interação forte entre atributos.

Passamos agora para a Etapa 2 do procedimento, onde aplicamos SAID com diferentes valores de n , a fim de determinar a melhor função de distância, dentro desta família.

Escolhemos valores de n entre 3 a 15 e novamente comparamos as seqüências obtidas com as seqüências dos radiologistas. A Figura 7.14 apresenta as curvas resultantes para os diferentes valores de n .

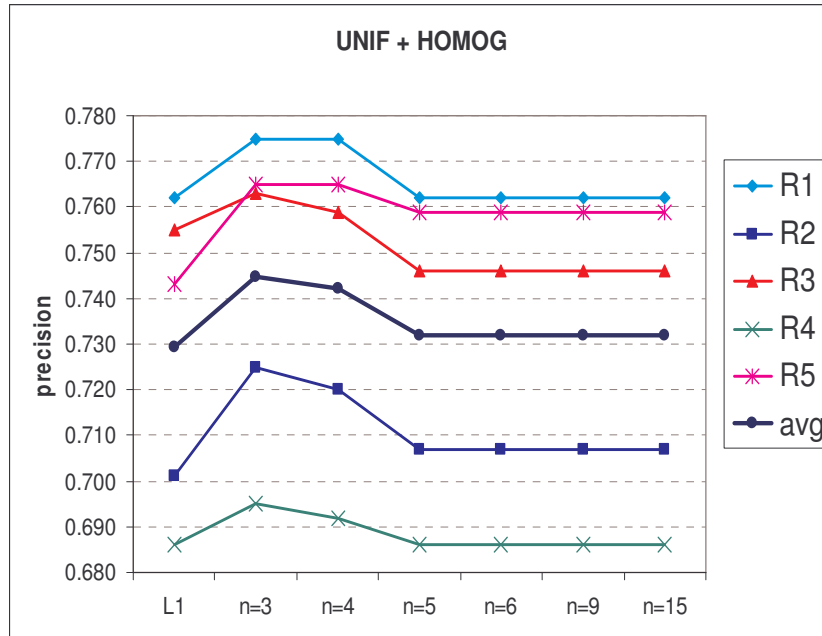


Figura 7.14. Valores de precisão para determinar a melhor função de distância SAID, considerando o parâmetro n .

Analisando estas curvas, podemos ver que os valores $n=3$ e $n=4$ apresentam desempenhos semelhantes, enquanto para valores maiores do que 4 o desempenho diminui. Com isto, concluímos que, neste contexto, usando os descritores Uniformidade e Homogeneidade para representar a textura das imagens, a melhor função de distância é a SAID com $n = 3$.

Realizamos também experimentos com as mesmas características do experimento descrito aqui, porém acrescentando dois novos descritores de textura: Variância e Entropia. Os resultados obtidos na Etapa 1 estão mostrados no gráfico da Figura 7.15.

Com esses resultados, vemos que, para esse contexto, com esses atributos, a função L1 atinge desempenho médio ligeiramente melhor do que a SAID com $n=3$. Para confirmar este resultado, executamos a Etapa 2 variando n de 3 a 15 e verificamos que valores de n maiores que 3 não geraram resultados melhores para a função SAID. Dessa forma, podemos concluir que a melhor função de distância para esse caso é a L1.

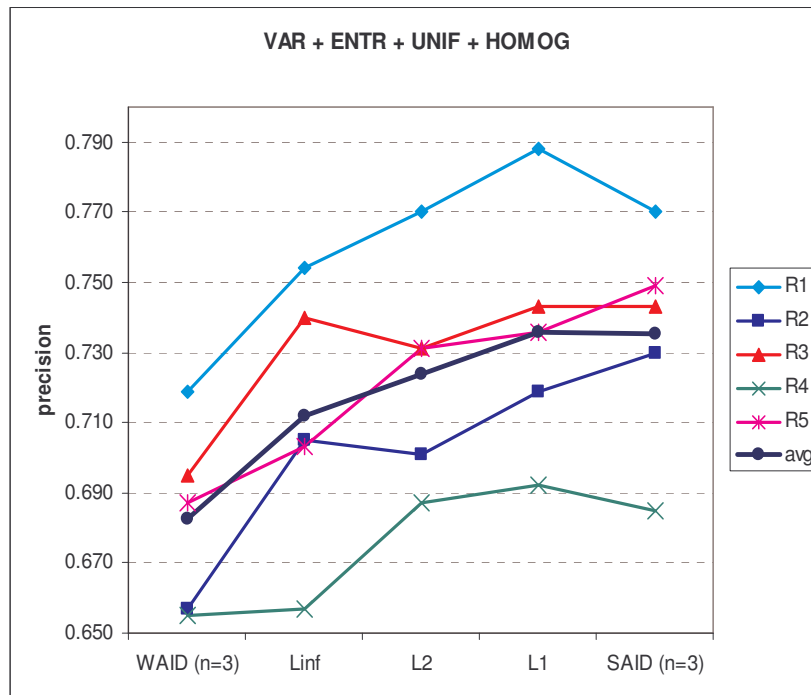


Figura 7.15. Valores de precisão utilizando 4 atributos de textura

7.4.3. Similaridade Perceptual de Imagens Médicas com Atributos Variados

Nesse experimento, o objetivo é testar as funções de distância na sua capacidade em discriminar imagens oriundas de diferentes regiões do corpo humano. Foi utilizada uma base de imagens médicas que consiste de 704 imagens de ressonância magnética, separadas em 8 classes: cabeça axial, cabeça sagital, cabeça coronal, pélvis axial, abdômen axial, abdômen coronal, angiograma e coluna sagital.

Cada imagem foi segmentada em 5 regiões e de cada região foram extraídos 6 atributos: a dimensão fractal, a coordenada x do centro de massa, a coordenada y do centro de massa, a massa, o nível de cinza médio e o coeficiente linear da linha de ajuste usado para a obtenção da dimensão fractal. Informações detalhadas dos procedimentos e técnicas empregadas podem ser encontradas em [Balan, A. G. R. '05]. Com isso, para cada imagem foi gerado um vetor de características composto por 30 atributos (6 atributos para cada uma das 5 regiões da imagem).

Consultas aos vizinhos mais próximos foram executadas, tomando-se aleatoriamente imagens de todas as classes e usando as funções de distância das famílias AID e L_p . Valores de precisão e revocação foram calculados e as médias para todas as classes foram usadas para gerar curvas de revocação *versus* precisão para cada função de distância. As curvas obtidas são mostradas na Figura 7.16.

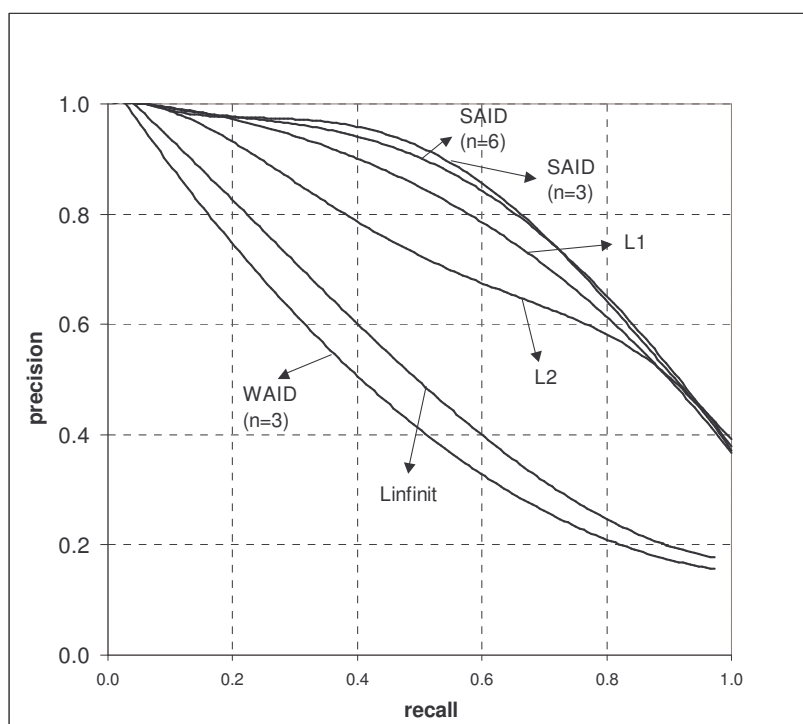


Figura 7.16. Revocação versus precisão para as diferentes funções de distância

Analisando essas curvas de revocação versus precisão, podemos ver que os melhores resultados foram alcançados usando a SAID, o que leva a concluir que trata-se de um contexto caracterizado por interação forte entre atributos. Além disso, SAID com $n=3$ gerou melhores resultados do que com $n=6$. Outros valores de n foram também testados, mas os melhores resultados ocorreram quando utilizamos $n=3$.

7.5. Conclusões

Os resultados apresentados na Seção 7.4 mostram que as famílias de distâncias propostas se aplicam a vários contextos, com diferentes naturezas de descritores para as imagens. Dependendo do contexto, se a interação entre atributos for forte ou, no outro extremo, fraca, as famílias AID são mais eficazes do que a família L_p para a recuperação por similaridade de imagens.

Para que a efetividade dessas funções seja largamente comprovada, outros experimentos necessitam ser realizados e analisados. Porém, os resultados obtidos até aqui são promissores e mostram que estamos diante de uma interessante contribuição para a redução do *gap* semântico que envolve os sistemas de recuperação de imagens por conteúdo.

Capítulo 8

8. CONCLUSÃO

Neste Capítulo apresentamos as principais contribuições e resultados obtidos com o projeto de doutorado. As publicações geradas também são listadas. Discutimos algumas questões que surgiram ao longo do desenvolvimento do projeto e fazemos sugestões de direcionamentos para trabalhos futuros.

8.1. Considerações Iniciais

A proposta inicial desse trabalho foi contribuir para a área de recuperação por conteúdo de imagens médicas, fornecendo subsídios para permitir uma melhor aproximação entre a concepção de similaridade dos métodos comparativos e aquela praticada pelos especialistas da área. A execução desta proposta deu-se através da implementação de métodos comparativos envolvendo os principais descritores relacionados às características intrínsecas das imagens, além de funções de distância relacionadas a um aspecto importante inerente à percepção humana.

8.2. Principais Contribuições e Resultados

Especificamente, foram propostos, implementados e avaliados três métodos comparativos e uma família de funções de distância, assim como algumas ferramentas computacionais para permitir a extração de características, a aplicação dos métodos/funções e a execução de testes com consultas por similaridade. Numa avaliação global, podemos afirmar que os objetivos iniciais que motivaram o seu desenvolvimento foram alcançados.

Resumidamente, podemos descrever essas contribuições, juntamente com os resultados obtidos em cada uma, a saber:

- Um método comparativo para cor, que extrai histogramas métricos das imagens e utiliza a *Global Warp Distance* para compará-los. Esta métrica flexibiliza a comparação local entre os *bins* (ou *buckets*) dos histogramas, aplicando um procedimento que "procura" na vizinhança de cada ponto de um histograma o melhor ajuste para o respectivo ponto do outro histograma. Ela considera também,

na geração do valor da distância, algumas características globais dos histogramas: simetria, curtose e número de picos. O uso da *Global Warp Distance*, para a recuperação por similaridade de imagens médicas gerais, promove uma redução de resultados falsos negativos quando as imagens da base apresentam discrepâncias de brilho e/ou contraste. Além disso, obtém-se também uma redução de resultados falsos positivos quando se busca por imagens pertencentes à mesma categoria anatômica da imagem de referência.

- Um método comparativo para forma, que utiliza momentos de Zernike como descritores. Para permitir que a recuperação das imagens seja realizada de modo concomitante com a classificação das mesmas, o método prevê uma avaliação de momentos relevantes, empregando um algoritmo para mineração de regras de associação, com o qual pode-se identificar quais atributos são relevantes para a classificação. Além disso, o método faz uso de um algoritmo para redução da dimensionalidade, o qual baseia-se na teoria de fractais para determinar a interdependência entre atributos, gerando, assim, um conjunto mínimo de momentos representativos. Os resultados do uso desse método para recuperação e classificação de imagens contendo regiões de interesse de nódulos da mama demonstraram a sua eficácia. Os resultados mostraram também que um conjunto mínimo composto de 16 momentos gera resultados melhores do que um conjunto inicial de 256 momentos, apoiando a asserção de que a inclusão de atributos correlacionados no vetor de características da imagem contribui para a degradação dos resultados de recuperação por conteúdo.
- Um método comparativo para textura, que extrai matrizes de co-ocorrência das imagens em 4 diferentes direções e 5 diferentes vizinhanças, e calcula valores de seis dos descritores propostos por Haralick, assim como de um novo descritor proposto – Histograma de Contraste – e também de uma combinação de descritores. O método prevê um ajuste para compensar possíveis defasagens de rotação entre as imagens comparadas. Avaliações comparativas entre os descritores, almejando a discriminação entre diferentes tecidos do corpo humano, mostra que o Histograma de Contraste gera melhores resultados do que os demais avaliados e também que uma combinação de descritores gera resultados ainda melhores. Resultados de revocação *versus* precisão com o uso dessa combinação mostraram que o método é eficiente para a discriminação de tecidos.
- Uma família de funções de distância (família AID) que permite a escolha da

função de distância ideal para um determinado contexto, de acordo com o grau de influência da interação entre os atributos na percepção de similaridade nesse contexto. Essas funções de distância são de ordem quadrática e podem ser subdivididas em duas outras famílias: a família SAID que produz resultados melhores quando a interação entre atributos é forte e a família WAID que produz resultados melhores quando a interação entre atributos é fraca. A composição dessas duas famílias com a família L_P permite disponibilizar ao usuário um escopo de funções de distância que podem ser testadas e avaliadas para se determinar qual delas melhor se ajusta ao contexto em que estão sendo aplicadas. Resultados obtidos com um conjunto sintético de imagens apresentando variações controladas de cor mostram que, nesse contexto, as funções WAID aproximam-se mais da percepção humana, quando comparadas com as funções SAID e as funções da família L_P . Em outro contexto, porém, os resultados obtidos em experimentos com imagens médicas reais, indicaram que as funções SAID aproximam-se mais da percepção dos radiologistas.

8.3. Publicações

Este projeto de doutorado gerou um conjunto de publicações e apresentações em eventos nacionais e internacionais, onde tivemos a oportunidade de discutir, junto à comunidade científica, a sua forma de abordagem e os resultados obtidos, tendo recebido opiniões, críticas e sugestões extremamente valiosas para o seu desenvolvimento. A relação de publicações é a seguinte:

- Artigos completos em eventos internacionais: [Felipe, J. C. '03], [Felipe, J. C. '05d], [Felipe, J. C. '05c], [Ribeiro, M. X. '05a] e [Felipe, J. C. '06b].
- Resumos em eventos internacionais: [Felipe, J. C. '06a] e [Felipe, J. C. '06c].
- Artigos completos em eventos nacionais: [Felipe, J. C. '02], [Felipe, J. C. '04] e [Felipe, J. C. '05a].
- Resumos em eventos nacionais: [Felipe, J. C. '05b] e [Ribeiro, M. X. '05b].

8.4. Propostas de Trabalhos Futuros

A partir do desenvolvimento realizado e das discussões geradas neste trabalho colocamos, a seguir, a indicação de novos trabalhos que venham complementar ou dar continuidade a este.

Avaliação de Outras Funções de Distância

O universo de funções de distância é amplo. Diversas das funções de distância descritas no Capítulo 3 podem ser usadas junto aos descritores de imagens no processo de comparação. Cabe explorar estas funções junto à busca por similaridade de imagens médicas, em contextos gerais e específicos, a fim de identificar as funções de distância que mais bem se ajustam a eles, para gerar melhores resultados.

Modelagem do Ambiente de Consultas e Análise de Imagens Médicas

Como pudemos observar nos experimentos que envolveram a participação de radiologistas (Seção 4.4.3 e Seção 7.4.2), o consenso nas avaliações de imagens depende fortemente de que haja um objetivo bem definido. Quando os radiologistas eram inquiridos a avaliar a similaridade entre as imagens sem uma definição de objetivos, cada um deles tendia a utilizar um critério diferente. Isto leva a deduzir que a forma de "olhar" a imagem, depende do tipo de imagem e também de um objetivo clínico específico. Portanto, cada diferente tarefa do ambiente de análise deve ter seus próprios parâmetros de representação das imagens e avaliação da similaridade entre elas.

Essa idéia foi confirmada quando exploramos diferentes contextos de análise de imagens para definir a melhor função de distância a ser usada (Seção 7.4). Constatamos que o desempenho obtido com cada função depende estreitamente dos atributos escolhidos e do grau de interação entre os mesmos.

Em outras palavras, o que propomos aqui é um estudo desse ambiente, nos seus diferentes contextos, no sentido de se levantar as diferentes tarefas e os diferentes procedimentos de análise. Uma vez mapeado o ambiente, para cada tarefa e para cada procedimento, as ferramentas de auxílio à análise podem ser customizadas com a definição de extratores específicos, métodos comparativos e funções de distância apropriados para cada possível cenário.

Método de Acesso que Utilize Funções de Distância não Métricas

Durante o levantamento e estudo de funções de distância, pudemos nos deparar com diversas funções de distância que, apesar de não serem métricas, mostraram-se bastante eficientes para a recuperação de imagens, em variadas situações. Especificamente quando se buscava por funções de distância que se aproximassem da percepção humana, muitas vezes as funções resultantes desobedeciam a propriedade de simetria e, principalmente, a propriedade da desigualdade triangular. No caso das funções de distância propostas neste trabalho não foi diferente, uma vez que a família de funções SAID não obedece à desigualdade triangular.

Com base nessas ocorrências, acabamos por repensar nossa visão inicial de trabalho que resumia-se em buscar por funções de distância que fossem obrigatoriamente métricas, para que elas pudessem ser utilizadas de forma otimizada pelos métodos de acesso métricos, nos processos de indexação em bases de dados. Encontrar funções de distância semanticamente eficientes não é uma tarefa fácil. Por isso, pensamos que funções não métricas que sejam perceptualmente eficazes não devem ser descartadas. O outro lado da questão, ou seja, das estruturas de indexação, é que deve ser reconsiderado.

Propomos aqui um desafio para os pesquisadores da área de bancos de dados, que trabalham com aperfeiçoamento de métodos de acesso: encontrar novas formas de poda das árvores de busca, que prescindam da exigência da propriedade triangular para a sua correta execução. Com essa evolução, os sistemas de bases de dados poderão usufruir do uso de funções de distância muito mais ricas e eficientes na redução do *gap* semântico.

Novas Análises com as Funções das Famílias SAID e WAID

Os experimentos realizados com as famílias de funções propostas SAID e WAID foram de caráter preliminar. Novas análises perceptuais precisam ser realizadas com o objetivo de quantificar como o *gap* semântico é afetado por tais distâncias. A grande dificuldade encontrada nesses experimentos reside no fato de que, com imagens reais, torna-se difícil isolar atributos para poder avaliar a sua influência. O caminho inicial para isso é a escolha criteriosa de conjuntos de imagens que apresentem alguns atributos (sejam de textura, de cor ou de forma) bem definidos, de modo que a sua influência possa ser controlada nos experimentos.

Avaliação das famílias SAID e WAID junto a Métodos de Acesso Métricos

A família WAID, por ser métrica, permite sua utilização em métodos de acesso métricos existentes. A família SAID, que não é métrica, necessita da definição de um limite inferior (*lower bound*) para poder ser utilizada nessas estruturas.

Em ambos os casos, faz-se necessária a realização de testes para verificar o desempenho desses métodos com essas funções, comparados com as funções atualmente implementadas.

Estudo da Integração entre Dados Pictóricos e Textos Associados às Imagens

Uma vez que existe uma quantidade razoável de informações textuais (oriundas de prontuários, exames, cabeçalhos DICOM, etc.) associadas às imagens médicas, parece muito promissor o estudo da integração entre esses dois tipos de informações. Já existem trabalhos voltados a essa tarefa, considerando-se imagens gerais. No caso das imagens médicas, há uma série de peculiaridades que podem potencializar os efeitos dessa integração, como, por exemplo, o uso de padrões de classificação de doenças.

Os progressos apresentados por pesquisas na área de Recuperação da Informação, no sentido de determinar ligações automáticas entre documentos, podem vir a ser utilizados no ambiente radiológico considerando-se o tipo de dado "imagem".

Mineração de Imagens

Assim como para a determinação de ligações automáticas envolvendo arquivos de imagens, outra área que pode ser explorada com sucesso é a de mineração de dados e imagens. A aplicação de técnicas de mineração de dados a dados textuais concomitantemente com dados pictóricos deve vir a revelar conjuntos de regras mais precisas a respeito de paciente, imagens e laudos.

Bases de Imagens Médicas Compartilhadas

A maior dificuldade encontrada para a realização deste trabalho foi conseguir conjuntos adequados de imagens para a execução dos experimentos. Apesar de podermos contar com o apoio incondicional do CCIFM, que permitiu nosso acesso à sua base de imagens, a grande dificuldade reside no fato de que as imagens são organizadas de acordo com os prontuários, de modo que a sua separação em grupos funcionais ou por ocorrência de anomalias tem que ser feita manualmente. Sabendo que a validação de métodos demanda grandes volumes de imagens (da ordem de milhares, de preferência), o processo de separação manual acaba por tornar-se inviável.

Considerando que esse problema é compartilhado por toda a comunidade de pesquisa com imagens médicas, do nosso ponto de vista, a solução para o problema é a formação de uma rede compartilhada de imagens, da qual participariam os diversos centros de pesquisa na área. A idéia não é nova. O INCOR tem proposto uma estrutura para compartilhamento de imagens na internet. Uma vez colocada em prática esta estrutura, resta aos centros de pesquisa organizarem e disponibilizarem suas bases de imagens. É uma tarefa árdua, porém, quando estiver funcional, os benefícios serão compartilhados e toda a comunidade poderá usufruir de bases extensas e diversificadas, que nos permitiriam validar, com maior propriedade, os novos métodos propostos, assim como realizar comparações efetivas com outros métodos existentes.

8.5. Considerações Finais

Chegamos ao final deste projeto, certos de que os esforços despendidos para a sua concretização geraram alguns frutos. Grande parte do que nos propusemos a desenvolver pôde ser levado a cabo. Nosso desejo é que as Ciências Médicas possam se beneficiar com esta pequena contribuição das Ciências da Computação, de maneira que, no final, o grande beneficiado seja a sociedade e, claro, o ser humano.

Referências Bibliográficas

REFERÊNCIAS BIBLIOGRÁFICAS

- [ACR '05] ACR. Breast Imaging Reporting and Data System (BI-RADS). <http://www.birads.at>. Acessado em 10/10/2005. 2005.
- [Aggarwal, C. C. '01] Aggarwal, C. C., A. Hinneburg e D. A. Keim. On the Surprising Behavior of Distance Metrics in High Dimensional Spaces. 8th International Conference on Database Theory (ICDT 2001), London, England. 420-434. 2001.
- [Aggarwal, C. C. '03] Aggarwal, C. C. Towards Systematic Design of Distance Functions for Data Mining Applications. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003), Washington, DC, USA. 9-18. 2003.
- [Akleman, E. '99] Akleman, E. e J. Chen. Generalized Distance Functions. Intl. Conf. on Shape Modeling and Applications, Aizu-Wakamatsu. 1999.
- [Aksoy, S. '98] Aksoy, S. e R. M. Haralick. Textural Features for Image Database Retrieval. IEEE Workshop on Content-Based Access of Image and Video Libraries, Santa Barbara, California. pp. 45-49. 1998.
- [Alferez, R. '05] Alferez, R., Y. F. Wang e L. Jiao. "An Affine-Invariant Tool for Retrieving Images from Homogeneous Databases." Multimedia Tools and Applications **25**(1): 133-159. 2005.
- [Alvarenga, A. V. '02] Alvarenga, A. V., A. F. C. Infantosi, C. M. d. Azevedo, et al. Contour Detection of Breast Ultrasound Tumor Images Using Morphological Operators. 12th Nordic Baltic Conference on Biomedical Engineering and Medical Physics, Reykjavik, Iceland. 78-79. 2002.
- [Andre, T. C. S. S. '02] Andre, T. C. S. S. e A. C. R. Silva-Filho. Segmentation of Digitized Mammograms Using Self-Organizing Maps in a Breast Cancer Computer Aided Diagnosis System. 7th Brazilian Symposium on Neural Networks (SBRN 2002), Recife, Brazil, IEEE Computer Society. 11-14. 2002.
- [Antani, S. K. '03] Antani, S. K., L. R. Long, G. R. Thoma, et al. Evaluation of Shape Indexing Methods for Content-Based Retrieval of X-Ray Images. 15th Annual Symposium on Electronic Imaging, Storage and Retrieval for Media Databases. 2003.
- [Antani, S. K. '04] Antani, S. K., L. R. Long e G. R. Thoma. Content-Based Image Retrieval for Large Biomedical Image Archives. 11th World Congress on Medical Informatics (MEDINFO), San Francisco, CA, USA. 829-833. 2004.
- [Antani, S. K. '05] Antani, S. K., M. Natarajan, J. L. Long, et al. "Developing a comprehensive system for content-based image retrieval of image and text from a national survey." SPIE Medical Imaging **5748**: 152-161. 2005.
- [Araujo, M. R. B. '02] Araujo, M. R. B. Extending SQL to Support Image Content-based Retrieval. IASTED International Conference Information Systems and Databases (ISDB 2002), Tokyo, Japan. pp. 19-24. 2002.
- [Aslandogan, Y. A. '99] Aslandogan, Y. A. e C. T. Yu. "Techniques and Systems for Image and Video Retrieval." IEEE Transactions on Knowledge and Data Engineering **11**(1): 56-63. 1999.
- [Aumann, Y. '99] Aumann, Y. e Y. Lindell. A Statistical Theory for Quantitative Association Rules. Fifth SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, California. 1999.

- [Bach, J. R. '96] Bach, J. R., C. Fuller, A. Gupta, et al. Virage Image Search Engine: an Open Framework for Image Management. SPIE Conference on Storage and Retrieval for Image and Video Databases IV. pp. 76-87. 1996.
- [Balan, A. G. R. '03] Balan, A. G. R. Técnicas de Segmentação de Imagens Aéreas para Contagem de População de Aves. Dissertação de Mestrado. Instituto de Ciências Matemáticas e de Computação - ICMC, Universidade de São Paulo. São Carlos, SP, Brazil. 2003.
- [Balan, A. G. R. '05] Balan, A. G. R., A. J. M. Traina, C. T. Jr., et al. Fractal Analysis of Image Textures for Indexing and Retrieval by Content. 18th IEEE Symposium on Computer-Based Medical Systems (CBMS), Dublin, Ireland. 581-586. 2005.
- [Baucher, S. '91] Baucher, S. The watershed transformation applied to image segmentation. 10th Pfefferkorn Conf. on Signal and Image Processing in Microscopy and Microanalysis, Cambridge, UK. 299-314. 1991.
- [Beckmann, N. '90] Beckmann, N., H.-P. Kriegel, R. Schneider, et al. The R*-tree: An Efficient and Robust Access Method for Points and Rectangles. ACM Int'l Conference on Data Management (SIGMOD). 322-331. 1990.
- [Beucher, S. '91] Beucher, S. The watershed transformation applied to image segmentation. 10th Pfefferkorn Conf. on Signal and Image Processing in Microscopy and Microanalysis, Cambridge, UK. 299-314. 1991.
- [Beyer, K. '99] Beyer, K., J. Godstein, R. Ramakrishnan, et al. When is "Nearest Neighbor" Meaningful? International Conference on Database Theory (ICDT), Jerusalem, Israel. 217-235. 1999.
- [Boukerroui, D. '99] Boukerroui, D., O. Basset, A. Baskurt, et al. Segmentation of Echocardiographic Data. Multiresolution 2D and 3D Algorithm Based on Grey Level Statistics. Medical Image Computing and Computer-Assisted Intervention - MICCAI'99, Second International Conference, Cambridge, UK, Springer-Verlage. pp. 516-523. 1999.
- [BRODATZ '05] BRODATZ. Brodatz Textures. <http://sipi.usc.edu/services/database/database.cgi?volume=textures>. Acessado em 17/10/2005. 2005.
- [Bueno, J. M. '02a] Bueno, J. M., A. J. M. Traina, C. Traina Jr., et al. cbPACS: PACS com Suporte à Recuperação de Imagens Médicas Baseada em Conteúdo. VIII Congresso Brasileiro de Informática em Saúde, Natal - Brazil. 271-276. 2002a.
- [Bueno, J. M. '02b] Bueno, J. M. Suporte à Recuperação de Imagens Médicas Baseada em Conteúdo Através de Histogramas Métricos. Tese de doutorado. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. São Carlos - Brasil. 2002b.
- [Bueno, J. M. '02c] Bueno, J. M., F. Chino, A. J. M. Traina, et al. How to Add Content-based Image Retrieval Capability in a PACS. IEEE International Conference on Computer Based Medical Systems - CBMS, Maribor, Slovenia. pp. 321-326. 2002c.
- [Burkhard, W. A. '73] Burkhard, W. A. e R. M. Keller. "Some Approaches to Best-Match File Searching." Communications of the ACM **16**(4): 230-236. 1973.
- [Bushberg, J. T. '02] Bushberg, J. T., J. A. Seibert, E. M. Leidholdt-Jr, et al. The Essential Physics of Medical Imaging. Second Edition Philadelphia, PA, USA, Lippincott Williams & Wilkins. 2002.
- [Cabena, P. '97] Cabena, P., P. Hadjinian, R. Stadler, et al. Discovering Data Mining - From Concept to Implementation New Jersey, USA, Prentice Hall PTR. 1997.
- [Carson, C. '02] Carson, C., S. Belongie, H. Greenspan, et al. BlobWorld: Image Segmentation Using Expectation-Maximization and Its Application to Image

- Querying. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. **24**: pp. 509-516. 2002.
- [Castanón, C. A. B. '03] Castanón, C. A. B. Extração de Características em Imagens Médicas Utilizando Wavelets. Dissertação de mestrado. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. São Carlos - Brasil. 92 pp. 2003.
- [Castleman, R. '96] Castleman, R. Digital Image Processing, Prentice-Hall Inc. 1996.
- [Celebi, E. '00] Celebi, E. e A. Alpkocak. Clustering of Texture Features for Content-Based Image Retrieval. ADVIS - Advances in Information Systems, Izmir, Turkey, Springer-Verlag. pp. 216-225. 2000.
- [Chang, E. Y. '00] Chang, E. Y., B. Li e C. Li. Toward Preception-Based Image Retrieval. IEEE Workshop on Content-Based Access of Image and Video Libraries, Hilton Head - USA. pp. 101-105. 2000.
- [Chávez, E. '01] Chávez, E., G. Navarro, R. Baeza-Yates, et al. "Searching in Metric Spaces." ACM Computing Surveys. 2001.
- [Ciaccia, P. '97] Ciaccia, P., M. Patella e P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. Intl. Conf. on Very Large Databases (VLDB), Athens, Greece. pp. 426-435. 1997.
- [Cohen, S. '99] Cohen, S. Finding Color and Shape Patterns in Images. PhD thesis. Stanford University. 1999.
- [Comer, D. '79] Comer, D. "The Ubiquitous B-Tree." ACM Computing Surveys **11**(2): pp. 121-137. 1979.
- [Coppini, G. '03] Coppini, G., S. Diciotti, M. Falchini, et al. "Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms." IEEE Transactions on Information Technology in Biomedicine **7**(4): 344-357. 2003.
- [Corner, M. L. '00] Corner, M. L. e E. J. Delp. "The EM/MPM Algorithm for Segmentation of Textured Imagens: Analysis and Further Experimental Results." IEEE Trans. Image Processing **9**(10): 1731-1744. 2000.
- [Costa, L. d. F. '00] Costa, L. d. F. e R. M. Cesar. Shape Analysis and Classification. 1st editon, CRC Press. 2000.
- [DBLP '05] DBLP. Database Systems and Logic Programming. <http://dblp.uni-trier.de/>. acessado in September, 2005. 2005.
- [Ding, S. '04] Ding, S., J. Yang e Q. Xu. Color Texture Analysis Using Wavelet-Based Hidden Markov Model. Australian Conference on Artificial Intelligence, Cairns, Australia. 1043-1048. 2004.
- [Do, M. N. '02] Do, M. N. e M. Vetterli. "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance." IEEE Transactions on Image Processing **11**(2): pp. 146-158. 2002.
- [Duncan, J. S. '00] Duncan, J. S. e N. Ayache. "Medical Image Analysis: Progress over Two Decades and the Challenges Ahead." IEEE PAMI **22**(1): pp. 85-105. 2000.
- [El-Naqa, I. '02] El-Naqa, I., Y. Yang, N. P. Galatsanos, et al. Content-based image retrieval for digital mammography. International Conference on Image Processing, New York, NY. 141-144. 2002.
- [Emine Krichen, L. A. '05] Emine Krichen, L. A., S. Garcia-Salicetti e B. Dorizzi. Specific Texture Analysis for Iris Recognition. Audio- and Video-Based Biometric Person Authentication: 5th International Conference (AVBPA), Hilton Rye Town, NY, USA. 23-30. 2005.
- [Faloutsos, C. '96] Faloutsos, C. Searching Multimedia Databases by Content Boston, MA, Kluwer Academic Publishers. 1996.
- [Faloutsos, C. '05] Faloutsos, C. Tutorial: Mining Time Series Data. 20th Brazilian Symposium on Databases, Uberlândia, MG, Brazil. 2005.

- [Feder, J. '96] Feder, J. "Image Recognition and Content-Based Retrieval for the World Wide Web." Advance Imaging **11(1)**: pp. 26-28. 1996.
- [Felipe, J. C. '02] Felipe, J. C. e A. J. M. Traina. Utilizando Características de Textura para Identificação de Tecidos em Imagens Médicas. II Workshop de Informática Médica (WIM 2002), Gramado - Brasil. 2002.
- [Felipe, J. C. '03] Felipe, J. C., A. J. M. Traina e C. Traina-Jr. Retrieval by Content of Medical Images Using Texture for Tissue Identification. 16th IEEE Symposium on Computer-Based Medical Systems (CBMS 2003), New York - NY. 175-180. 2003.
- [Felipe, J. C. '04] Felipe, J. C. e A. J. M. Traina. Warp Metric Distance: Aprimorando o Uso de Histogramas de Intensidade para Busca por Similaridade em Imagens Médicas. IX Congresso Brasileiro de Informática em Saúde (CBIS 2004), Ribeirão Preto, SP, Brazil. 2004.
- [Felipe, J. C. '05a] Felipe, J. C., A. J. M. Traina, C. Traina-Jr, et al. Discriminação de Aspectos Malignos em Massas Tumoriais de Mamografias Usando Características de Forma das Imagens. V Workshop de Informática Médica (WIM 2005), Porto Alegre, RS, Brazil. 2005a.
- [Felipe, J. C. '05b] Felipe, J. C. e A. J. M. Traina. Uma Nova Medida de Similaridade entre Histogramas Aplicada à Busca por Conteúdo em Imagens Médicas. II Simpósio de Instrumentação e Imagens Médicas (SIIM 2005), São Pedro, SP, Brazil. 2005b.
- [Felipe, J. C. '05c] Felipe, J. C., A. J. M. Traina e C. Traina-Jr. Global Warp Metric Distance: Boosting Content-based Image Retrieval through Histograms. IEEE International Symposium on Multimedia (ISM 2005), Irvine, CA, USA. (to appear). 2005c.
- [Felipe, J. C. '05d] Felipe, J. C., J. B. Olioti, M. X. Ribeiro, et al. A Low-cost Approach for Effective Shape-based Retrieval and Classification of Medical Images. The First IEEE International Workshop on Multimedia Information Processing and Retrieval (MIPR 2005), Irvine, California, USA. (to appear). 2005d.
- [Felipe, J. C. '06a] Felipe, J. C., A. J. M. Traina, A. G. R. Balan, et al. A New Family of Perceptual Distance Functions for Similarity Retrieval of Images. The 21st Annual ACM Symposium on Applied Computing (SAC 2006), Dijon, France. (to appear). 2006a.
- [Felipe, J. C. '06b] Felipe, J. C., A. J. M. Traina, C. Traina-Jr, et al. Effective Shape-based Retrieval and Classification of Mammograms. The 21st Annual ACM Symposium on Applied Computing (SAC 2006), Dijon, France. (to appear). 2006b.
- [Felipe, J. C. '06c] Felipe, J. C., A. J. M. Traina e C. Traina-Jr. A New Similarity Measure for Histograms Applied to Content-based Retrieval of Medical Images. The 21st Annual ACM Symposium on Applied Computing (SAC 2006), Dijon, France. (to appear). 2006c.
- [Figueiredo, J. C. B. '01a] Figueiredo, J. C. B. e S. S. Furuie. Statistical distances to detect changes in the normal behavior of ECG-Holter signals. SPIE - Medical Imaging: Ultrasonic Imaging and Signal Processing, San Diego, CA, USA, SPIE PRESS. 2001a.
- [Figueiredo, J. C. B. '01b] Figueiredo, J. C. B., S. S. Furuie, M. A. Gutierrez, et al. PACS-based model for telemedicine. SPIE - Medical Imaging 2001: PACS and Integrated Medical Information Systems: Design and Evaluation, San Diego, CA, USA. 2001b.
- [Figueiredo, J. M. '05] Figueiredo, J. M. Formalização do domínio imagem para buscas por conteúdo em SGBDs relacionais. Tese de Doutorado. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. São Carlos. 125 p. 2005.
- [Foley, J. D. '95] Foley, J. D., A. v. Dam, S. K. Feiner, et al. Computer Graphics: Principles and Practice in C. 2nd edition, Addison-Wesley Pub Co. 1995.

- [Furuie, S. S. '99] Furuie, S. S., M. A. Gutierrez, N. B. Bertozzo, et al. "Archiving and Retrieving Long-Term Cineangiographic Images in a PACS." Computers in Radiology **26**: 435-438. 1999.
- [Gaede, V. '98] Gaede, V. e O. Günther. "Multidimensional Access Methods." ACM Computing Surveys **30**(2): pp. 170-231. 1998.
- [Gibbs, A. L. '02] Gibbs, A. L. e F. E. Su. "On Choosing and Bounding Probability Metrics." International Statistical Review **70**(3): 419-435. 2002.
- [Giger, M. L. '01] Giger, M. L., N. Karssemeijer e S. G. Armato-III. "Computer-Aided Diagnosis in Medical Imaging." IEEE Trans. Med. Imaging **20**(12): 1205-1208. 2001.
- [Gonzales, R. C. '87] Gonzales, R. C. e P. Wintz. Digital Image Processing. 2nd, Adison-Wesley Publishing Company. 1987.
- [Gusfield, D. '99] Gusfield, D. Algorithms on Strings, Trees and Sequences New York, NY, USA, Cambridge University Press. 1999.
- [Guttman, A. '84] Guttman, A. R-Tree : A dynamic Index Structure for Spatial Searching. ACM Int'l Conference on Data Management (SIGMOD), Boston, MA, ACM Press. pp. 47-57. 1984.
- [Hand, D. J. '01] Hand, D. J., H. Mannila e P. Smyth. Principles of Data Mining Cambridge, MA, MIT Press. 2001.
- [Haralick, R. M. '73] Haralick, R. M., K. Shanmugan e I. Dinstein. "Textural Fetures for Image Classification." IEEE Transactions on Systems, Man and Cybernetics **3**(6): pp. 610-621. 1973.
- [Hellerstein, J. M. '95] Hellerstein, J. M., J. F. Naughton e A. Pfeffer. Generalized Search Trees for Database Systems. Intl. Conf. on Very Large Databases (VLDB), Zurich, Switzerland, Morgan Kaufmann. 562-573. 1995.
- [Hertz, T. '04] Hertz, T., A. Bar-Hillel e D. Weinshall. Learning Distance Functions for Image Retrieval. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), Washington, DC, USA. 570-577. 2004.
- [Hiransakolwong, N. '05] Hiransakolwong, N., K. A. Hua, S. Koompaiojn, et al. An Adaptive Distance Computation Technique for Image Retrieval Systems. ACM Symposium on Applied Computing, Santa Fe, NM, USA. 1195-1199. 2005.
- [Hjaltson, G. '03] Hjaltson, G. e H. Samet. "Index-Driven Similarity Search in Metric Spaces." ACM Transactions On Database Systems **28**(4): 517-580. 2003.
- [Hludov, S. '99] Hludov, S., C. Meinel, G. Noelle, et al. PACS for Teleradiology. 12th IEEE Symposium on Computer-Based Medical Systems (CBMS'99), Stamford, CT, USA. 6-11. 1999.
- [Honda, M. O. '01] Honda, M. O. Elaboração de uma Base de Conhecimentos para Auxílio ao Diagnóstico através da Comparação Visual de Imagens Mamográficas. Dissertação de mestrado. Escola de Engenharia de São Carlos, São Paulo University. São Carlos - Brasil. 2001.
- [Iakovidis, D. K. '05] Iakovidis, D. K., D. E. Maroulis, S. A. Karkanis, et al. A Comparative Study of Texture Features for the Discrimination of Gastric Polyps in Endoscopic Video. 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05), Dublin, Ireland. 575-580. 2005.
- [IARL '05] IARL. Digital Database for Screening Mammography. <http://marathon.csee.usf.edu/Mammography/>. acessado em 13/10/2005. 2005.
- [Inoue, M. '05] Inoue, M. e N. Ueda. Retrieving Lightly Annotated Images using Image Similarities. ACM Symposium on Applied Computing, Santa Fé - USA. 1031-1037. 2005.

- [Jacobs, D. W. '00] Jacobs, D. W., D. Weinshall e Y. Gdalyahu. "Classification with Nonmetric Distances: Image Retrieval and Class Representation." IEEE Trans. Pattern Analysis and Machine Intelligence **22**(6): 583-600. 2000.
- [Joo, S. '04] Joo, S., Y. S. Yang, W. K. Moon, et al. "Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features." IEEE Trans. Med. Imaging **23**(10): 1292-1300. 2004.
- [Kak, A. '02] Kak, A. e C. Pavlopoulou. Content-Based Image Retrieval from Large Medical Databases. First International Symposium on 3D Data Processing Visualization and Transmission, Padova, Italy. pp. 138-147. 2002.
- [Kalinin, M. '05] Kalinin, M., D. S. Raicu, J. D. Furst, et al. A Classification Approach for Anatomical Regions Segmentation. IEEE International Conference on Image Processing (ICIP), Genoa, Italy. 2005.
- [Kan, C. '01] Kan, C. e M. D. Srinath. Combined Features of Cubic B-Spline Wavelet Moments and Zernike Moments for Invariant Character Recognition. IEEE International Conference on Information Technology: Coding and Computing (ITCC'01), Las Vegas, NV. 2001.
- [Kao, O. '00] Kao, O. e I. I. Tendresse. CLIMS - A System for Image Retrieval by Using Colour and Wavelet Features. ADVIS, LNCS 1909. pp. 238-247. 2000.
- [Keogh, E. '01a] Keogh, E. A Tutorial on Indexing and Mining Time Series Data. IEEE International Conference on Data Mining, San Jose - USA. 2001a.
- [Keogh, E. '01b] Keogh, E. e M. J. Pazzani. Derivative Dynamic Time Warping. First SIAM International Conference on Data Mining, Chicago - USA. 2001b.
- [Keogh, E. '02a] Keogh, E. On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. ACM SIGKDD '02, Alberta - Canada. 2002a.
- [Keogh, E. '02b] Keogh, E. Exact Indexing of Dynamic Time Warping. 28th VLDB Conference, Hong Kong - China. 2002b.
- [Kohavi, R. '97] Kohavi, R., D. Sommerfield e J. Dougherty. "Data Mining Using MLC++ - A Machine Learning Library in C++." International Journal on Artificial Intelligence Tools **6**(4): 537-566. 1997.
- [Kom, F. '01] Kom, F., B.-U. Pagel e C. Faloutsos. On the "Dimensionality Curse" and the "Self-Similarity" Blessing. IEEE Transactions on Knowledge and Data Engineering. **vol.13**: pp. 96-101. 2001.
- [Korn, P. '98] Korn, P., N. Sidiropoulos, C. Faloutsos, et al. "Fast and Effective Retrieval of Medical Tumor Shapes." IEEE Transactions on Knowledge and Data Engineering **10**(6): 889-904. 1998.
- [Kurani, A. '04] Kurani, A., D. H. Xu, J. D. Furst, et al. Co-occurrence matrices for volumetric data. 7th IASTED International Conference on Computer Graphics and Imaging - CGIM 2004, Kauai, Hawaii, USA. 2004.
- [Lado, M. J. '01] Lado, M. J., A. J. Méndez, P. G. Tahoces, et al. Computer-Aided Diagnosis: Application of Wavelet Transform to the Detection of Clustered Microcalcifications in Digital Mammograms. 2nd International Symposium on Medical Data Analysis, Madrid - Spain. pp. 140-145. 2001.
- [Lee, C.-S. '99] Lee, C.-S., H.-J. Tschai, Y.-H. Kuo, et al. PACS: Construction and Application to Medical Image Enhancemet. Third International Conference on Knowledge-Based Intelligent Information Engineering Systems. pp. 246-249. 1999.
- [Lehmann, T. M. '03a] Lehmann, T. M., B. B. Wein e H. Greenspan. Integration of Content-based Image Retrieval to Picture Archiving and Communication Systems. Medical Informatics Europe (MIE 2003), St Malo, France. CD-ROM. 2003a.

- [Lehmann, T. M. '03b] Lehmann, T. M., M. O. Güld, C. Thies, et al. Content-based Image Retrieval in Medical Applications for Picture Archiving and Communication Systems. SPIE 2003. 440-451. 2003b.
- [Lehmann, T. M. '05] Lehmann, T. M., D. Beier, C. Thies, et al. Segmentation of Medical Images Combining Local, Regional, Global and Hierarchical Distances into a Bottom-up Region Merging Scheme. SPIE 2005. 546-555. 2005.
- [Leite, N. J. '01] Leite, N. J. Introdução ao Processamento e Análise de Imagens. Campinas, SP, IC - UNICAMP: 29. 2001.
- [Li, H. '04] Li, H., M. L. Giger, O. I. Olopade, et al. Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms. 18th International Congress of Computer Assisted Radiology and Surgery (CARS 2004), Chicago, USA. 878-881. 2004.
- [Li, J. '00] Li, J., J. Z. Wang e G. Wiederhold. IRM: Integrated Region Matching for Image Retrieval. ACM Multimedia International Conference. pp. 147-156. 2000.
- [Li, Q. '03] Li, Q., F. Li, J. Shiraishi, et al. "Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules." Medical Physics **30**(10): 2584-2593. 2003.
- [M. Heath K. Bowyer, D. K. '00] M. Heath K. Bowyer, D. K., R. Moore e P. K. Jr. The Digital Database for Screening Mammography. 5th International Workshop on Digital Mammography, Toronto, Canada, Medical Physics Publishing (Madison, WI). 2000.
- [Manjunath, B. S. '96] Manjunath, B. S. e W. Y. Ma. "Texture Features for Browsing and Retrieval of Image Data." IEEE Transactions on Patterns Analysis and Machine Intelligence **18**(8): pp. 837-842. 1996.
- [Marques, P. M. A. '02] Marques, P. M. A., M. H. Honda, J. A. H. Rodrigues, et al. "Recuperação de Imagens Baseada em Conteúdo: Uso de Atributos de Textura para Caracterização de Microcalcificações Mamográficas." Radiologia Brasileira **35**(2): 93-98. 2002.
- [Marques, P. M. d. A. '00] Marques, P. M. d. A., A. C. Santos, J. E. Júnior, et al. "Implantação de um Sistema de Informação em Radiologia em Hospital Univeristário." Radiologia Brasileira **33**: pp. 155-160. 2000.
- [Marsh, A. '97] Marsh, A. EUROMED - The Creation of a Telemedical Information Society. 10' IEEE Symposium on Computer Based Medical Systems, Maribor, Slovenia. pp. 86-91. 1997.
- [Mlsna, P. A. '04] Mlsna, P. A. e N. M. Sirakov. Intelligent Shape Feature Extraction and Indexing for Efficient Content-Based Medical Image Retrieval. IEEE Southwest Symposium on Image Analysis and Interpretation, Lake Tahoe, Nevada, USA. 2004.
- [Morioka, C. A. '05] Morioka, C. A., S. El-Saden, W. Pope, et al. Integration of HIS/RIS clinical document with PACS image studies for neuroradiology. SPIE International Symposium on Medical Imaging, San Diego, CA, USA. 2005.
- [Mudigonda, N. '00] Mudigonda, N., R. Rangayyan e J. Desautels. "Gradient and Texture Analysis for the Classification of Mammographic Masses." IEEE Transactions on Medical Imaging **19**(10): pp. 1032-1043. 2000.
- [Müller, H. '04] Müller, H., N. Michoux, D. Bandon, et al. "A review of content-based image retrieval systems in medical applications - clinical menefits and future directions." International Journal of Medical Informatics **73**: 1 - 23. 2004.
- [NEMA '03] NEMA. DICOM Standard. <http://medical.nema.org/>. acessado em 11/11/2005. 2003.
- [Niblack, W. '93] Niblack, W., R. Barber, W. Equitz, et al. Querying Images by Content, Using Color, Texture and Shape. SPIE Conference on Storage and Retrieval for Image and Video Databases. pg 173-187. 1993.

- [Ogle, V. E. '95] Ogle, V. E. e M. Stonebraker. "Chabot: Retrieval from a Relational Database of Images." Computer: pp. 40-48. 1995.
- [Panchal, R. '04] Panchal, R. e B. Verma. A Fusion of Neural Network Based Auto-associator and Classifier for the Classification of Microcalcification Patterns. 11th International Conference on Neural Information Processing (ICONIP'04), India. 794-799. 2004.
- [Pentland, A. '96] Pentland, A., R. W. Picard e S. Sclaroff. "Photobook: Content-Based Manipulation of Image Databases." International Journal of Computer Vision **18**(3): pp. 233-254. 1996.
- [Qamra, A. '05] Qamra, A., Y. Meng e E. Y. Chang. "Enhanced Perceptual Distance Functions and Indexing for Image Replica Recognition." IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **27**(3): 379-391. 2005.
- [Qian, G. '04] Qian, G., S. Sural, Y. Gu, et al. Similarity Between Euclidean and Cosine Angle Distance for Nearest Neighbor Queries. ACM Symposium on Applied Computing, Nicosia, Cyprus. 1232-1237. 2004.
- [Quinlan, J. R. '86] Quinlan, J. R. Induction of Decision Trees. Machine Learning Magazine. **1**(1): 81-106. 1986.
- [Ribeiro, M. X. '05a] Ribeiro, M. X., J. C. Felipe, A. G. R. Balan, et al. Mining Statistical Association Rules to Select the Most Relevant Medical Image Features. The First International Workshop on Mining Complex Data (MCD 2005) in conjunction with The Fifth IEEE International Conference on Data Mining (ICDM 2005), Houston, Texas, USA. (to appear). 2005a.
- [Ribeiro, M. X. '05b] Ribeiro, M. X., J. C. Felipe e A. J. M. Traina. Seleção de Atributos Relevantes para Busca por Similaridade e Classificação de Imagens Médicas. I Workshop de Visão Computacional, Piracicaba, SP, Brasil. 2005b.
- [Rosa, N. A. '02] Rosa, N. A., R. F. Santos, J. M. Bueno, et al. Sistema de Recuperação de Imagens Similares em um Hospital Universitário. VIII Congresso Brasileiro de Informática em Saúde, Natal - RN - Brazil. pp. 613-618. 2002.
- [Rosh, E. '75] Rosh, E. "Cognitive Reference Points." Cognitive Psychology **7**: pp. 532-547. 1975.
- [Rubner, Y. '01] Rubner, Y. e C. Tomasi. Perceptual Metrics for Image Database Navigation Boston - USA, Kluwer Academic Publishers. 2001.
- [Rui, Y. '97] Rui, Y., T. Huang e S. Mehrotra. Content-Based Image Retrieval with Relevance Feedback in Mars. IEEE International Conference on Image Processing. 1997.
- [Russel, S. J. '03] Russel, S. J. e P. Norvig. Artificial Intelligence - A Modern Approach. 2nd editon, Prentice Hall. 2003.
- [Saccavini, C. '01] Saccavini, C., P. Mosca, R. Stramare, et al. Low Cost PACS system: An Opensource Project Approach. 13th European Congress of Radiology, Vienna, Austria. 2001.
- [Samet, H. '95] Samet, H. Spatial Data Structures in Modern Database Systems: The Object Model, Interoperability and Beyond. W. Kim, Ed. Addison-Wesley/ACM Press: pp. 361-385. 1995.
- [Sclaroff, S. '94] Sclaroff, S. e A. P. Pentland. On Modal Modeling for Medical Images: Underconstrained Shape Description and Data Compression. IEEE Workshop on Biomedical Image Analysis, Seattle, WA, USA. 70-79. 1994.
- [Sclaroff, S. '97] Sclaroff, S., L. Taycher e M. L. Cascia. ImageRover: A Content-Based Image Browser for the World Wide Web. IEEE Workshop on Content-Based Access of Image and Video Libraries. 1997.

- [Semler, L. '05] Semler, L., L. Dettori e J. Furst. Wavelet-Based Texture Classification of Tissues in Computed Tomography. 18th IEEE International Symposium on Computer-Based Medical Systems(CBMS'05), Dublin, Ireland. 265-270. 2005.
- [Shasha, D. '90] Shasha, D. e T. L. Wang. "New techniques for best-match retrieval." ACM Transactions on Information Systems **8**(2): pp. 140-158. 1990.
- [Shiraishi, J. '03] Shiraishi, J., H. Abe, R. Engelmann, et al. "Computer-aided Diagnosis for Distinction between Benign and Malignant Solitary Pulmonary Nodules in Chest Radiographs." Radiology **227**: 469-474. 2003.
- [Siegel, E. L. '99] Siegel, E. L. "Current State of the Art and Future Trends." Filmless Radiology: pp. 3-20. 1999.
- [Smeulders, A. W. M. '00] Smeulders, A. W. M., M. Worring, S. Santini, et al. "Content-Based Image Retrieval at the End of the Early Years." IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(12). 2000.
- [Smith, J. R. '97] Smith, J. R. Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression. PhD thesis. Columbia University. 1997.
- [Smith, S. W. '99] Smith, S. W. The Scientist and Engineer's Guide to Digital Signal Processing, California Technical Publishing. 1999.
- [Sonka, M. '99] Sonka, M., V. Hlavac e R. Boyle. Image Processing, Analysis, and Machine Vision. 2nd ed Pacific Grove, CA, USA, Brooks/Cole Publishing Company. 1999.
- [Sorid, D. '00] Sorid, D. e S. K. Moore. The Virtual Surgeon. IEEE Spectrum. **July 2000**: pp. 26-31. 2000.
- [Souza, E. P. M. '02] Souza, E. P. M., C. Traina-Jr, A. J. M. Traina, et al. How to Use Fractal Dimension to Find Correlations between Attributes. First Workshop on Fractals and Self-similarity in Data Mining (in conjunction with 8th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining), Alberta, Canada. 26-30. 2002.
- [Spertus, E. '05] Spertus, E., M. Sahami e O. Buyukkocuten. Evaluating Similarity Measures: A Large Scale Study in the Orkut Social Network. 11th ACM SIGKDD international conference on Knowledge discovery in data mining KDD '05, Chicago, IL. USA. 678-684. 2005.
- [Spiegel, M. R. '76] Spiegel, M. R. EstatísticaSão Paulo, SP, Editora McGraw-Hill do Brasil Ltda. 1976.
- [Stan, D. '03] Stan, D. e I. K. Sethi. "eID: A System for Exploration of Image Databases." Information Processing and Management Journal **39**(3): 335-361. 2003.
- [Stehling, R. O. '02] Stehling, R. O. Recuperação por Conteúdo em Grandes Coleções de Imagens Heterogêneas. Tese de doutorado. Instituto de Computação, Universidade de Campinas. Campinas - Brasil. 2002.
- [Stollnitz, E. J. '95] Stollnitz, E. J., T. D. DeRose e D. H. Salesin. "Wavelets for Computer Graphics: A Primer." IEEE Computer Graphics and Applications **15**(3): pp. 76-84. 1995.
- [Sung, M. Y. '00] Sung, M. Y., M. S. Kim, M.-W. Sung, et al. CoMed: A Real-Time Collaborative Medicine System. 13th IEEE Symposium on Computer-Based Medical Systems (CBMS'00), Houston, Texas, IEEE Computer Society. 2000.
- [Takahashi, N. '00] Takahashi, N., M. Iwasaki, T. Kunieda, et al. "Image Retrieval Using Spatial Intensity Features." Signal Processing Image Communication(no. 16): pp. 45-57. 2000.
- [Tech, C. H. '88] Tech, C. H. e R. T. Chin. "On Image Analysis by Methods of Moments." IEEE Transactions on Pattern Analysis and Machine Intelligence **10**(4): 496-513. 1988.
- [Theodoridis, S. '99] Theodoridis, S. e K. Koutroumbas. Pattern RecognitionNew York, Academic Press. 1999.

- [Traina, A. J. M. '03a] Traina, A. J. M., C. Traina-Jr, J. M. Bueno, et al. "Efficient Content-based Image Retrieval through Metric Histograms." World Wide Web Journal **6**(2): 157-185. 2003a.
- [Traina, A. J. M. '03b] Traina, A. J. M., N. A. Rosa e C. Traina-Jr. Integrating Images to Patient Electronic Medical Records through Content-based Retrieval Techniques. 16th IEEE Symposium on Computer-based Medical Systems (CBMS'2003), New York, NY, USA. 163-168. 2003b.
- [Traina, A. J. M. '04] Traina, A. J. M., A. G. R. Balan, L. M. Bortolotti, et al. Content-based Image Retrieval Using Approximate Shape of Objects. 17th IEEE Symposium on Computer-Based Medical Systems (CBMS'2004), Bethesda, MD, USA. 91-96. 2004.
- [Traina-Jr, C. '00a] Traina-Jr, C., A. J. M. Traina, B. Seeger, et al. Slim-Trees: High Performance Metric Trees Minimizing Overlap Between Nodes. Intl. Conf. on Extending Database Technology, Konstanz, Germany, Springer Verlag. pp. 51-65. 2000a.
- [Traina-Jr, C. '00b] Traina-Jr, C., A. J. M. Traina, L. Wu, et al. Fast Feature Selection Using Fractal Dimension. XV Brazilian Database Symposium, João Pessoa - PA - Brazil. pp. 158-171. 2000b.
- [Tsang, W. '05] Tsang, W., A. Corboy, K. Lee, et al. Texture-Based Image Retrieval for Computerized Tomography Databases. 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05), Dublin, Ireland. 593-598. 2005.
- [Tversky, A. '77] Tversky, A. "Features of Similarity." Psychological Review **84**(4): pp. 327-352. 1977.
- [Tversky, A. '82] Tversky, A. e I. Gati. "Similarity, Separability and the Triangle Inequality." Psychological Review, **89**(1): pp. 123-154. 1982.
- [Twa, M. D. '03] Twa, M. D., S. Parthasarathy, T. W. Raasch, et al. Decision Tree Classification of Spatial Data Patterns From Videokeratography Using Zernike Polynomials. SIAM International Conference on Data Mining, San Francisco - CA. 2003.
- [Uhlmann, J. K. '91] Uhlmann, J. K. "Satisfying General Proximity/Similarity Queries with Metric Trees." Information Processing Letter **40**(4): pp. 175-179. 1991.
- [Vailaya, A. '01] Vailaya, A., M. A. T. Figueiredo, A. K. Jain, et al. "Image Classification for Content-based Indexing." IEEE Transactions on Image Processing **10**(1): 117-130. 2001.
- [Vasconcelos, N. '00] Vasconcelos, N. e A. Lippman. A Unifying View of Image Similarity. Intl. Conf. on Pattern Recognition (ICPR). 2000.
- [Vasconcelos, N. '04] Vasconcelos, N. "On the efficient evaluation of probabilistic similarity functions for image retrieval." IEEE Transactions on Information Theory **50**(7): 1482-1496. 2004.
- [Verma, B. '01] Verma, B. e J. Zakos. "A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques." IEEE Transactions on Information Technology in Biomedicine **5**(1): 46-54. 2001.
- [Vince, D. '00] Vince, D., K. Dixon, R. Cothren, et al. "Comparison of Texture Analysis Methods for the Characterization of Coronary Plaques in Intravascular Ultrasound Images." Computerized Medical Imaging and Graphics(no. 24): pp. 221-229. 2000.
- [Wilson, D. R. '97] Wilson, D. R. e T. R. Martinez. "Improved Heterogeneous Distance Functions." Journal of Artificial Intelligence Research **6**: 1-34. 1997.
- [Yin, T.-K. '04] Yin, T.-K. e N.-T. Chiu. "A computer-aided diagnosis for locating abnormalities in bone scintigraphy by a fuzzy system with a three-step minimization approach." IEEE Trans. Med. Imaging **23**(5): 639-654. 2004.

- [Zhang, D. S. '01] Zhang, D. S. e G. Lu. Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study. IEEE International Conference on Multimedia and Expo, Tokyo, Japan. 317-320. 2001.
- [Zhang, D. S. '03] Zhang, D. S. e G. Lu. Evaluation of Similarity Measurement For Image Retrieval. IEEE International Conference on Neural Networks & Signal Processing (ICNNSP03), Nanjing, China. 928-931. 2003.
- [Zhang, H. '02] Zhang, H. "Relevance Feedback in CBIR." Visual and Multimedia Information Management: pp. 21-36. 2002.
- [Zhang, P. '05] Zhang, P., K. Kumar e B. Verma. A Hybrid Classifier for Mass Classification with Different Kinds of Features in Mammography. Second International Conference Fuzzy Systems and Knowledge Discovery (FSKD 2005), Changsha, China. 316-319. 2005.
- [Zhou, X. S. '02] Zhou, X. S. e T. S. Huang. "Unifying Keywords and Visual Contents in Image Retrieval." IEEE MultiMedia **9**(2 (April)): 23-33. 2002.

Apêndice A

A. PROCESSAMENTO DE IMAGENS

Neste Apêndice, são apresentados resumidamente alguns conceitos envolvidos no processamento e na análise de imagens digitais, em especial algumas das técnicas que foram utilizadas pelos métodos propostos e implementados ao longo do projeto de doutorado.

A.1 Introdução

Uma imagem digital é usualmente o resultado de um processo de discretização de uma imagem contínua $I(x,y)$, a qual é armazenada no computador como uma matriz bidimensional $I(m,n)$. Cada elemento (m,n) da matriz corresponde a um pixel da imagem, cujo valor de intensidade ou luminância ficará armazenado nesse elemento. Uma vez que a luminância seja quantizada em N_g valores discretos de níveis de cinza, N_g irá identificar a "profundidade" da imagem. O valor de N_g é normalmente uma potência de 2, e comumente 256 (8 bits) para armazenamento digital. Esta definição aplica-se a imagens monocromáticas. Quando temos mais de uma cor, o grau da matriz irá aumentar de acordo com a quantidade de atributos usados pelo modelo de representação adotado.

Mesmo para uma imagem de dimensões reduzidas, a grande quantidade de pixels necessários para representá-la torna praticamente inviável o tratamento computacional de análise baseado diretamente nos valores de intensidade de cada pixel. Por isso, buscam-se formas de representação que sejam mais práticas e que, ao mesmo tempo, sejam capazes de manter as características originais da imagem. Os atributos utilizados para a representação da imagem, usualmente são oriundos do tratamento de seus descritores, ou então da aplicação de transformações sobre a imagem. As Seções subsequentes abordam essas formas de representação.

A.2 Descritores

Descritores são estímulos sensoriais que, quando devidamente modelados numericamente, podem caracterizar uma imagem, permitindo, assim, que a mesma seja representada e manipulada computacionalmente mantendo as suas características.

Os descritores para os quais as comunidades de Processamento de Imagens e Visão Computacional já conseguiram desenvolver modelos com um relativo grau de maturidade são: cor, textura e forma [Aslandogan, Y. A. '99].

A.2.1 Cor

Cor é uma propriedade relacionada com a capacidade dos objetos de refletir ondas eletromagnéticas de diferentes comprimentos, num espectro visível que varia de aproximadamente 400 nm até 700 nm. O olho humano detecta cores como sendo combinações das cores primárias vermelho, verde e azul, as quais para o propósito de padronização, têm sido definidas como possuindo respectivamente comprimentos de onda de 700 nm, 546.1 nm e 435.8 nm [Foley, J. D. '95].

Modelos de representação de cor

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.

O modelo de representação de cores mais popular é o **RGB** (*red* - vermelho, *green* - verde, *blue* - azul), no qual um pixel pode ser associado a um vetor tridimensional contendo as respectivas intensidades de cor, sendo (0,0,0) a cor preta, (k,k,k) a cor branca, (k,0,0) o vermelho puro, e assim por diante, onde k é o valor de quantização das cores (normalmente variando de 0 a 255). Isto implica em um espaço de cor de k^3 . As cores secundárias são combinações de duas cores primárias puras: magenta = azul + vermelho, amarelo = vermelho + verde e ciano = azul + verde. Os tons de cinza são obtidos variando-se o valor de k e atribuindo-se este valor para as três intensidades primárias.

Existem outros modelos de representação de cor. O mais simples talvez seja o **CMY** (*cyan*, *magenta*, *yellow*), que é baseado nas cores secundárias e é usado para construir um esquema de cores subtrativas, no qual a cor que vemos (refletida) em um objeto é aquela que resultou da absorção de outras cores pelo objeto (por isso se diz que é um esquema subtrativo). Por exemplo, um objeto amarelo (cor da luz refletida) possui esta cor porque ele absorve a componente azul da luz branca. Esta abordagem de "pigmento" é utilizada para misturar cores em equipamentos tais como impressoras.

O modelo **YIQ** é uma transformação linear simples do modelo RGB [Sonka, M. '99]. Ele é muito útil em transmissão de TV a cores, devido ao fato de o valor do componente Y sozinho caracterizar a informação necessária para apresentar a imagem de forma

monocromática. Além disso, este modelo explora propriedades do sistema visual humano, em particular a percepção da luminância (a energia percebida da fonte de luz).

O modelo alternativo de maior relevância para o processamento de imagens é o **HSI** (*hue* - matiz, *saturation* - saturação, *intensity* - brilho). Matiz refere-se à cor, ou seja, o comprimento de onda dominante; a saturação mede a sua diluição pela luz branca, caracterizando os tons claros e escuros. Além disso, ao isolar o componente de brilho (ou intensidade) dos outros dois componentes (que refletem a percepção humana), o modelo HSI torna a representação bastante útil para algoritmos de processamento de imagens, para realizar o realce da imagem com pequena alteração da cor. Este modelo também possui fórmulas de conversão com o modelo RGB.

O modelo **CIELab** é um modelo perceptivamente uniforme, ou seja, qualquer pequena perturbação ocorrida em uma cor irá produzir a mesma alteração na percepção em qualquer localização do espaço de cores [Rubner, Y. '01]. O componente L define a luminância, através de um eixo que varia do branco ao preto, e os outros dois componentes definem cor, sendo um eixo a que varia do azul ao amarelo, e um eixo b que varia do verde ao vermelho. Esse modelo é independente de dispositivo.

Histograma de Intensidades

Histograma de intensidades é a estrutura mais utilizada para a representação da distribuição global de cores de uma imagem. Um histograma de intensidades (ou de cor) $h_i(z)$ de uma imagem quantifica a frequência de ocorrência dos valores de luminância (z) apresentada pelos pixels da imagem. O histograma de uma imagem com L níveis de cinza será um vetor com L elementos (os *bins*⁴ do histograma). Normalmente os valores de frequência de um histograma são normalizados como frações centesimais do total de níveis de intensidades ou de cores, de forma que a soma de todas as frequências seja igual a 1. A forma clássica de se apresentar um histograma é através de um gráfico com os níveis de cinza nas abscissas e as respectivas frequências nas ordenadas. A Figura A.1 apresenta uma imagem tomográfica de crânio humano e o respectivo histograma de intensidades.

Os histogramas de intensidades são invariantes a rotação e translação (considerando-se a translação de objetos dentro da imagem) e também a escala, se eles estiverem normalizados.

Cada imagem possui apenas um histograma, porém o mesmo histograma pode estar representando diferentes imagens. Basta que as mesmas possuam as mesmas frequências de intensidade de pixels, independentemente de como os pixels se encontram distribuídos.

⁴ Um bin corresponde à medida de frequência de cada nível de intensidade mapeado pelo histograma.

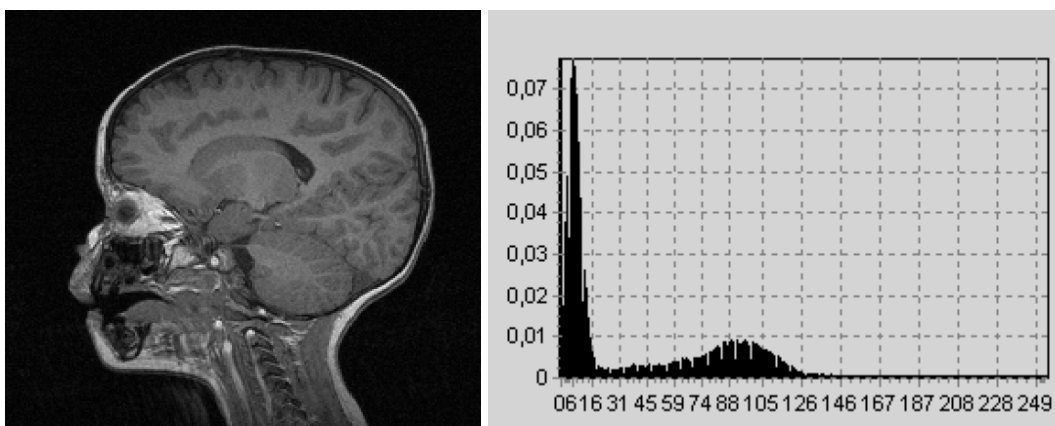


Figura A.1. Imagem de TC e seu histograma de intensidades

Além de ser uma importante ferramenta para o descritor cor na representação da imagem, o histograma é também útil em processos tais como manipulação de contraste e segmentação da imagem, como será visto nas Seções seguintes.

A.2.2 Textura

A textura de uma região de uma imagem é determinada pela forma como os níveis de cinza estão distribuídos nos pixels dessa região. Embora não haja uma definição clara de "textura", ela pode ser reconhecida e diferenciada pelo olhar humano, através de características tais como "fina" ou "grosseira", "lisa" ou "áspera", "homogênea" ou "não homogênea" e assim por diante [Theodoridis, S. '99]. A Figura A.2 apresenta alguns exemplos de diferentes texturas gerais. A Figura A.3 apresenta alguns exemplos de texturas de tecidos do corpo humano.

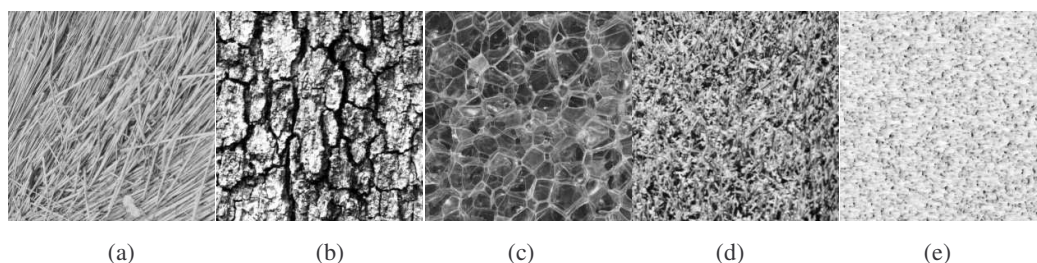


Figura A.2. Texturas da base de dados Brodatz [BRODATZ '05]: (a) capim; (b) casca de árvore; (c) bolhas de plástico; (d) grama; (e) areia

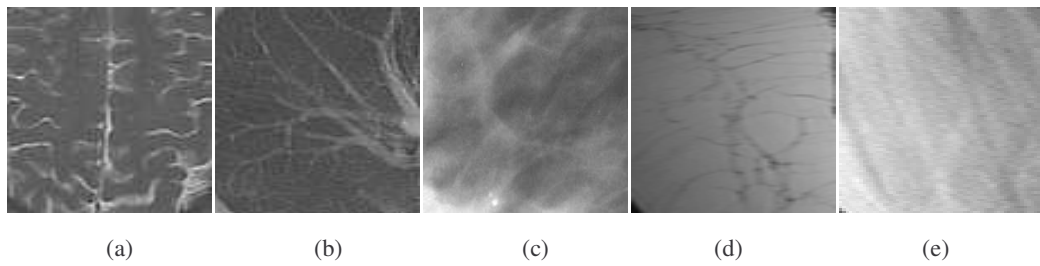


Figura A.3. Texturas de tecidos do corpo humano: (a) cérebro; (b) pulmão; (c) mama; (d) gordura; (e) osso

As três principais abordagens usadas no processamento de imagens para descrever a textura de uma região são [Gonzales, R. C. '87]: estatística, estrutural e espectral. A abordagem estatística considera a distribuição dos tons de cinza e o inter-relacionamento entre eles. As técnicas estruturais, por outro lado, lidam com o arranjo espacial de primitivas de imagens regulares. A abordagem espectral é baseada em propriedades de espectros de frequência, obtidos através de transformadas tais como a de Fourier e a de Wavelets.

Abordagem estatística

Esta abordagem tem sido amplamente utilizada [Sonka, M. '99], por gerar resultados satisfatórios com um custo computacional reduzido. A análise é realizada a partir da matriz de co-ocorrência da imagem, na qual os índices das linhas e colunas representam os diferentes valores de níveis de cinza, computando-se a frequência com que os mesmos ocorrem dois a dois em uma direção específica e a uma certa distância de vizinhança. O valor $P(i,j)$ armazenado em uma linha i e coluna j consiste no número de vezes em que pixels de valor i possuem vizinho de valor j na direção especificada e segundo a distância sobre as quais se refere aquela matriz. As direções usualmente tratadas são as de 0° , 45° , 90° e 135° e as distâncias são escolhidas de acordo com a granulosidade das imagens manipuladas, normalmente variando de 1 a 5. Para cada direção e cada distância, gera-se uma matriz de co-ocorrência. Como exemplo, suponhamos a imagem mostrada na Figura A.4a. Para a direção de 0 grau (horizontal à esquerda e à direita) e a distância 1, teremos a matriz de co-ocorrência mostrada na Figura A.4b. Posteriormente, a matriz sofre um processo de normalização, dividindo-se cada elemento pelo número máximo de elementos em cada posição, resultando em uma matriz cuja soma será igual a 1.

0	0	0	★	1	2
1	1	★	0	★	1
2	2	1	★	0	0
1	2	0		2	0
0	0	★	1	★	0
					★

Figura A.4a. Uma imagem com níveis de cinza 0, 1 e 2. Em destaque, as 7 vizinhanças entre pixels de valores 0 e 1, com direção 0° e distância 1.

	0	1	2
0	8	7	3
1	7	4	3
2	3	3	2

Figura A.4b. Matriz de co-ocorrência P da imagem, na direção 0° e distância 1. O valor 7, em destaque, é o valor de co-ocorrência entre os pixels de valores 0 e 1 com direção 0° e distância 1 na imagem.

De forma geral, o problema consiste em analisar um conjunto de matrizes de co-ocorrência para se caracterizar a textura à qual elas se referem, utilizando-se um ou mais descritores. O que estamos chamando, nesta Seção, de "descritores", na verdade são "sub-descritores" do descritor textura, ou seja, características que revelam diferentes especificidades da textura, a partir da co-ocorrência entre os pixels da imagem. Por exemplo, o descritor "Energia" é obtido da matriz de co-ocorrência pela somatória dos quadrados dos seus valores. Se a imagem possuir os seus diferentes valores de pixels bem distribuídos, os valores na matriz de co-ocorrência terão pouca discrepância e, assim, o valor resultante do cálculo da Energia será baixo. Dessa forma, o descritor Energia consiste em uma representação numérica do grau de uniformidade da imagem.

Em [Haralick, R. M. '73], um conjunto de 14 descritores é proposto. A Tabela A.1 mostra os mais utilizados. $P(i,j)$ é o valor do elemento (i,j) na matriz de co-ocorrência.

Normalmente, a escolha dos descritores ou características de imagens a serem adotados baseia-se em testes empíricos sobre um domínio específico, verificando-se quais deles produzem resultados mais satisfatórios.

Abordagem estrutural

Esta abordagem de textura é indicada quando são analisadas imagens com padrões fixos e repetitivos de textura. A idéia básica é que uma ou mais texturas primitivas simples podem ser usadas para formar padrões mais complexos de textura, de acordo com alguma regra que irá limitar o número de possíveis arranjos da(s) primitiva(s). A textura da imagem é então descrita pelo número e pelos tipos de primitivas, assim como por sua organização espacial.

Tabela A.1. Descritores de textura

Descritor	Equação	Significado
Variância	$\sum_i \sum_j (i-j)^2 P(i, j)$	nível de contraste da imagem
Entropia	$\sum_i \sum_j P(i, j) \log P(i, j)$	suavidade da imagem
Energia	$\sum_i \sum_j P^2(i, j)$	uniformidade da imagem
Homogeneidade	$\sum_i \sum_j P(i, j) / (1 + i - j)$	homogeneidade da distribuição dos pixels
Momento 3ª ordem	$\sum_i \sum_j (i-j)^3 P(i, j)$	nível de distorção da imagem
Variância Inversa	$\sum_i \sum_j P(i, j) / (i-j)^2$	nível inverso de contraste

Duas propriedades que podem ser utilizadas para caracterizar este tipo de textura são o **nível de cinza ou cor**, que representa as propriedades de intensidade dos pixels na primitiva, e a **estrutura**, que representa o relacionamento espacial entre as primitivas. Se as primitivas da imagem são pequenas e a diferença tonal entre primitivas vizinhas é alta, o resultado é uma textura fina. Se as primitivas são maiores, o resultado é uma textura áspera.

Os métodos para descrever texturas sintáticas são baseados numa analogia entre as relações espaciais das primitivas e a estrutura de uma linguagem formal. Uma gramática é construída para cada classe de textura presente em um conjunto de treinamento. O processo de reconhecimento é então uma análise sintática das "palavras" que descrevem a textura [Sonka, M. '99].

Abordagem espectral

Esta abordagem faz uso de espectros relacionados a transformadas (ver Seção A.4), como, por exemplo, o espectro de Fourier, com o qual pode-se descrever a direcionalidade de padrões bidimensionais periódicos ou quase periódicos em uma imagem [Gonzales, R. C. '87]. Os padrões globais de textura são facilmente identificados por picos de alta energia no espectro.

A detecção das características desejadas é feita por expressão do espectro em coordenadas polares, gerando duas funções: $S_2(r)$ que representa o valor do raio fixando-se a direção; e $S_r(\theta)$ que representa a direção, fixando-se o raio. A descrição da energia espectral de textura pode ser obtida a partir destas funções, para uma imagem completa ou uma região da mesma. Medidas típicas que representam as características de textura são: (a) localização de valores de pico, que fornecem as direções principais do padrão de textura; (b) média e variância da amplitude e das variações axiais e (c) a distância entre o valor médio e o valor mais alto da função.

A.2.3 Forma

Em muitas aplicações de análise de imagens, o reconhecimento de forma representa um problema de grande interesse. Especificamente em imagens médicas, podemos destacar a tarefa de classificar a malignidade de um nódulo, de acordo com seu formato. Nódulos com formato irregular possuem grande probabilidade de serem malignos, enquanto aqueles com formatos mais regulares, geralmente são benignos.

Definir a forma de um objeto é uma tarefa bastante difícil. A forma é usualmente representada verbalmente ou por figuras, e usam-se termos tais como alongado, arredondado, pontudo, etc. Considerando a necessidade de computação de contornos, surge a necessidade de se descrever formas as mais complexas com suficiente precisão. Embora haja muitos métodos práticos de descrição de forma, não há uma metodologia aceita de maneira unânime.

Antes de se aplicarem os algoritmos de caracterização de forma, geralmente a imagem passa por um processo de segmentação, para isolar o objeto que se deseja avaliar, seguido de um processo de extração de contorno [Costa, L. d. F. '00].

Uma característica de extrema importância dos sistemas de identificação de forma é que eles sejam invariantes a transformações geométricas (translação, rotação e escala) [Theodoridis, S. '99]. A caracterização de forma de uma região ou de um objeto pode ser obtida através de várias técnicas. A seguir três delas serão descritas sucintamente.

Transformada de Fourier

Conforme descrição teórica que será apresentada na Seção A.4, para um conjunto de N pontos pertencentes ao contorno de um objeto, a Transformada Discreta de Fourier (DFT) é composta por um conjunto de N coeficientes (números complexos) f_L , também conhecidos como os descritores de Fourier do contorno. Uma vez conhecidos todos os coeficientes f_L , a operação inversa permite recuperar os pontos originais e reconstruir o contorno. Porém, como

nosso objetivo não é reconstituir o contorno, mas apenas representá-lo com relativa precisão, um número reduzido de coeficientes é normalmente escolhido.

Operações de normalização aplicadas aos coeficientes podem tornar a DFT invariante às transformações geométricas.

Cadeia de Códigos

Esta é uma das técnicas mais amplamente utilizadas para descrição de forma [Theodoridis, S. '99]. A curva de contorno é aproximada por uma seqüência de segmentos de linha retos e conectados, dos quais são conhecidos o comprimento e a direção. Cada segmento é codificado de acordo com um número que corresponde à sua direção, e essa seqüência (no sentido horário) forma a chamada "cadeia de códigos".

Uma forma de se otimizar a construção da cadeia de códigos, eliminando ruídos também, consiste em se "quadricular" a imagem e considerar o contorno como uma linha que passa pelo centro de todos os quadrados onde existe algum ponto pertencente ao contorno original. O contorno obtido será uma aproximação do contorno original.

Algumas métricas podem ser construídas a partir desse modelo:

- (a) **freqüência direcional:** para cada direção, computa-se o número de códigos da cadeia que apresentam esta direção. Depois divide-se o valor resultante pelo total de códigos da cadeia.
- (b) **curvatura:** quantifica-se a ocorrência de diferentes concavidades dos ângulos de encontro de cada par de segmentos, considerando-se uma volta no sentido horário pelo contorno. Para oito direções, dezesseis atributos podem ser gerados, sendo oito para os cantos côncavos e oito para os cantos convexos.

Quando são computados os valores de área e perímetro do objeto, é possível fazer uso de alguns atributos que caracterizam a sua similaridade com uma circunferência. Isto é obtido por meio de medidas como, por exemplo:

- (a) **diâmetro efetivo:** corresponde ao diâmetro de uma circunferência que possua a mesma área. É calculado por:

$$d = 2\sqrt{\frac{A}{\pi}} \quad (\text{A.1})$$

onde: A é a área do objeto

- (b) **circularidade:** quanto mais próximo de 1, mais próximo da circunferência está o objeto:

$$c = \frac{4\pi A}{P^2} \quad (\text{A.2})$$

onde: P é o perímetro do contorno

Momentos Geométricos

Este grupo de medidas fornece informações valiosas sobre a imagem e são atributos populares na área de reconhecimento de padrões. Sua capacidade de representatividade reside no fato de que os momentos permitem a reconstrução da imagem a partir dos seus valores. Assim, cada coeficiente de momento agrega uma certa quantidade da informação a respeito da imagem. Alguns momentos serão abordados com detalhes na Seção A.5.

A.3 Operações em Imagens

A fim de facilitar a análise visual e a manipulação, diversas operações podem ser executadas sobre as imagens, por meio de alterações específicas nos valores de intensidade de seus pixels.

A.3.1 Realce de Contraste (*Stretching*)

O contraste da imagem pode ser alterado por meio da operação de realce de contraste, que consiste em se definir um intervalo de tons de cinza e ajustar ao mesmo o intervalo total de variação de tons de cinza da imagem, redistribuindo proporcionalmente os pixels. A Figura A.5 mostra a mesma imagem da Figura A.1 com realce do contraste e a Figura A.6 mostra a alteração executada em seu histograma.

A.3.2 Filtragem

Há ainda uma série de outras operações que podem ser aplicadas às imagens, a fim de modificá-las, tornando-as adequadas para objetivos específicos. Dentre estas operações, podemos citar a **convolução**, que consiste em se aplicar sobre cada pixel alterações que consideram os valores dos seus vizinhos [Castleman, R. '96]. A forma como os valores dos vizinhos são considerados é definida através de uma matriz (geralmente de pequena ordem) que contém os valores de ponderação de cada posição relativa ao valor central (pixel que está sendo modificado). Esta matriz é chamada de máscara. Diferentes máscaras produzirão diferentes efeitos sobre a imagem final.

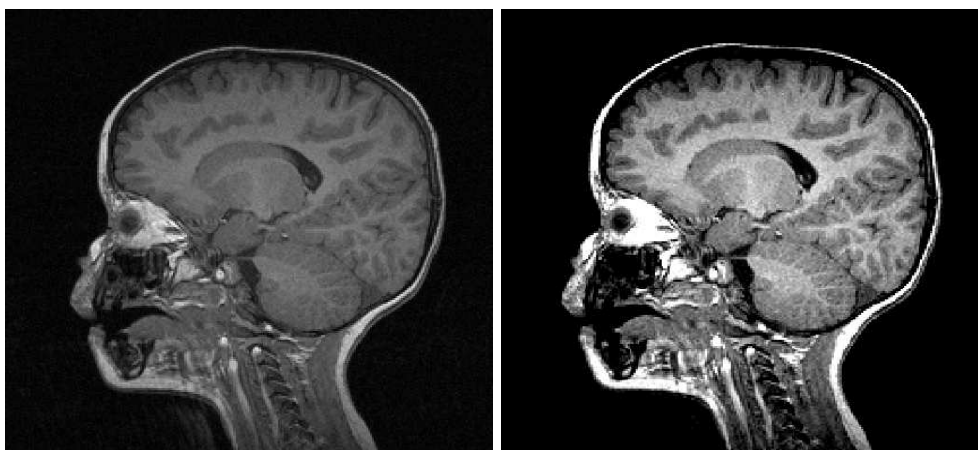


Figura A.5. Imagem da Figura A.1 (esquerda) após processo de realce de contraste (direita)

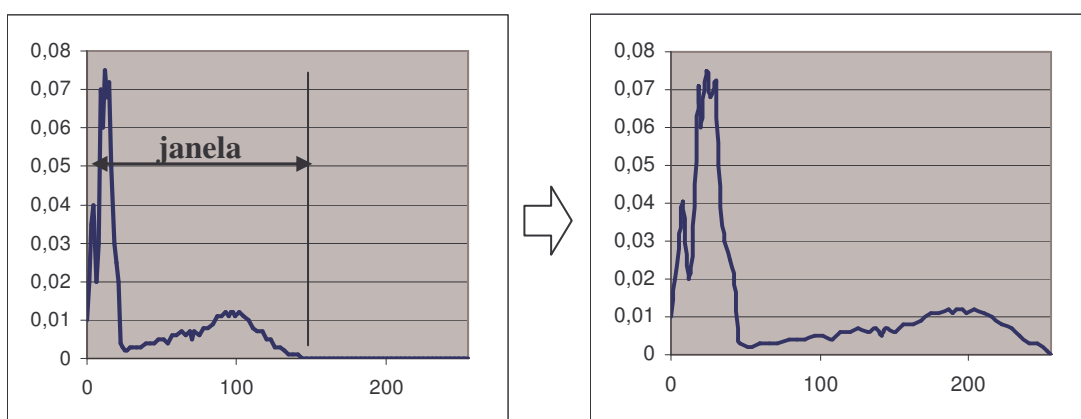


Figura A.6. "Janelamento" para realce de contraste da imagem da Figura A.1

Muitas dessas máscaras podem ser usadas como **filtros**. Um exemplo é o filtro **passa-baixa**, com o qual o valor do pixel é substituído pela média ponderada dos valores dos pixels de sua vizinhança. O seu efeito é a suavização da imagem, porém incorrendo em perda de nitidez do contorno. Outro filtro bastante usado é o filtro de **mediana**, com o qual o valor do pixel é substituído pelo valor mediano dos seus vizinhos. É eficiente para a eliminação de ruídos brancos, com pouca perda de contorno. A Figura A.7 mostra três imagens: a imagem original da Figura A.1, esta imagem após acréscimo de ruído e a imagem resultante da operação com um filtro de mediana sobre a imagem acrescida de ruído. Outro tipo de máscara importante é o filtro **passa-alta**, o qual produz o realce de bordas da imagem.

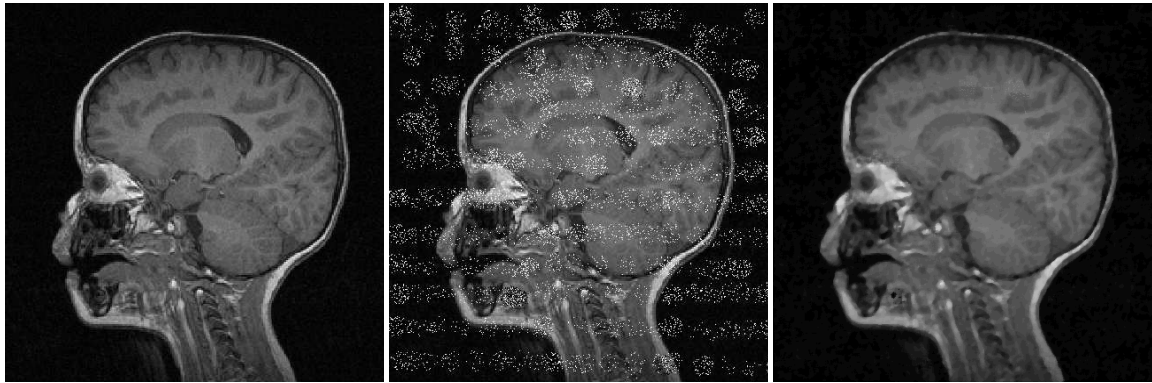


Figura A.7. Imagem da Figura A.1 (à esquerda), acrescida de ruído (centro) e resultante de aplicação do filtro da mediana com tamanho de vizinhança 2 (à direita) sobre a imagem com ruído

A.3.3 Operações Algébricas

Um grupo de operações importante é o das chamadas **operações algébricas**, que são definidas como a interação direta entre os pixels de duas (ou mais) imagens, produzindo uma terceira. As operações algébricas podem ser: (a) adição; (b) subtração; (c) multiplicação e (d) divisão; onde o valor de cada pixel da imagem final é resultado da referida operação sobre os respectivos valores nas imagens originais. A adição pode ser utilizada para gerar a média de múltiplas imagens de uma mesma cena, ou então para gerar o efeito de sobreposição de imagens, e também para atenuar os efeitos de ruídos aleatórios aditivos. A subtração pode ser utilizada para a remoção de padrões indesejáveis (por exemplo, artefatos presentes na imagem) e também para a detecção de mudanças na mesma cena (por exemplo, o resultado da sobreposição de imagens de angiografia com e sem contraste). Multiplicação e divisão podem ser utilizadas, por exemplo, para corrigir possíveis defeitos de um digitalizador em que a sensibilidade do sensor de luz varie conforme a posição dos pontos da imagem.

A.3.4 Segmentação

A segmentação é um processo que consiste na divisão da imagem em regiões distintas, considerando para imagens em tons de cinza, duas propriedades básicas: descontinuidade e similaridade. A partir de um processo de segmentação, é possível a delimitação de componentes da imagem, os quais podem ser utilizados para a classificação, e o reconhecimento de padrões de forma e textura, entre outros.

O processo de segmentação pode ser automático ou semi-automático (com a interação do usuário) e, segundo [Müller, H. '04], processos de segmentação totalmente automáticos para imagens médicas ainda é um problema não resolvido.

Segundo [Gonzales, R. C. '87], de acordo com o processo e o resultado final, o processo de segmentação pode ser classificado como: limiarização (*thresholding*), segmentação baseada em bordas e segmentação baseada em regiões. Porém, existem outros métodos, como, por exemplo, o baseado em cores e em algoritmos genéticos.

Limiarização (Thresholding). A limiarização é o método mais simples e intuitivo de segmentação de imagens. Basicamente, todos os pixels que apresentam valores de intensidade dentro de uma faixa específica são considerados como pertencentes a uma mesma região. Este método é geralmente aplicado a imagens simples que possuem áreas homogêneas sob um plano de fundo uniforme.

Métodos baseados na detecção de bordas. Os métodos de segmentação baseados na detecção de bordas envolvem basicamente a localização de regiões da imagem onde ocorrem variações bruscas de tons de cinza. Estas discontinuidades de tons podem ocorrer na forma de pontos isolados, linhas, segmentos ou curvas e, a partir delas, são formados os contornos (ou bordas) dos elementos contidos na imagem. A Figura A.8 mostra uma imagem e o resultado do processo de detecção de bordas aplicado à mesma. Nem todas as imagens apresentam níveis de discontinuidade suficientes para que se processe a detecção de bordas, inviabilizando, nesse caso, a aplicação desse método.



Figura A.8. (a) Imagem original; (b) Imagem com bordas detectadas.

[Baucher, S. '91]

Métodos baseados em regiões. Enquanto os métodos de limiarização e os métodos baseados em bordas consideram o problema de definir regiões significativas dentro da imagem a partir das diferenças nas tonalidades dos pixels, os métodos baseados em regiões segmentam a imagem a partir da detecção de similaridades entre os valores de tons de cinza dos pixels. Duas técnicas mais conhecidas desta categoria são denominadas de Crescimento de Região (*Region Growing*) e Divisão e Fusão (*Split and Merge*) [Balan, A. G. R. '03].

A Figura A.9 mostra uma imagem contendo um corte sagital de crânio humano obtido por ressonância magnética e o resultado da aplicação sobre a mesma do método conhecido como EM/MPM [Corner, M. L. '00].

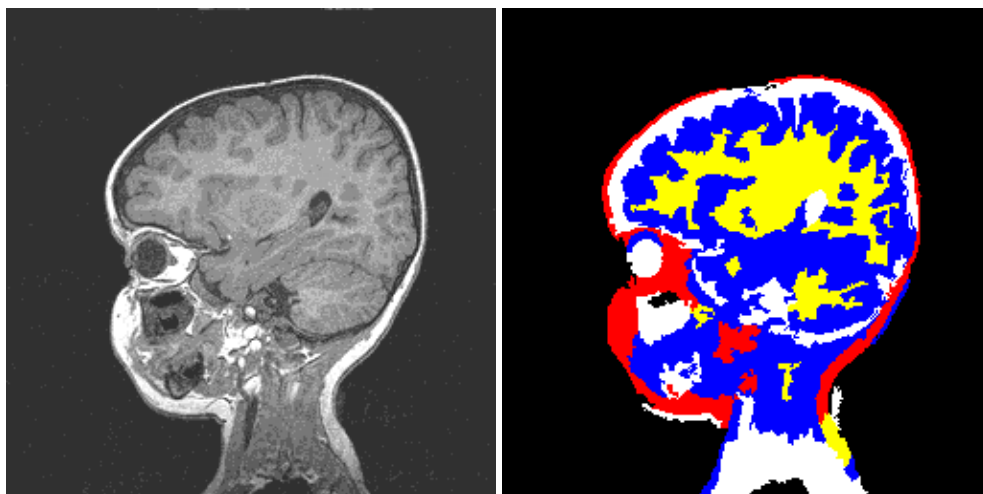


Figura A.9. (a) Imagem original; (b) Imagem segmentada em regiões pelo método EM/MPM [Traina, A. J. M. '04]

A.3.5 Morfologia Matemática

Morfologia Matemática (MM) é uma técnica não linear de processamento de imagens que define um conjunto de procedimentos que auxiliam nos processos de segmentação e de extração de características de uma imagem [Sonka, M. '99] [Beucher, S. '91]. Seu fundamento encontra-se na teoria dos conjuntos e na geometria integral, sendo utilizada atualmente para formalizar diversos problemas práticos e teóricos de processamento de imagens [Leite, N. J. '01].

A base da MM são as chamadas transformações morfológicas, que consistem de comparação entre a imagem original e uma imagem de referência, chamada de "elemento estruturante", resultando em alterações na imagem original de acordo com o tipo de operação que se executa.

Operações Morfológicas

Há duas operações elementares da MM: erosão e dilatação. Essas operações constituem a base para a construção de transformações mais complexas, por meio do encadeamento desses operadores.

Seja X o conjunto de pontos iniciais que constituem uma imagem binária a ser analisada, e B uma pequena imagem (elemento estruturante) representando um certo conhecimento geométrico e/ou topológico de X , ao qual é associado um ponto *origem* que representa a posição de referência para a aplicação das transformações.

A **dilatação** de X por B , $\delta_B(X)$, é dada pela união de todos os pontos x de \mathfrak{R}^2 tal que o elemento estruturante B , centrado em x , intercepta X .

A **erosão** de X por B , $\varepsilon_B(X)$, é definida pelo conjunto de pontos de x em \mathfrak{R}^2 tal que B , ao ser transladado para a posição x , esteja totalmente incluído no conjunto X .

A Figura A.10 apresenta exemplos de dilatação e erosão de uma imagem X por um elemento estruturante B .

Duas operações importantes, que se obtêm a partir de combinações de erosão e dilatação, são a **abertura** e o **fechamento** morfológicos. A abertura γ_B é caracterizada por uma erosão seguida por uma dilatação e o fechamento φ_B por uma dilatação seguida de uma erosão:

$$\gamma_B = \delta_B \circ \varepsilon_B \quad (\text{A.3})$$

$$\varphi_B = \varepsilon_B \circ \delta_B \quad (\text{A.4})$$

A abertura é anti-extensiva e o fechamento é extensivo. Em termos visuais, a abertura regulariza os contornos e elimina pequenas "ilhas" e "cabos" estreitos da imagem, enquanto o fechamento elimina pequenos "lagos" e "canais".

Outros exemplos de operações morfológicas elementares são o *gradiente morfológico*, o *chapéu mexicano claro* e o *chapéu mexicano escuro*, que podem ser empregados, entre outros, na definição do esqueleto morfológico de uma imagem [Leite, N. J. '01].

Segmentação Morfológica

Uma poderosa ferramenta de segmentação, baseada em MM, é a chamada transformação **Linha Divisora de Águas (LDA)** ou *watershed*. O processo de operação da LDA pode ser resumido da seguinte forma: suponhamos uma imagem em níveis de cinza que represente uma superfície topográfica, onde os níveis de cinza indicam as diferentes altitudes.

Imaginemos que os mínimos de cada bacia são perfurados e que essa superfície seja gradualmente imersa em um lago. À medida que a água preenche a superfície topográfica, dois ou mais fluxos vindos de mínimos diferentes podem se unir. Os diques construídos na superfície para evitar que tal junção ocorra constituem a LDA da imagem.

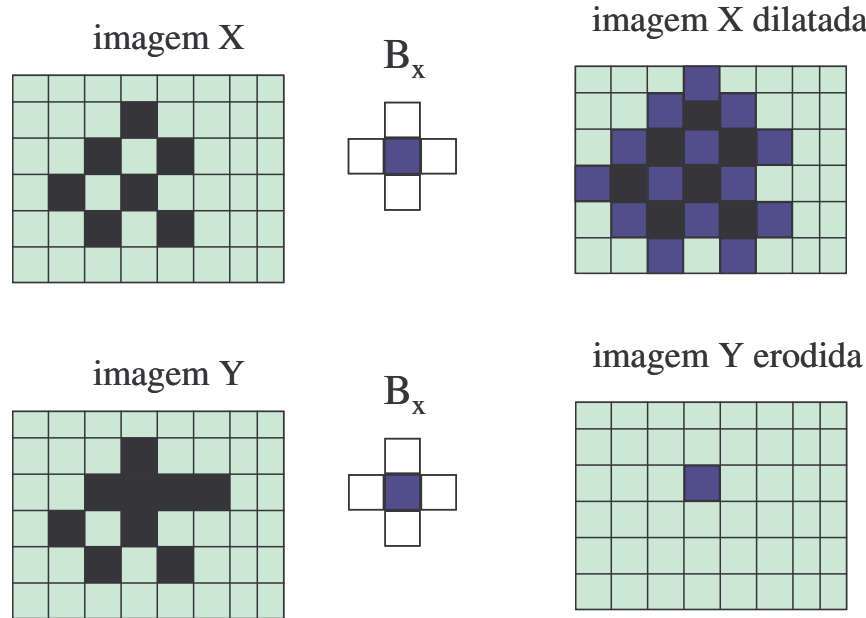


Figura A.10. Operações de dilatação e erosão [Leite, N. J. '01]

O resultado da aplicação pura desse algoritmo é uma super-segmentação da imagem. Na prática, para se obter uma segmentação coerente, aplicam-se operações em duas etapas principais: a primeira para obter marcadores associados às regiões que pretendemos segmentar e a segunda calculando a LDA sobre uma imagem que considera a informação destes marcadores. Os marcadores podem ser obtidos, por exemplo, a partir de uma operação de limiarização da image e uma imagem de relevo adequada à definição das regiões pode ser obtida com a operação de gradiente sobre a imagem original.

A.4 Transformadas

As transformações em imagens são o resultado da aplicação de técnicas que basicamente mudam a forma de representação das mesmas. A representação original – através da quantização de valores de pixels – é substituída, por exemplo, por um conjunto de coeficientes obtidos a partir da análise da imagem no domínio da frequência, que se opera através da transformação da mesma em uma soma de funções específicas. Essa nova

representação traz a vantagem de manter as informações contidas na imagem ("energia") com perdas em níveis aceitáveis, utilizando-se uma quantidade muito menor de atributos [Faloutsos, C. '96]. Essa característica é de importância crucial tanto para a área de compressão de imagens, quanto para a área de CBIR.

Quando se constroem as funções de base de uma transformada, a principal consideração diz respeito a propriedades locais dessas funções no domínio do tempo e da frequência. Os sinais que estamos considerando aqui são imagens 2D em tons de cinza. Procuramos, por transformadas que possam efetivamente representar variações de cor em qualquer região espacial local da imagem, de forma que coeficientes escolhidos dessas transformadas possam ser usados como atributos para formar a assinatura da imagem.

A.4.1 Transformada Discreta de Fourier

A idéia por trás da Transformada de Fourier baseia-se no teorema de Fourier, para o qual toda função contínua pode ser considerada como uma soma de funções senoidais [Smith, S. W. '99]. Para o caso discreto, que é aquele que interessa à aplicação em imagens digitais, a Transformada Discreta de Fourier (*Discrete Fourier Transform* - DFT) de n pontos de um sinal dado pelo vetor $x = [x_i]$, $i = 0, \dots, n-1$ é definida como sendo um vetor (seqüência) X de n números complexos X_f , $f = 0, \dots, n-1$, dada por

$$X_f = 1/\sqrt{n} \sum_{i=0}^{n-1} x_i \exp(-j2\pi f_i/n) \quad f = 0, 1, \dots, n-1 \quad (\text{A.5})$$

onde: j é a unidade imaginária

De forma compacta, $x \Leftrightarrow X$ denota o par DFT.

O sinal x pode ser recuperado pela transformada inversa:

$$x_i = 1/\sqrt{n} \sum_{f=0}^{n-1} X_f \exp(-j2\pi f_i/n) \quad f = 0, 1, \dots, n-1 \quad (\text{A.6})$$

X_f é um número complexo, com exceção de X_0 , que é real se o sinal de x é real.

A técnica que caracteriza a DFT consiste em decompor um sinal em funções senoidais e co-senoidais de várias frequências, múltiplas da frequência básica $1/n$. A razão do seu sucesso é que certas operações, como filtragem, remoção de ruído, convolução e correlação podem ser executadas mais convenientemente no domínio da frequência. Além disso, o domínio da

freqüência destaca algumas propriedades dos sinais, tais como a periodicidade, que pode ser associada, por exemplo, à textura da imagem.

A DFT possui algumas propriedades que são de singular importância para o tratamento de imagens: 1) a DFT preserva a energia (quadrado do comprimento) do sinal; 2) a DFT também preserva a distância euclidiana; 3) a DFT na sua versão *Fast Fourier Transform* (FFT) requer tempo computacional de ordem $O(n \log n)$.

A.4.2 Transformada Discreta do Co-Seno

Para nossos propósitos, a transformada ideal deve concentrar a energia no menor número possível de coeficientes iniciais, para a maioria dos sinais de interesse. Para diversos sinais reais, valores sucessivos são correlacionados: por exemplo, em uma imagem, se um pixel é escuro, há grande probabilidade de que seus vizinhos também sejam escuros. Nesses casos, a Transformada Discreta do Co-Seno (DCT) alcança uma concentração de energia que é melhor do que a da DFT, além de ser muito próxima do valor ótimo [Faloutsos, C. '96].

Além disso, a DCT evita problemas de "vazamento de freqüência" que contaminam a DFT quando o sinal de entrada possui uma tendência. A DCT resolve esse problema de forma inteligente, refletindo conceitualmente a seqüência original no eixo do tempo em torno do último ponto e tomando a DFT na seqüência resultante. Exatamente pelo fato do sinal ser simétrico, todos os coeficientes serão números reais. Sendo assim, precisamos manter apenas as primeiras amplitudes. Assim como acontece com a DFT, a complexidade da DCT é $O(n \log(n))$.

A.4.3 Wavelets

A Transformada de *Wavelets* consegue evitar o problema de "vazamento de freqüência" de forma ainda mais eficaz. Se considerarmos, por exemplo, o caso da função impulso, tanto na DFT quanto na DCT ela possui amplitudes diferentes de zero em todas as freqüências. Portanto, algo que usa um valor único para ser descrito no domínio do tempo irá exigir vários valores no domínio da freqüência. O problema é que a DFT não possui localização temporal: cada um dos seus coeficientes fornece informações sobre todos os instantes de tempo. Uma solução parcial seria a chamada "Transformada por Janela de Fourier" (WFT), com a qual a seqüência temporal é dividida em janelas e a transformada de Fourier da porção "visível" do sinal é calculada para cada posição da janela. Assim, um impulso no domínio do tempo terá um vazamento de freqüência restrito.

No domínio do tempo, cada valor fornece todas as informações sobre aquele instante (mas nenhuma informação a respeito da frequência). A DFT possui coeficientes que dão todas as informações sobre uma dada frequência, mas ela precisa de todas as frequências para recuperar o valor em um dado instante no tempo. Nesse sentido, WFT possui natureza intermediária. Porém, uma vez definido o tamanho da janela na WFT, ele permanecerá constante para todas as frequências. Pode acontecer que algumas aplicações precisem de um enfoque mais flexível, onde o tamanho da janela possa mudar, com o objetivo de detectar o conteúdo local da frequência. Esse problema é tratado e é a base da **Transformada Discreta de Wavelets (DWT)**.

Diversas transformadas discretas de Wavelets foram propostas. A mais simples de descrever e codificar é a **DWT de Haar** [Stollnitz, E. J. '95]. Ignorando temporariamente algumas constantes de proporcionalidade, a transformada de Haar opera na totalidade do sinal, dando a soma e a diferença entre as partes esquerda e direita, focando recursivamente em cada uma das metades e computando a diferença de suas sub-divisões, até que atinja um intervalo com um único exemplo.

Consideremos um sinal de entrada x de tamanho n (n é potência de 2):

Nível 0: Tomam-se os dois primeiros valores x_0 e x_1 e computa-se a sua soma $s_{0,0}$ e a sua diferença $d_{0,0}$. Repete-se esse procedimento para todos os outros pares de pontos (x_{2i} , x_{2i+1}). Assim, $s_{0,i} = C * (x_{2i} + x_{2i+1})$ e $d_{0,i} = C * (x_{2i} - x_{2i+1})$, onde C é uma constante de proporcionalidade. Os valores $s_{0,i}$ ($0 \leq i \leq n/2$) constituem uma versão "plana" (baixa frequência) do sinal, enquanto os valores $d_{0,i}$ representam o conteúdo de alta frequência do mesmo.

Nível 1: Considerando os valores "planificados" de $s_{0,i}$, repete-se o processo do nível 0 com os mesmos, obtendo-se uma versão "mais plana" do sinal: $s_{1,i}$ e as respectivas diferenças $d_{1,i}$ ($0 \leq i \leq n/4$).

E assim sucessivamente, até que se atinja um sinal de tamanho 2.

A DWT Haar do sinal original x é a coleção de todos os valores de diferenças $d_{l,i}$ em cada nível L e posição i , mais a componente plana $s_{L,0}$ do último nível L ($L = \log_2(n) - 1$).

Um valor interessante para a constante C é $1/\sqrt{2}$, porque ele faz com que a matriz de transformação seja ortonormal. Assim, a transformada de Haar será calculada como:

$$d_{l,i} = 1/\sqrt{2} (s_{l-1,2i} - s_{l-1,2i+1}) \quad l = 0, \dots, L, \quad i = 0, \dots, n/2^{l+1} - 1 \quad (\text{A.7})$$

com

$$s_{l,i} = 1/\sqrt{2} (s_{l-1,2i} + s_{l-1,2i+1}) \quad l = 0, \dots, L, \quad i = 0, \dots, n/2^{l+1} - 1 \quad (\text{A.8})$$

e a condição inicial

$$s_{-1,i} = x_i$$

Esse procedimento é usado por todas as transformadas de *wavelets*: inicia-se no nível mais baixo, aplicando-se duas funções em janelas sucessivas do sinal: a primeira função realiza uma planificação, do tipo de uma média ponderada, enquanto a segunda função realiza uma subtração ponderada. A versão planificada (e reduzida) do sinal é recursivamente devolvida ao processo, até que o sinal resultante seja suficientemente pequeno.

Dentre as transformadas de *Wavelets* propostas, a mais popular é a **Transformada de Daubechies**, a qual apresenta uma melhor concentração de energia, preservando melhor a informação de tendência dos sinais, quando comparada à transformada de Haar.

A DWT Daubechies-4 opera com 4 pontos consecutivos:

$$s_{0,i} = h_0x_{2i} + h_1x_{2i+1} + h_2x_{2i+2} + h_3x_{2i+3} \quad i = 0, \dots, n/2 \quad (\text{A.9})$$

$$d_{0,i} = h_3x_{2i} - h_2x_{2i+1} + h_1x_{2i+2} - h_0x_{2i+3} \quad i = 0, \dots, n/2 \quad (\text{A.10})$$

onde

$$\begin{aligned} h_0 &= (1 + \sqrt{3})/(4\sqrt{2}), & h_1 &= (3 + \sqrt{3})/(4\sqrt{2}), \\ h_2 &= (3 - \sqrt{3})/(4\sqrt{2}), & h_3 &= (1 - \sqrt{3})/(4\sqrt{2}) \end{aligned} \quad (\text{A.11})$$

A DWT Haar corresponde à DWT Daubechies-2.

A complexidade computacional da DWT é $O(n)$, ou seja, ela se mostra mais rápida do que as transformadas vistas anteriormente. Além disso, a DWT atinge concentrações de energia ainda melhores do que a DFT e a DCT.

A.5 Momentos

Seja $I(x,y)$ uma função contínua que representa os valores de luminância em cada posição (x,y) da imagem. O seu **momento geométrico** de ordem $p+q$ é definido como:

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q I(x, y) dx dy \quad (\text{A.12})$$

Para imagens digitais, compostas por valores discretos, as integrais da equação A.12 podem ser substituídas por somatórias em x e y . Dessa forma, \mathbf{m}_{00} será a somatória dos valores de luminância de todos os pixels, \mathbf{m}_{10} a variância das linhas e \mathbf{m}_{01} a variância das colunas.

Uma propriedade desejável em reconhecimento de padrões é a invariância a transformações geométricas. Os momentos, como definidos através da equação A.12, dependem das coordenadas do objeto de interesse dentro da imagem. Portanto eles não são invariantes. Esse problema pode ser contornado através da definição de combinações apropriadas de versões normalizadas.

A.5.1 Momentos Centrais

Definimos **momentos centrais** como:

$$\mu_{pq} = \iint I(x, y)(x - \bar{x})^p (y - \bar{y})^q dx dy \quad (\text{A.13})$$

$$\text{onde: } \bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

\bar{x} e \bar{y} representam o "centro de massa" da imagem $I(x,y)$. Os momentos centrais são invariantes a translações.

Definimos **momentos centrais normalizados** como:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (\text{A.14})$$

$$\text{onde: } \gamma = \frac{p + q + 2}{2}$$

Os momentos centrais normalizados são invariantes a translação e a escala.

Ao "binarizar" a região ou objeto cujo contorno deseja-se representar, fazendo o valor de $I(x,y)$ igual a 1 na região interna e igual a 0 na região externa, obtém-se um meio rápido de calcular os momentos, considerando-se apenas os pixels internos ao contorno:

$$m_{pq} = \sum_x \sum_y x^p y^q \quad (\text{A.15})$$

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q \quad (\text{A.16})$$

Dos momentos de segunda e terceira ordem, podemos derivar um conjunto de sete momentos que são invariantes a rotação, translação e escala. Eles são conhecidos como **momentos de Hu**. Sua formulação pode ser encontrada em [Sonka, M. '99] e [Theodoridis, S. '99].

A.5.2 Momentos de Zernike

Uma classe mais complexa de momentos invariantes são os chamados **momentos de Zernike**, que consistem em projeções sobre funções polinomiais complexas, as quais formam uma base ortogonal. Os momentos de Zernike se mostram bastante eficazes na representação de formas complexas, compostas de várias regiões disjuntas.

Os polinômios de Zernike constituem-se em um modelo matemático preciso que captura o formato global, ao mesmo tempo em que preserva informações suficientes, por meio de harmônicas. Eles formam um conjunto ortogonal completo no interior do círculo unitário $x^2 + y^2 = 1$, representando a projeção do mapa de pixels da imagem sobre essas funções de base e, se adequadamente implementadas, podem ser invariantes a escala, rotação e translação.

A representação polar dos momentos de Zernike de ordem n e repetição l é definida como:

$$A_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{nl}(r, \theta)]^* f(r \cos \theta, r \sin \theta) r \, dr \, d\theta \quad (\text{A.17})$$

com $n - l = \text{par}$ e $|l| \leq n$

onde r e θ são as coordenadas polares dos pixels

$f(r \cos \theta, r \sin \theta)$ é o valor do brilho no pixel representado por r e θ

$V_{nl} = \sqrt{(R_{nl} \cos(l\theta))^2 + (R_{nl} \sin(l\theta))^2}$ é o polinômio de Zernike e

$$R_{nl}(r) = \sum_{s=0}^{\frac{n-|l|}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|l|}{2} - s\right)! \left(\frac{n-|l|}{2} - s\right)!} r^{n-2s} \quad \text{é o polinômio radial ortogonal}$$

A quantidade de raízes de $R_{nl}(r)$ representa a sua capacidade de descrever componentes de alta frequência espacial da imagem [Kan, C. '01]. Na Figura A.11 podemos

observar esta propriedade, pela análise visual da representação espacial das séries relativas a cada ordem dos momentos.

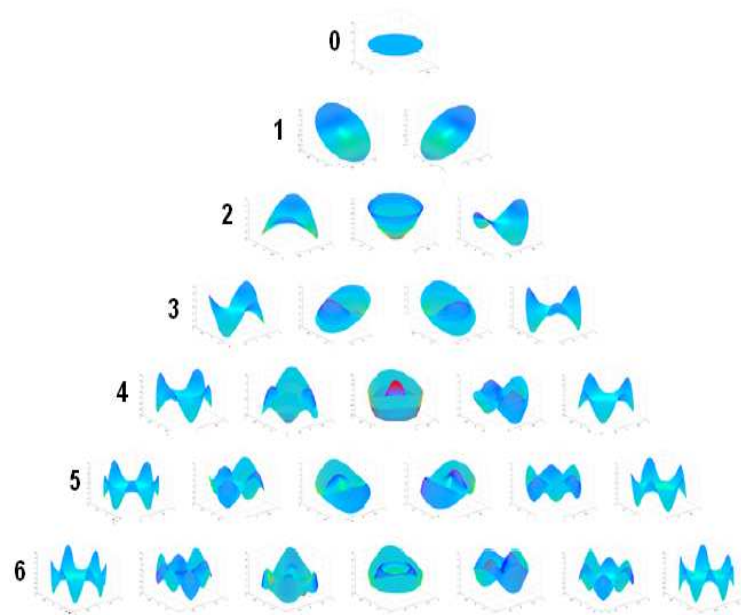


Figura A.11. Representação espacial das séries polinomiais de Zernike até 6ª ordem [Twa, M. D. '03]

Em [Tech, C. H. '88] são avaliados vários tipos de momentos, levando em consideração a redundância da informação e a sensibilidade ao ruído, e conclui-se que os momentos de Zernike apresentam melhores resultados para a representação de imagens.

Em [Zhang, D. S. '01] são realizadas comparações de diversos descritores de forma que têm sido freqüentemente adotados em CBIR: descritores de Fourier, descritores de espaço de escala de curvatura, momentos de Zernike e descritores de *grid*. As potencialidades e limitações de cada um deles são analisadas a partir de parâmetros tais como invariância, robustez, complexidade computacional e medição da similaridade perceptual. O conjunto de momentos de Zernike foi o descritor que atingiu os melhores resultados na avaliação geral.

A.6 Considerações Finais

Este Apêndice contém uma revisão sobre Processamento de Imagens, abordando as suas principais técnicas e, em particular, as técnicas utilizadas nas implementações realizadas durante o projeto de doutorado.

