

Márcio Júnio Ribeiro Ferreira

**Um Modelo de Recuperação de Imagens por Conteúdo
Através da Quantização do Espectro de Fourier**

Uberlândia-MG

Dezembro / 2005

Márcio Júnio Ribeiro Ferreira

Um Modelo de Recuperação de Imagens por Conteúdo Através da Quantização do Espectro de Fourier

Dissertação apresentada ao Programa de
Pós-Graduação da Faculdade de Ciência da
Computação da Universidade Federal de
Uberlândia como requisito para obtenção do
grau de Mestre em Ciência da Computação.

Orientadora:

Prof. Dra Celia Aparecida Zorzo Barcelos

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Uberlândia-MG

Dezembro 2005

Márcio Júnio Ribeiro Ferreira

Um Modelo de Recuperação de Imagens por Conteúdo Através da Quantização do Espectro de Fourier

Dissertação apresentada ao Programa de Pós-Graduação da Faculdade de Ciência da Computação da Universidade Federal de Uberlândia como requisito para obtenção do grau de Mestre em Ciência da Computação.

Uberlândia, 19 de Dezembro de 2005.

Prof. Dra. Celia A. Zorzo Barcelos - UFU

Prof. Dr. Junior Berrera - USP

Prof. Dr. Ilmério Reis da Silva - UFU

Universidade Federal de Uberlândia

Autor: **Márcio Júnio Ribeiro Ferreira**
Título: **Um Modelo de Recuperação de Imagens por Conteúdo Através da Quantização do Espectro de Fourier**
Faculdade: **Ciência da Computação**
Grau: **Mestrado**

Fica garantido à Universidade Federal de Uberlândia o direito de circulação e impressão deste material para fins não comerciais, bem como o direito de distribuição por solicitação de qualquer pessoa ou instituição.

O autor reserva para si qualquer outro direito de publicação deste material.

*Dedico esta vitória aos meus pais, **Francisco e Maria**, aos meus irmãos, **Ângela e Marcos**, e à **Mylene**, a quem tanto amo e admiro. Esta não é apenas uma conquista minha mas sim, de todos vocês, por tudo que fizeram por mim. Obrigado.*

Agradecimentos

Em primeiro lugar agradeço à **Deus**, que me deu a força necessária para nunca desistir, mesmo nos momentos difíceis.

Aos meus pais e irmãos, por estarem sempre me apoiando e, pelo imenso amor dedicado à mim. Obrigado pelo seu infinito desejo de que tudo desse certo.

À minha orientadora, **Prof. Dra Celia Aparecida Zorzo Barcelos**, por sua estimada ajuda na orientação e fortalecimento deste trabalho.

À **Mylene Lemos Rodrigues**, cuja contribuição e incentivo foram inestimáveis durante todas as etapas deste trabalho.

Aos meus caros colegas de curso, **Ivan e Alexandre** (Chucky), pelo companherismo, troca de experiências e compartilhamento de idéias.

Aos demais colegas, **professores, dirigentes e funcionários** da pós-graduação do curso de Ciência da Computação da Universidade Federal de Uberlândia . Em especial ao **Prof. Dr Ilmério Reis da Silva**, pela sua colaboração.

E finalmente, à todas as demais pessoas que não foram mencionadas mas que, direta ou indiretamente contribuíram com o meu sucesso, me ajudando, me apoiando ou simplesmente torcendo por mim. Compartilho com todos vocês as minhas conquistas durante este período.

Muito obrigado à todos!

“Não há caminho para a paz, a paz é o caminho.”

Mahatma Gandhi

Resumo

A recuperação de imagens é uma importante área de pesquisa em Processamento Digital de Imagens e Visão Computacional, que encontra aplicações nas mais variadas áreas, como diagnóstico de imagens médicas, prevenção ao crime, identificação pessoal (impressão digital), propriedade intelectual, etc. Os sistemas de recuperação de imagens por conteúdo (CBIR-*Content-Based Image Retrieval*) têm como objetivo recuperar imagens armazenadas em coleções de imagens que sejam mais similares à uma imagem consulta escolhida pelo usuário, com base nas características extraídas automaticamente das imagens. O surgimento de sistemas CBIR pode ser justificado pelo fato de que os métodos tradicionais de indexação de imagens baseados em texto consomem bastante tempo e requerem considerável esforço manual na indexação de grandes coleções. As características visuais mais exploradas em CBIR são a cor, a textura e a forma. Em relação à textura, existem três abordagens principais: a abordagem estatística, a estrutural e a 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 estruturais, enquanto que a abordagem espectral é baseada em propriedades de espectros de frequência, obtidos através de transformadas tais como a de Fourier e a de Wavelets. Seguindo as idéias apresentadas por Shapiro e Brady e por Carcassoni, Ribeiro e Hancock, neste trabalho explora-se como a estrutura modal dos padrões, tomados no espaço da frequência das texturas, pode ser utilizada para fins de reconhecimento. Carcassoni, Ribeiro e Hancock apresentaram uma variação do método de correspondência modal de Shapiro e Brady, que visa realizar casamento entre conjuntos de pontos através da comparação dos autovetores da matriz que mede a inter-relação entre estes pontos (matriz proximidade). Carcassoni, Ribeiro e Hancock utilizaram um descritor de texturas baseado nos picos do espectro de potência da imagem, para representá-la. Neste trabalho foi utilizada uma variação da técnica de quantização de Lloyd, realizada à partir do espectro de potência da imagem, para obter a representação da mesma. Para comprovar a eficiência do método, diversos experimentos foram realizados em uma coleção de imagens regulares, não regulares, homogêneas e não homogêneas. A coleção é formada por imagens de tecidos, papéis de parede, paisagens, veículos, madeira, tijolos, construções, etc, extraídas de diversos bancos de dados. Os resultados obtidos pelo método proposto são comparados com o trabalho de Carcassoni et al. e com o método da matriz de co-ocorrência de níveis de cinza, de Haralick, que é um método de abordagem estatística bastante conhecido e utilizado na extração de padrões de textura. A performance dos três métodos comparados foi medida através de gráficos de *precision* e *recall*, que constituem uma importante ferramenta na análise de performance de sistemas de recuperação de informações.

Palavras-chave : recuperação de imagens por conteúdo - quantização - espectro de potência - análise modal - banco de dados de imagens.

Abstract

Image retrieval is an important research area in Digital Image Processing and Computational Vision that can be applied in many areas such as medical images diagnosis, crime prevention, personal identification (finger-print), intellectual property, etc. The content-based image retrieval systems (CBIR) has as the main goal of retrieving images in image database that are more similar with a query image chosen by the user, based on the features automatically extracted from the images. The appearance of the CBIR systems can be justified by the fact that traditional indexation methods based on text, require much more time and efforts in the indexation process for huge images databases. The most explored visual features in CBIR are color, texture and shape. Concerning to texture, there are three main approaches: a statistical, a structural and a spectral one. The statistical approach considers the color distribution and their inter-relationship. The structural approach, by the other side, works with spatial arrange of structural primitives, while the spectral approach is based on the spectral frequency properties, obtained through transformations such as Fourier and Wavelets. Following the ideas presented by Shapiro and Brady and Carcassoni, Ribeiro and Hancock, this work explores how the modal structure of the pattern, taken in the textures' frequency space can be used for retrieval purposes. Carcassoni, Ribeiro and Hancock presented a variation of the correspondence method of Shapiro and Brady, that aims to match point sets by comparing the eigenvectors of a matrix that measures the inter-relationship between the pairwise points (proximity matrix). Carcassoni, Ribeiro and Hancock introduces a texture descriptor based on the image power spectrum peaks, with the aim of represent it. In this work, was used a variation of the Lloyd's quantization technique from the image power spectrum to represent it. With the aim of verifying the method efficiency, several experiments were carried out using regular, non-regular, homogeneous and non-homogeneous textures. The image collection is composed of images such as tissue, fabric paper, landscapes, vehicles, wooden floor, bricks, buildings images, etc, that were extracted from several images database. The results obtained by the proposed method are compared with the Carcassoni's method and also with the gray level co-occurrence matrix method of Haralick, that is a well-known and a method widely used for texture feature extraction. The performance of the three compared methods is measured by the commonly used retrieval performance measurement, precision and recall, which is considered one of the most important techniques for performance analysis of any retrieval systems.

Key-words : content-based image retrieval - quantization - power spectrum - modal analysis - image database.

Sumário

Lista de Figuras	xii
Lista de Tabelas	xv
1 Introdução	1
1.1 Motivação	3
1.2 Organização do Trabalho	4
1.3 Considerações Finais do Capítulo	5
2 Recuperação de Imagens por Conteúdo	6
2.1 Introdução	6
2.2 Aplicações em CBIR	9
2.3 Consultas em Sistemas de Recuperação de Imagens	10
2.4 Extração de Características	10
2.5 Atributos das Imagens	10
2.6 Recuperação Utilizando Cores	11
2.7 Recuperação Utilizando Texturas	12
2.8 Recuperação Utilizando Formas	13
2.9 Resumo das Características de Baixo Nível	15
2.10 Softwares Comerciais CBIR Disponíveis	16
2.10.1 QBIC	16
2.10.2 MARS	17
2.10.3 VisualSEEK e WebSEEK	17

2.11	Considerações Finais do Capítulo	17
3	Fundamentos Teóricos	19
3.1	Introdução	19
3.2	Conceitos em Imageamento Digital	19
3.2.1	Pixel	19
3.2.2	Imagem Digital e Valores Padrões	19
3.2.3	Definição de Textura e seus Padrões	20
3.3	Conceitos Matemáticos	23
3.3.1	Transformada de Fourier	23
3.3.1.1	Transformada Discreta de Fourier	23
3.3.1.2	Algumas Propriedades da Transformada Bidimensional de Fourier	24
3.3.1.3	Transformada Rápida de Fourier	25
3.3.1.4	Fase e Magnitude do Espectro	25
3.3.2	Base Ortonormal	26
3.3.3	Autocorrelação	26
3.3.4	Espectro de Potência	26
3.3.5	Autovalor e Autovetor	27
3.3.6	Norma	27
3.3.7	Similaridade	27
3.4	Considerações Finais do Capítulo	28
4	Trabalhos Correlatos	30
4.1	Introdução	30
4.2	Método de Correspondência de Shapiro e Brady	32
4.2.1	O Algoritmo de Shapiro e Brady	33
4.2.2	Exemplo de Correspondência entre Pontos Padrões	34

4.2.3	Resultados Experimentais	35
4.3	Método de Correspondência Espectral de Carcassoni	37
4.3.1	O Algoritmo de Carcassoni	37
4.3.2	A Representação do Espectro de Potência	37
4.3.3	Cálculo da Matriz Proximidade dos Picos Dominantes	38
4.3.4	Representação Modal	39
4.3.5	Cálculo da Matriz Centróide	39
4.3.6	Cálculo da Similaridade	40
4.3.7	Resultados Experimentais	40
4.4	Matrizes de Co-ocorrência de Níveis de Cinza	41
4.4.1	Algoritmo do Método GLCM	43
4.4.2	Resultados Experimentais	44
4.5	Técnica de Lloyd	44
4.6	Considerações Finais do Capítulo	45
5	Método Proposto	47
5.1	Introdução	47
5.2	Fluxograma do Método Proposto	47
5.3	A Representação do Espectro	48
5.4	Modelo de Quantização	49
5.5	Construção da Relação entre os Pontos Representantes da Imagem	52
5.6	Cálculo da Similaridade	53
5.7	Resultados Experimentais	54
5.8	Considerações Finais do Capítulo	54
6	Resultados Obtidos	56
6.1	Introdução	56

6.2	Banco de Dados de Imagens	56
6.3	Grupos de Experimentos	57
6.3.1	Primeiro Grupo de Experimentos	59
6.3.2	Segundo Grupo de Experimentos	60
6.3.3	Terceiro Grupo de Experimentos	63
6.4	Parâmetros do Algoritmo Proposto	66
6.5	Tempo de Execução	67
6.6	Considerações Finais do Capítulo	67
7	Análise dos Resultados	72
7.1	Introdução	72
7.2	Precision e Recall	72
7.3	Gráficos de Precision e Recall	72
7.4	Análise dos Resultados	73
7.4.1	Conjunto de Relevantes	73
7.4.2	Performance no Primeiro Grupo de Experimentos	74
7.4.3	Performance no Segundo Grupo de Experimentos	74
7.4.4	Performance no Terceiro Grupo de Experimentos	75
7.5	Considerações Finais do Capítulo	77
8	Conclusão, Contribuições e Trabalhos Futuros	78
	APÊNDICE A - Experimentos Adicionais	80
	Número de Representantes	80
	APÊNDICE B - Sistema CBIR Proposto	81
	Interface do Usuário	81
	Referências	83

Lista de Figuras

1.1	Abordagens em Recuperação de Imagens.	2
2.1	Diagrama típico de sistemas CBIR.	7
2.2	A difícil tarefa de anotação de uma imagem rica em conteúdo.	8
2.3	Imagem de uma edificação e seu histograma de cores (à direita).	12
2.4	Imagem original (à esquerda) e a mesma após o algoritmo de detecção de bordas de Canny.	14
3.1	Exemplo de imagem digital de dimensão 256×256	20
3.2	Exemplo de textura.	22
3.3	Exemplo de texturas regulares.	22
3.4	Exemplo de texturas não regulares.	22
3.5	Imagem original (à esquerda), a sua magnitude do espectro (centro) e a sua fase (à direita). A transformada inversa de Fourier utiliza ambas, a magnitude e fase para retornar ao domínio espacial da imagem.	25
4.1	Pontos padrões da figura de uma mão e a sua forma rotacionada em 45° . .	36
4.2	Casamento de todos os padrões selecionados ao longo da figura de uma mão e sua forma rotacionada.	36
4.3	Exemplo de casamento entre a figura de uma casa e sua forma em escala. .	36
4.4	Exemplo de texturas regulares utilizadas por Carcassoni.	41
4.5	Resultados obtidos pelo método de Carcassoni em um banco de dados homogêneo.	41
4.6	Resultados obtidos pelo método da matriz de co-ocorrência de Haralick. . .	44
5.1	Diagrama do processo de representação e consulta de imagens no modelo proposto.	48

5.2	Elementos do processo de quantização.	50
5.3	Experimentos realizados utilizando o método proposto.	55
6.1	Amostra da coleção de imagens utilizada nos experimentos.	58
6.2	Resultados obtidos pelos três métodos comparados: (a-d) mostra os resultados do método proposto, (e-h) os resultados do método de Carcassoni e (i-l) os resultados obtidos pelo método GLCM, de Haralick.	60
6.3	Resultados obtidos pelos métodos comparados utilizando a imagem de um tecido silcado como consulta.	61
6.4	Resultados obtidos pela utilização de uma imagem de papel de parede como imagem consulta.	62
6.5	Resultados obtidos pelos métodos comparados utilizando toda coleção de imagens, onde a imagem de uma paisagem é utilizada como consulta. . . .	63
6.6	Neste experimento foram obtidos resultados expressivos por todos os métodos.	64
6.7	Neste experimento uma imagem de paisagem com a figura de uma árvore em seu conteúdo foi utilizada como consulta.	65
6.8	A imagem consulta é uma rua com calçamento em pedras.	66
6.9	Imagens de edificações foram recuperadas por todos os métodos, porém o método proposto trouxe um número maior deste tipo de imagens.	67
6.10	Resultados obtidos pelos métodos comparados onde uma imagem de paisagem com ruído Gaussiano de 0 db foi utilizada como consulta.	68
6.11	Neste experimento temos como imagem consulta uma paisagem danificada por um risco branco ao longo da textura.	69
6.12	Resultados obtidos à partir da consulta feita com a imagem de um tecido danificado com manchas circulares.	70
6.13	Neste experimento a imagem de uma árvore ampliada (<i>zoom</i>) foi utilizada como consulta. O método de Carcassoni e o método proposto foram capazes de recuperar a versão normal da imagem (sem <i>zoom</i>).	71
7.1	Exemplo de imagens relevantes definidas para alguns experimentos realizados. Na coluna (a) temos as imagens consulta.	74

7.2	Curvas de PR x RE para os três métodos comparados utilizando um banco de dados homogêneo.	75
7.3	Curvas de PR x RE para os três métodos comparados utilizando todas as imagens da coleção.	76
7.4	Curvas de PR x RE para os três métodos comparados no terceiro grupo de experimentos. As imagens consulta apresentam diferentes níveis de ruído Gaussiano, estragos, etc.	76
A.1	Resultados obtidos utilizando diferentes valores para S.	80
B.1	Interface do usuário do sistema CBIR proposto.	81
B.2	<i>Ranking</i> do método proposto retornado pela submissão de uma imagem consulta.	82

Lista de Tabelas

2.1	Algumas abordagens na extração de características de baixo nível, suas vantagens e desvantagens.	16
3.1	Valores padrões para imagens utilizados em PDI.	20
4.1	Algumas características que podem ser calculadas da matriz de co-ocorrência.	43
6.1	Divisão por classe das imagens dos bancos de dados utilizados nos experimentos.	57

1 Introdução

Imagens têm atualmente um papel crucial em diversas áreas como medicina, jornalismo e propaganda, desenho arquitetônico e de engenharia, prevenção ao crime, aplicações militares, propriedade intelectual, moda e design de interiores, segurança, identificação pessoal, geo-processamento e sistemas de sensoriamento remoto, educação e treinamento, entretenimento, etc. Um impulso na revolução de imageamento digital foi dado com a expansão da utilização dos computadores, onde surgiram técnicas para captura, armazenamento, processamento e transmissão de imagens. Outro ponto marcante desta expansão foi sem dúvida a criação da *World Wide Web* em meados dos anos 90 e, seu estrondoso crescimento desde então, tornando possível o acesso as mais variadas formas de mídia em todo canto do planeta e intensificando ainda mais o estímulo à exploração de imagens digitais.

Um segmento de pesquisa em imagens que vem crescendo bastante é o da recuperação de imagens por conteúdo (cor, textura e forma), cuja extração das características da imagem é feita de forma automática, diferentemente dos métodos tradicionais de indexação que consomem bastante tempo e esforço nesta tarefa. A recuperação de imagens em grandes coleções pode ser realizada através da navegação (*browsing*) na coleção, onde o usuário inspeciona toda ou parte da coleção à procura daquelas imagens que atendam às suas necessidades de informação. Porém, a forma mais comum e eficiente é onde o usuário expressa as suas necessidades através de uma consulta, na forma de frase, palavra-chave (*keywords*) ou modelo, e fica a cargo do sistema de recuperação encontrar na coleção de imagens aquelas que supostamente satisfazem às necessidades de informação do usuário.

Existem atualmente duas abordagens principais em recuperação de imagens: uma baseada em texto (*text-based*) e outra baseada em conteúdo (*content-based*). Estas abordagens podem ser divididas em 4 categorias: baseada em atributo (*attribute-based*), baseada em anotação (*annotation-based*), baseada em reconhecimento de objetos (*object recognition-based*) e baseada em características de baixo nível (*low-level feature-based*). A Figura 1.1 mostra a distribuição das abordagens dentro da área de recuperação de imagens.

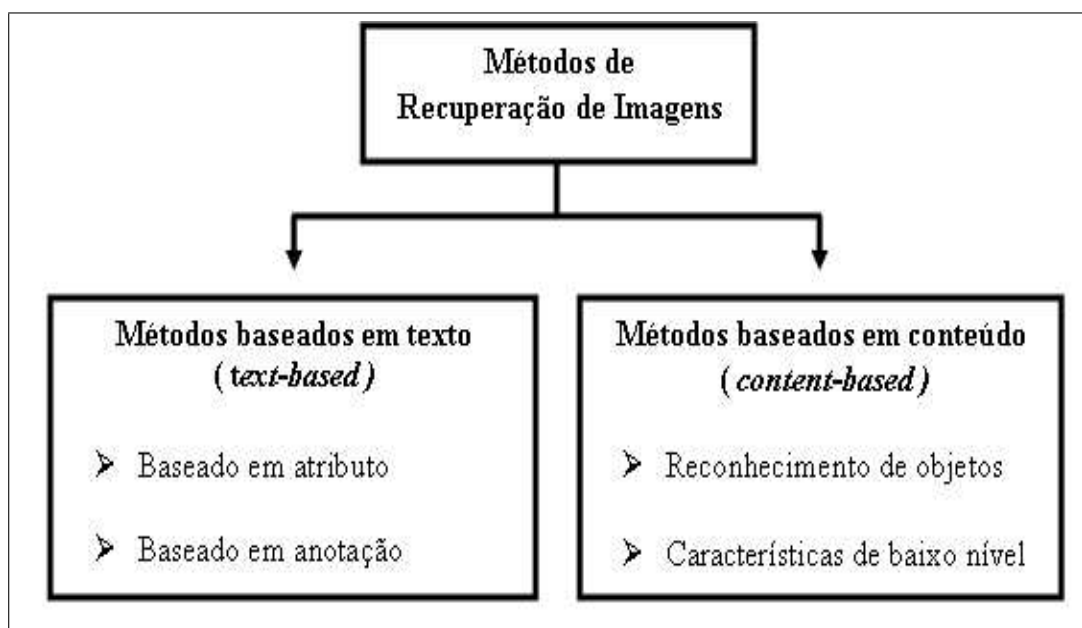


Figura 1.1: Abordagens em Recuperação de Imagens.

Na abordagem baseada em atributo, o conteúdo da imagem é representado como um conjunto de atributos extraídos manualmente da imagem, como nome, categoria, autor, assunto, origem, data de criação, etc. Esta abordagem utiliza os métodos de indexação e consulta dos sistemas de gerenciamento de banco de dados (SGBD) tradicionais, que oferecem além de mecanismos de indexação, uma linguagem de consulta bastante poderosa, o SQL (*Structural Query Language*). Um dos maiores problemas desta abordagem é que os atributos podem não descrever de forma apropriada o conteúdo da imagem.

Na abordagem baseada em anotação, a representação da imagem é feita através de um texto que descreve o seu conteúdo. Esta abordagem utiliza nas fases de indexação e recuperação das imagens, métodos tradicionais de recuperação de informação. As consultas são realizadas através de palavras-chaves ou frases, onde normalmente são utilizados operadores booleanos. Uma vantagem desta abordagem é a possibilidade da captura da abstração de uma imagem. Uma desvantagem é que o trabalho de anotação é feito de forma manual, o que normalmente requer profissionais com um alto conhecimento do domínio das imagens sob anotação, além de um tempo considerável dependendo do tamanho da coleção. Um dos grandes desafios do processo de anotação refere-se a como realizá-la de forma eficiente, completa e consistente.

A abordagem baseada em reconhecimento de objetos trabalha com o reconhecimento e interpretação de cenas, pessoas, ações, objetos, etc, presentes na imagem, onde estão envolvidos diversos conceitos relacionados à inteligência humana. As pesquisas nesta área ainda se encontram em estágios iniciais.

A abordagem baseada nas características de baixo nível (cor, textura e forma) realiza de forma automática a indexação das imagens. Este processo automático pode reduzir consideravelmente o tempo necessário para indexar uma coleção, se tornando assim viável a sua utilização em grandes coleções de imagens.

1.1 Motivação

Os sistemas de recuperação de imagens por conteúdo têm por objetivo resolver alguns dos problemas apresentados nas abordagens baseadas em texto, como a sua inviabilidade de utilização em grandes coleções de imagens, devido ao vasto e muitas vezes especializado trabalho manual requerido para descrever as imagens da coleção e, ao complexo e nem sempre completo processo de representação da imagem através de anotação, que como fica sujeito à subjetividade da percepção humana, pode interferir negativamente no processo de recuperação de imagens relevantes.

Apesar de mais de dez anos de pesquisas na recuperação de imagens por conteúdo, o interesse ainda continua crescente por parte da comunidade científica devido a ambos, o constante e crescente aumento das coleções de imagens e a inviabilidade dos métodos tradicionais de indexação. Diversas abordagens e modelos já foram propostos para tratar estes problemas, porém, há ainda espaço para a criação de novos modelos e o melhoramento dos atuais, de forma que as imagens recuperadas em uma consulta possam reproduzir com maior exatidão a percepção de similaridade humana. Com base nestas necessidades, esta dissertação tem por objetivo principal estudar alguns dos modelos de recuperação de imagens por conteúdo já propostos, com o intuito de contribuir na pesquisa e no desenvolvimento de novas técnicas. Com o intuito de alcançar o objetivo proposto, é apresentada neste trabalho uma variação ao modelo de recuperação proposto por Carcassoni et al. [9], onde os autores exploraram como o reconhecimento de texturas pode ser alcançado através da análise modal dos padrões dos picos do espectro de potência das imagens. Carcassoni et al. utilizaram o modelo de correspondência modal de Shapiro e Brady [34], localizando os centros dos clusters dos picos do espectro de potência através da informação contida na estrutura modal da matriz proximidade destes picos. Carcassoni utilizou em [9], os picos de maior magnitude como os representantes da imagem. Entretanto, considerando que duas texturas diferentes podem possuir os N primeiros picos do espectro de potência com praticamente a mesma magnitude, o que pode enfraquecer o método e levar à erros de similaridade entre imagens, é proposto nesta dissertação um modelo de representação da imagem baseada na quantização dos picos espectrais para se obter os N representantes

da imagem, ao invés da seleção dos picos de maior magnitude do espectro de potência proposto por Carcassoni. O modelo de quantização é uma variação da técnica de quantização [37] desenvolvida por Lloyd em [25].

Para comprovar a eficiência do novo modelo, diversos experimentos foram realizados utilizando um banco de dados de imagens contendo cerca de 800 imagens. Curvas de *Precision* e *Recall* foram utilizadas para auferir a qualidade do novo método, uma vez que elas são bastante utilizadas em modelos de recuperação de informações (RI) baseados em texto e vêm se tornando também um padrão na avaliação de sistemas de recuperação de imagens.

Os resultados obtidos pelo método proposto foram comparados com os do modelo de Carcassoni et al. [9] e com os do modelo da matriz de co-ocorrência de níveis de cinza (GLCM) de Haralick [17], que compara algumas características calculadas à partir das matrizes de co-ocorrência das imagens, como a energia, entropia, correlação, contraste, etc.

1.2 Organização do Trabalho

O restante desta dissertação contém 7 capítulos, dispostos da seguinte forma:

- O **Capítulo 2** apresenta uma breve introdução aos sistemas de recuperação de imagens por conteúdo, abordando algumas técnicas de extração de características (cor, textura e forma). O capítulo ainda traz algumas informações sobre ferramentas de recuperação de imagens disponíveis comercialmente.
- O **Capítulo 3** apresenta alguns fundamentos em Processamento Digital de Imagens e alguns conceitos matemáticos relacionados ao conteúdo dos modelos dispostos neste trabalho.
- O **Capítulo 4** apresenta os modelos correlacionados ao trabalho proposto: a análise modal de Shapiro e Brady [34], o método de recuperação de imagens baseado na análise espectral, de Carcassoni et al. [9], o método da matriz de co-ocorrência de níveis de cinza, de Haralick [17] e o método de quantização de Lloyd [25].
- O **Capítulo 5** apresenta o modelo proposto.
- No **Capítulo 6** apresenta alguns dos resultados experimentais realizados.

- No **Capítulo 7** é realizada a análise de performance dos métodos comparados, através das medidas de performance *Precision* e *Recall*.
- O **Capítulo 8** apresenta as conclusões finais e contribuições deste trabalho. Propostas de melhoria e continuidade estão também dispostas neste capítulo.

1.3 Considerações Finais do Capítulo

As pesquisas em Processamento Digital de Imagens (PDI) tem um papel importante em diversas áreas e aplicações. Um dos segmentos de pesquisa em PDI que vem crescendo bastante está relacionado à recuperação de imagens, onde existem duas abordagens principais: uma onde a recuperação de imagens é baseada em texto e outra baseada em conteúdo, sendo que atualmente o foco de atenção está sendo dado à abordagem baseada em conteúdo, devido principalmente à proposta de tais sistemas em extrair de forma automática as características das imagens (como cor, textura e forma), viabilizando a sua utilização em grandes coleções de imagens.

O próximo capítulo introduz a recuperação de imagens por conteúdo e seus principais aspectos, as propriedades visuais comumente exploradas (cor, textura e forma) e algumas abordagens utilizando estas propriedades. O capítulo ainda comenta algumas características dos principais sistemas comerciais em recuperação de imagens atualmente disponíveis.

2 Recuperação de Imagens por Conteúdo

2.1 Introdução

Devido ao enorme e crescente volume de informações visuais disponíveis atualmente nas mais diversas áreas como medicina, aplicações militares, comércio, desenho arquitetônico e de engenharia, desenho de moda, entretenimento, propaganda, prevenção ao crime, etc, grande atenção tem sido dada aos sistemas de recuperação de imagens por conteúdo (CBIR - *Content-Based Image Retrieval*). Nos sistemas CBIR o usuário fornece uma imagem consulta ou modelo com o intuito de encontrar imagens similares ou relevantes dentre as imagens armazenadas no banco de dados de imagens. A imagem consulta e as imagens armazenadas necessitam estar representadas, ou seja, extraídas suas características para que possam ser comparadas umas com as outras. Existem duas abordagens principais em recuperação de imagens por conteúdo para representação de imagens: o reconhecimento de objetos e as características de baixo nível. Estas abordagens também podem ser classificadas em relação às características visuais exploradas, que são:

- i) Características de **baixo nível**: estão relacionadas à visão humana e operam com as primitivas cor, textura e forma. Elas são calculadas com base nos valores dos *pixels*.
- ii) Características de **alto nível**: estão relacionadas às características semânticas, como objetos e o seu significado, interpretação de cenas, ações, sentimentos, etc.

As características de alto nível constituem um dos principais problemas em CBIR, denominado gap semântico (*semantic gap*), uma vez que estas características não podem ser facilmente capturadas por modelos matemáticos. Esta dissertação se enquadra no grupo que explora as características de baixo nível, mais especificamente com a textura da imagem.

As técnicas de recuperação de imagens por conteúdo podem ser resumidas em:

1. Extrair as características visuais (cor, textura e forma) das imagens utilizando técnicas de PDI e Visão Computacional.

2. Representar as imagens através de uma estrutura (vetor de características) contendo os padrões extraídos do seu conteúdo visual.
3. Calcular/medir a similaridade entre os padrões extraídos da imagem consulta e das imagens armazenadas no banco de dados de imagens, classificando-as de forma que as imagens do topo da lista de classificação (*ranking*) sejam as mais similares em relação à imagem consulta.

Alguns sistemas de recuperação incorporam ainda um processo de realimentação de relevância (*relevance feedback*), que consiste na interação do usuário num processo de refinamento da consulta, podendo gerar resultados ainda mais significativos do ponto de vista visual e semântico. A Figura 2.1 mostra um diagrama típico para sistemas CBIR.

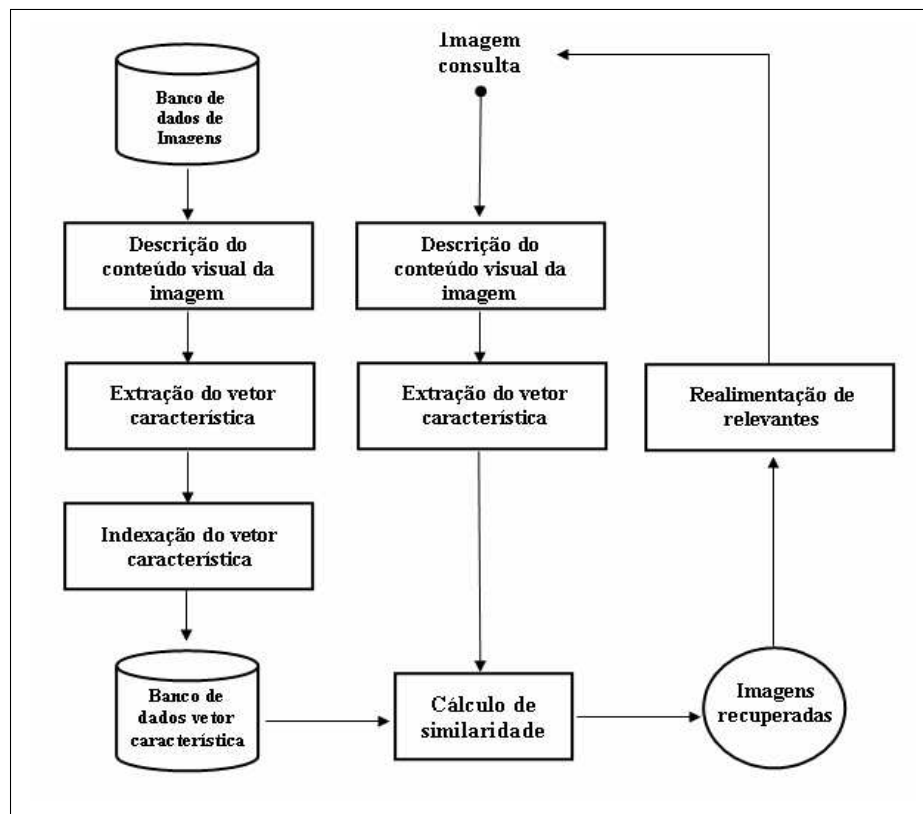


Figura 2.1: Diagrama típico de sistemas CBIR.

A recuperação de imagens tem duas linhas principais de pesquisa: a recuperação baseada em texto e a recuperação baseada em conteúdo. A indexação na recuperação baseada em texto pode ser feita por anotação ou por atributo e as consultas são realizadas por palavras-chaves ou frases. A consulta por texto em um banco de dados de imagens indexado por anotação é um processo rápido e confiável quando o processo de anotação das imagens foi realizado de forma eficiente e completa. Porém, o processo de anotação

requer considerável trabalho manual e um alto conhecimento sobre o domínio das imagens sob anotação. Outro problema deste modelo é que a anotação das imagens fica sujeita à subjetividade humana, podendo variar consideravelmente de indivíduo para indivíduo. O rico conteúdo visual de algumas imagens torna este processo ainda mais difícil, pois nem sempre é possível descrevê-las de forma completa. Considere como exemplo a imagem da Figura 2.2. Qual seria a melhor forma de descrever o seu conteúdo?

- Cidade européia?
- Montanhas?
- Inverno?
- Bela paisagem?
- Áustria?



Figura 2.2: A difícil tarefa de anotação de uma imagem rica em conteúdo.

Considerando as diferentes formas de percepção da imagem da Figura 2.2, notamos que nem sempre é possível encontrar uma descrição unânime sobre o conteúdo das imagens, o que constitui um dos principais problemas das abordagens baseadas em texto, a descrição da imagem.

As seções subseqüentes neste capítulo apresentam algumas das mais importantes aplicações nas mais diferentes áreas de pesquisa, envolvendo os sistemas de recuperação de imagens por conteúdo, bem como as formas de elaboração de consultas existentes em tais sistemas. Uma breve introdução aos atributos visuais mais explorados em CBIR e algumas das principais abordagens que exploram estes atributos visuais são sucintamente descritas. As seções subseqüentes ainda trazem alguns comentários acerca das mais importantes características encontradas em alguns dos principais softwares comerciais disponíveis em CBIR.

2.2 Aplicações em CBIR

Sistemas CBIR possuem aplicações em diversas áreas como:

1. **Medicina:** diagnóstico médico (reconhecimento de tumores, metastases, etc).
2. **Prevenção ao Crime:** reconhecimento de faces, impressão digital, etc.
3. **Militar:** reconhecimento de alvos por satélite, radares, etc.
4. **Observação Espacial:** observações por satélite para a agricultura, tráfego, desmatamento, etc.
5. **Propriedade Intelectual:** marcas de proteção legal (*trademark*), detecção de cópia de imagem, etc.
6. **Desenho de Arquitetura e Engenharia:** banco de dados CAD.
7. **Comercial:** moda, jornalismo, etc.
8. **Cultural:** galerias de arte, museus, etc.
9. **Educacional e Treinamento:** gráficos, slides, etc.
10. **Entretenimento:** foto, vídeo, filmes, etc.

As aplicações acima abrangem uma vasta lista de tópicos e compartilham informações com processamento de imagens e recuperação de informação. Dentre elas estão:

- O entendimento da necessidade de informação do usuário e o seu comportamento na busca destas informações;
- A identificação de formas apropriadas de descrever o conteúdo das imagens;
- A extração de características das imagens em seu estado original;
- Oferecer armazenamento compacto para grandes coleções de imagens;
- Consulta às imagens armazenadas de uma forma que reflita o julgamento de similaridade humano;
- Acesso eficiente às imagens armazenadas através do seu conteúdo;
- Oferecer interface humana amigável aos sistemas CBIR.

2.3 Consultas em Sistemas de Recuperação de Imagens

Em sistemas de recuperação de imagens, sejam eles baseados em texto ou conteúdo, a interface com o usuário consiste normalmente de duas partes: a primeira é da formulação da consulta e a segunda é da apresentação dos resultados. Existem basicamente três formas de formular uma consulta em sistemas de recuperação de imagens:

- i) Consulta direta* o usuário especifica uma consulta através de palavras-chaves ou características extraídas da imagem, como histograma de cores ou um descritor de textura.
- ii) Consulta por exemplo* o usuário fornece um esboço de onde as características são extraídas da mesma forma que das imagens armazenadas no banco de dados.
- iii) Navegação* o usuário explora todo ou parte do conteúdo do banco de dados, analisando imagem à imagem.

2.4 Extração de Características

Tipicamente a descrição da imagem, que é uma representação numérica da mesma, é armazenada em um vetor denominado vetor de características (*feature vector*), que pode ser multi-dimensional dependendo do número de características armazenadas. A construção do vetor de características das imagens armazenadas no banco de dados é um processo *off-line*, enquanto que o processo de consulta é um processo *on-line*. O processo de casamento (*matching*) realiza as comparações necessárias, retornando os índices das imagens recuperadas. Um outro processo se ocupa de buscar as imagens do banco pertencentes aos índices recuperados, enviando estas imagens à interface de visualização do usuário.

2.5 Atributos das Imagens

As técnicas de CBIR trabalham com o princípio de recuperar imagens da coleção de imagens cujas características extraídas mais se assemelham às da imagem consulta. O grupo de características mais explorado é o grupo de características de baixo nível (cor, textura e forma), as quais podem ser exploradas de forma individual ou coletiva, dependendo do modelo proposto.

2.6 Recuperação Utilizando Cores

As cores podem ser definidas como a sensação causada pela luz e sua interação com o olho e cérebro humano. O olho humano contém dois tipos de receptores visuais: bastonetes e cones. Os bastonetes são responsáveis pela luz fraca e são sensíveis à pequenas variações de luminosidade. Os cones são mais ativos na luz forte e são responsáveis pela visão colorida. Os cones do olho humano podem ser divididos em 3 categorias principais, que correspondem ao vermelho, verde e azul. Desta forma, as cores podem ser vistas como a combinação das cores primárias vermelho (*Red*), verde (*Green*) e azul (*Blue*).

A utilização de cores em processamento de imagem é motivada por dois fatores principais: (i) a cor é um descriptor bastante poderoso que facilita a identificação de objetos e a extração dos mesmos de uma cena e (ii) o olho humano consegue discernir entre milhares de variações de intensidades de cores, enquanto consegue distinguir apenas duas dezenas de tons de cinza.

A idéia básica em métodos de recuperação baseados em cores é a construção do histograma de cores da imagem, que mostra a proporção de *pixels* de cada cor dentro da imagem. A técnica de intersecção de histograma, que é uma das mais utilizadas, foi desenvolvida por Swain and Ballard [39]. Diversos sistemas utilizam o histograma de cores baseado em diferentes modelos de cores como RGB, HSV, HLS, etc, como em [10, 27, 38].

Em [2], é utilizado como descritor da imagem um vetor contendo a distribuição de cores baseada nos coeficientes *wavelets*. Bourgeois em [6], utiliza o gradiente de cores como padrões da imagem, que leva em consideração a curvatura das regiões da imagem e as cores encontradas em regiões adjacentes. Outros trabalhos utilizando cores incluem o correlograma de cores [19], que guarda a correlação espacial da cor, a matriz de adjacência [10], etc.

• Histograma de Cores

O histograma de cores é um método que descreve o conteúdo de cores de uma imagem. Ele conta o número de ocorrências de cada cor dentro de uma imagem. O histograma de cores é invariante a translação e rotação e pode até se tornar invariante à escala com a normalização do histograma. O principal problema do método de histograma de cores é que ele utiliza somente a informação da cor, desconsiderando as informações de textura e forma, o que pode levar a erros de similaridade. A Figura 2.3 mostra a imagem de uma edificação e seu respectivo histograma de cores.

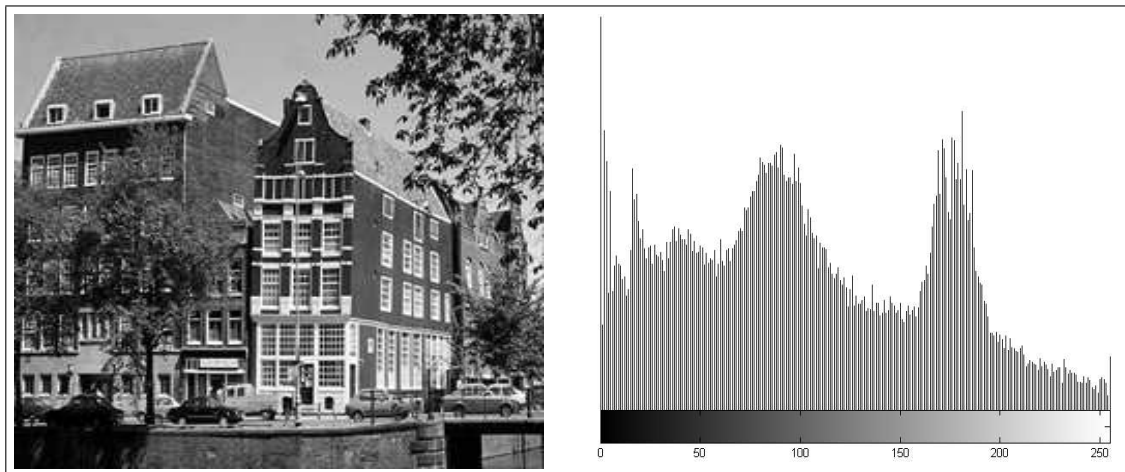


Figura 2.3: Imagem de uma edificação e seu histograma de cores (à direita).

2.7 Recuperação Utilizando Texturas

A análise de textura pode ser dividida em quatro categorias principais:

1. **Segmentação de Texturas:** consiste em encontrar texturas diferentes em uma imagem. A dificuldade consiste em conhecer, à priori, quantas texturas diferentes e seus tipos existem na imagem.
2. **Classificação de Texturas:** consiste em dizer à qual categoria uma dada textura pertence, como por exemplo pedras, grama, tecidos, nuvens, etc. Redes Neurais e Bayesianas podem ser utilizadas para reconhecer e classificar os padrões das texturas.
3. **Forma à partir da Textura:** consiste em encontrar as distorções da textura quando da projeção de objetos do mundo real, que são em 3D, em um plano de imagem 2D.
4. **Síntese da Textura:** consiste em sintetizar uma nova textura à partir de uma textura exemplo, ou seja, os algoritmos devem ser capazes de, à partir de uma textura exemplo, gerar uma quantidade ilimitada de novas imagens que não são necessariamente como a imagem original mas, que será percebida pelos humanos como se fosse a mesma textura.

As três principais abordagens utilizadas em PDI para representar a textura de uma região são: (i) estatística, (ii) estrutural e (iii) espectral. A abordagem estatística considera a distribuição dos tons de cinza e o inter-relacionamento entre eles, como as matrizes

de co-ocorrência [17] e modelos fractais [20]. As técnicas estruturais, por outro lado, descrevem a textura através de primitivas estruturais (círculos, hexágonos, pontos, etc) e sua disposição na imagem. Este método é mais eficiente quando aplicado à texturas muito 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*. Exemplos incluem filtros de Gabor [26] e a transformada de *wavelets* [22], etc.

Métodos de recuperação de imagens baseados em texturas envolvem a exploração de áreas que possuam cores em comum, como mar e céu, grama e folhas, etc, ou seja, apresentam padrões visuais que têm propriedades de homogeneidade mas não implicam na presença de uma única cor ou intensidade. O cálculo de similaridade é geralmente realizado utilizando a relação de brilho, aspereza, direção e periodicidade da textura.

• Modelos de Wavelets

As texturas podem ser modeladas como padrões quase-periódicos com representação espaço/frequência. A transformada de *Wavelet* transforma a imagem em uma representação multi-escala contendo ambas características, espacial e frequência. De acordo com esta transformação, uma função, que pode representar uma imagem, uma curva, um sinal, etc, pode ser descrita em termos de nível de aspereza.

A transformada de *Wavelet* utiliza funções conhecidas como *wavelets*. As *wavelets* são funções finitas no tempo. A transformada de Fourier converte um sinal em uma série contínua de ondas de seno, sendo cada uma delas de frequência e amplitude constante e infinita duração, sendo que a maioria dos sinais do mundo real (como música, imagens, etc), tem uma duração finita e mudanças abruptas na frequência. Já a transformada de *Wavelet* converte um sinal dentro de séries de *wavelets*, que podem ser armazenadas mais eficientemente devido ao tempo finito, aproximando-se mais dos sinais do mundo real.

Alguns exemplos de *wavelets* são Coiflet [11], Haar [16] e Daubechies [11], sendo que a Haar é a mais simples e utilizada, enquanto que a Daubechies tem estrutura fractal que é vital para as atuais aplicações envolvendo *wavelets*.

2.8 Recuperação Utilizando Formas

A recuperação de imagens utilizando formas consiste em identificar formas de objetos ou regiões presentes na imagem através da segmentação. O cálculo de similaridade entre as

imagens é feito analisando o conjunto de formas da imagem consulta e os conjuntos de formas das imagens armazenadas. A Figura 2.4 mostra um exemplo de segmentação de imagem utilizando o algoritmo de detecção de bordas de Canny.



Figura 2.4: Imagem original (à esquerda) e a mesma após o algoritmo de detecção de bordas de Canny.

Alguns modelos representam a forma pela excentricidade, circularidade [27], momentos invariantes [12], etc. Sangineto, em [31], propõe a junção de técnicas clássicas de alinhamento para a localização de formas. Esta estratégia inovadora é capaz de lidar com casamentos inexatos entre a forma procurada e a encontrada.

• Momentos Invariantes

Momentos invariantes, derivados por Hu [18], são úteis quando comparando duas imagens entre si ou com uma imagem padrão. Momentos invariantes são largamente utilizados em reconhecimento de padrões e análise de imagens. Existem duas abordagens principais: momentos geométricos e momentos de Zernike.

Momentos geométricos descrevem uma imagem como uma função numérica em relação à referência do eixo e é definida como:

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

onde $p, q = 0, 1, 2, \dots, \infty$ e $f(x, y)$ é a função de distribuição de densidade da imagem.

A forma de uma imagem pode ser representada em termos de sete funções definidas nos momentos invariantes ($\phi_1 - \phi_7$). As seis primeiras funções ($\phi_1 - \phi_6$) são invariantes à rotação e a última ϕ_7 é invariante à rotação e distorção.

Os momentos invariantes $\mu(i, j)$ de uma imagem $f(x, y)$ é dado por:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j f(x, y)$$

onde $\bar{x}$ e $\bar{y}$ indicam o centro de massa da forma.

Definindo $\gamma = (i + j)/2 + 1$ e $\eta_{ij} = \mu_{ij}/\mu_{00}^\gamma$, as funções invariantes são obtidas por:

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \\ \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) [3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + 3(\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\ &\quad [3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ \phi_6 &= (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2] - 3(\eta_{21} + \eta_{03})^2 - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \\ &\quad [3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \end{aligned}$$

Momentos de Zernike utilizam funções de bases ortogonais e são menos sensíveis à ruídos do que momentos geométricos. Eles são construídos utilizando um conjunto de polinômios complexos. Momentos de Zernike bidimensional são dados por:

$$A_{mn} = \frac{m+1}{\pi} \int_x \int_y f(x, y) [V_{mn}(x, y)]^* dx dy$$

onde

$$x^2 + y^2 \leq 1$$

$$m = 0, 1, 2, \dots, \infty$$

$f(x, y)$ é a função sendo descrita

* denota o complexo conjugado

n é um inteiro que representa a dependência angular ou rotação.

2.9 Resumo das Características de Baixo Nível

As características de baixo nível da imagem que têm sido amplamente utilizadas em sistemas CBIR, suas vantagens e limitações segundo Vailaya [41], podem ser encontradas na

tabela 2.1.

Atributo	Abordagens	Vantagens	Desvantagens
Cor	Histogramas e momentos de cor	São facilmente extraídos da imagem com baixo custo computacional. Representam as propriedades globais da imagem	Não representam a informação local.
Textura	Matrizes de co-ocorrência, descritores de Fourier, filtros de Gabor, etc.	Podem ser extraídas de forma automática da imagem. Detém informações globais e locais.	Podem ser computacionalmente caros e de difícil definição.
Forma	Aproximação poligonal, momentos invariantes e descritores de Fourier	Conseguem obter um alto nível de abstração em termos morfológicos dos objetos locais.	Não podem ser extraídos automaticamente sem a necessidade de bons algoritmos de segmentação.

Tabela 2.1: Algumas abordagens na extração de características de baixo nível, suas vantagens e desvantagens.

2.10 Softwares Comerciais CBIR Disponíveis

Após mais de uma década de intensa pesquisa em CBIR, alguns dos modelos propostos vêm deixando os laboratórios de pesquisa e se transformando em ferramentas comerciais. Dentre estas ferramentas podemos destacar algumas mais conhecidas e utilizadas como a QBIC [14, 27], a VisualSEEK [35] e WebSEEK [36] e a MARS [29, 30].

2.10.1 QBIC

Produzido pela IBM, a QBIC foi a primeira ferramenta disponível comercialmente para a recuperação de imagens e é provavelmente a melhor e mais conhecida dentre todas. Esta ferramenta oferece mecanismos de recuperação de imagens através da combinação de cor, textura, forma e palavra-chave. As consultas podem ser formuladas pela seleção de uma paleta de cores, imagem consulta ou através do esboço de uma forma desejada. Os padrões de textura são baseados no modelo de Tamura [40], composta por características como aspereza, contraste e direção. Os padrões de forma explorados são a área, circularidade, excentricidade, momentos de invariância, etc. Histogramas de cores são utilizados para recuperação baseada em cores.

2.10.2 MARS

A MARS (*Multimedia Analysis and Retrieval System*) foi criada pela Universidade de Illinois e posteriormente desenvolvida pela Universidade da Califórnia, ambas nos Estados Unidos. A principal característica desta ferramenta é de organizar várias características visuais dentro de uma arquitetura de recuperação mais significativa. A MARS suporta as características visuais de baixo nível e implementa uma arquitetura de realimentação de relevância em vários níveis durante o processo de recuperação das imagens. A MARS oferece também a possibilidade da consulta ser realizada através da descrição textual ao invés de uma imagem.

2.10.3 VisualSEEK e WebSEEK

Ambas ferramentas foram produzidas pela Universidade da Columbia, em Nova Iorque. VisualSEEK é um mecanismo de busca visual, enquanto que a WebSEEK é uma ferramenta de busca voltada para Web, onde é possível trabalhar com texto ou imagem. As buscas podem ser realizadas por regiões de cor, forma, localização espacial e por palavra-chave. Realimentação de relevância também é utilizada para refinar a consulta.

2.11 Considerações Finais do Capítulo

A recuperação de imagens por conteúdo consiste em propor soluções para o problema de recuperação de imagens em grandes bancos de dados utilizando as características visuais das imagens, em contraste com a maioria dos mecanismos de busca na Web (Google, por exemplo) que faz uso da recuperação de imagens baseada em texto (*text-based*), na qual as imagens são recuperadas baseando-se nos seus rótulos, descrições e texto ao redor da imagem. Apesar de ser um método rápido e confiável, ele está completamente dependente do texto. Em outros modelos, a recuperação é baseada em imagens anotadas, que requer que cada imagem do banco de dados passe pelo nem sempre trivial processo de anotação.

As pesquisas em CBIR avançaram muito desde o seu início, em meados dos anos 90, onde duas abordagens principais são exploradas: o reconhecimento de objetos e as características de baixo nível (cor, textura e forma). O problema consiste em, dada uma imagem consulta, extrair as suas propriedades visuais e verificar a similaridade (*matching*) entre a estrutura que representa a imagem consulta e aquelas estruturas que representam as imagens armazenadas no banco de imagens (vetor de características), retornando aquelas

imagens cujas estruturas sejam mais similares à da imagem consulta. Afim de alcançar tal objetivo, os modelos CBIR aplicam técnicas de Processamento de Imagens e Visão Computacional para indexação e recuperação de imagens onde o principal desafio é conseguir reproduzir em modelos matemáticos o conceito de similaridade adotado pelo cérebro humano, de forma que os resultados alcançados se aproximem cada vez mais do julgamento de similaridade do homem, satisfazendo desta forma as necessidades de informações do usuário.

No próximo capítulo serão apresentados alguns conceitos em imagem digital e fundamentos matemáticos inerentes aos trabalhos discutidos ao longo dessa dissertação.

3 Fundamentos Teóricos

3.1 Introdução

O Processamento Digital de Imagens, seja ele com a finalidade de melhorar ou interpretar imagens é de grande interesse da comunidade científica. Um segmento de pesquisa que compartilha informações entre PDI e Visão Computacional e que vem crescendo bastante é o que estuda a recuperação de imagens por conteúdo, mais conhecido como CBIR.

Os algoritmos dos sistemas CBIR utilizam uma gama enorme de conceitos em imageamento digital e modelos matemáticos. O conhecimento prévio de alguns destes conceitos e definições facilitará na compreensão dos métodos aqui apresentados. Desta forma, o corrente capítulo expõe de forma sucinta alguns dos principais conceitos descritos neste trabalho, divididos em conceitos em imageamento digital e conceitos matemáticos. Começamos pela definição de um elemento importante em imagens digitais, a *pixel*.

3.2 Conceitos em Imageamento Digital

3.2.1 Pixel

A palavra *pixel* vem da combinação de *picture* e *element*, que significa elemento de pintura e representa a menor unidade de informação em uma tela ou imagem e que possui uma cor.

3.2.2 Imagem Digital e Valores Padrões

Uma imagem digital monocromática pode ser definida através de uma função bidimensional de intensidade da luz $f(x,y)$. O valor das coordenadas x e y refere-se a intensidade de nível de cinza ou brilho da imagem naquele ponto. Desta forma, uma imagem digital pode ser considerada como sendo uma matriz cujos índices de linhas e colunas represen-

tam um ponto na imagem e, o valor do elemento (*pixel*) da matriz representa o nível de cinza naquele ponto. A Figura 3.1 mostra um exemplo de imagem digital de dimensão 256×256 , com 256 níveis de tonalidade de cinza.

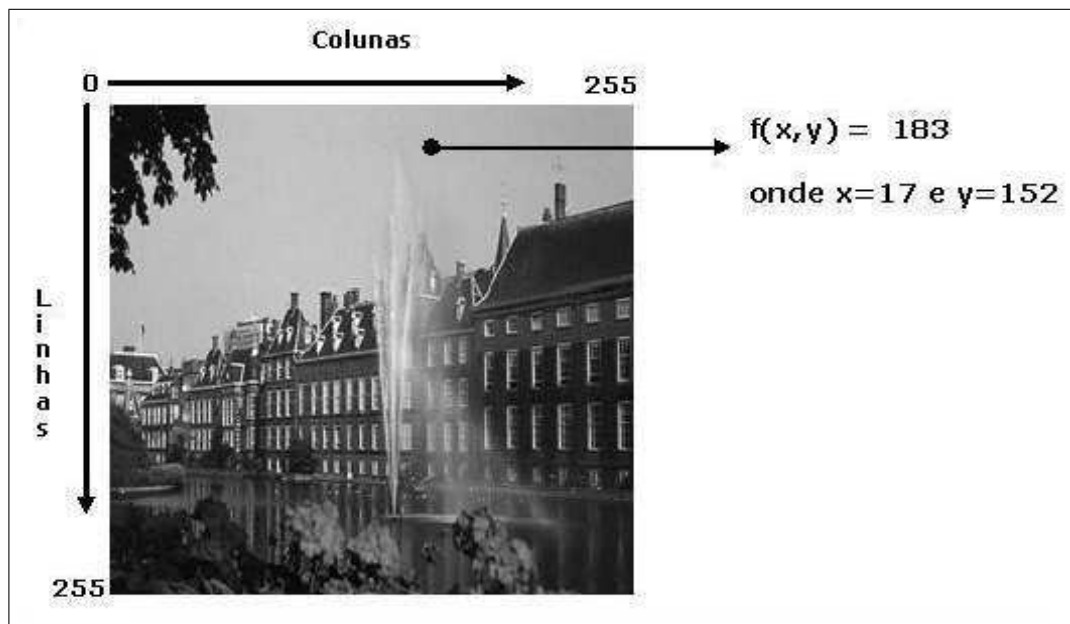


Figura 3.1: Exemplo de imagem digital de dimensão 256×256 .

Alguns dos valores padrões para imagens mais utilizados em Processamento Digital de Imagens estão dispostos na tabela 3.1. Estes valores foram estabelecidos por padronizações de vídeo, necessidades dos algoritmos, etc.

Parâmetro	Símbolo	Valores típicos
Linhas	M	256, 512, 525, 625, 1024, 1035
Colunas	N	256, 512, 768, 1024, 1320
Escala de níveis de cinza	L	2, 64, 256, 1024, 4096

Tabela 3.1: Valores padrões para imagens utilizados em PDI.

3.2.3 Definição de Textura e seus Padrões

Textura é um fenômeno bastante difundido, fácil de reconhecer e difícil de definir. Não existe ainda uma definição universalmente aceita, porém em alguns pontos há uma concordância entre os pesquisadores. São eles:

- Dentro de uma textura existe uma variação significativa nos níveis de intensidade de cor entre *pixels* vizinhos.

- Textura é uma propriedade homogênea em alguma escala espacial maior do que a da resolução da imagem.

Alguns pesquisadores descrevem uma textura como sendo uma grande quantidade de objetos pequenos, como por exemplo grama, folhagem, galhos, cabelos, etc. Há ainda aqueles que consideram que superfícies com padrões comuns que parecem uma grande quantidade de pequenos objetos, como por exemplo, manchas de animais como leopardos e chacais, listras de animais como zebras, padrões em casca de árvores, madeira, pele, etc.

Existem outras definições para textura em Processamento Digital de Imagens e Visão Computational presentes na literatura. Vejamos algumas delas:

Sklansky (1978), *Image Segmentation and Feature Extraction*: “Uma região, em uma imagem, tem uma textura constante se um conjunto de estatísticas locais ou outras propriedades locais da função da Figura são constantes, de lenta variação, ou de variação aproximadamente periódica.”

Jahne (1995), *Digital Image Processing*: “Modelos que caracterizam objetos são chamados texturas em Processamento de Imagens.”

Wilson e Spann (1988), *Image Segmentation and Uncertainty*: “Regiões de Textura são padrões espacialmente estendidos baseados na maior ou menor repetição precisa de alguma unidade celular (texton ou subpadrão).”

Gonzalez e Woods (1992), *Digital Image Processing*: “Nós intuitivamente vemos este descritor como provedor de uma medida de propriedades tal como suavidade, aspereza e regularidade.”

Considerando as definições acima, é seguro afirmar que a Figura 3.2 apresenta um exemplo de textura.

As texturas podem ser classificadas pela sua forma e geometria como regulares e irregulares e com base na sua distribuição de cores como homogêneas e não homogêneas. As texturas regulares apresentam um padrão repetitivo de formas, enquanto que as texturas homogêneas apresentam áreas de cor uniforme e poucas transições de tons na imagem, ou seja, uma maior regularidade e suavidade na distribuição de cores. A Figura 3.3 apresenta exemplos de texturas regulares, onde é possível perceber os padrões repetitivos presentes nas texturas.



Figura 3.2: Exemplo de textura.

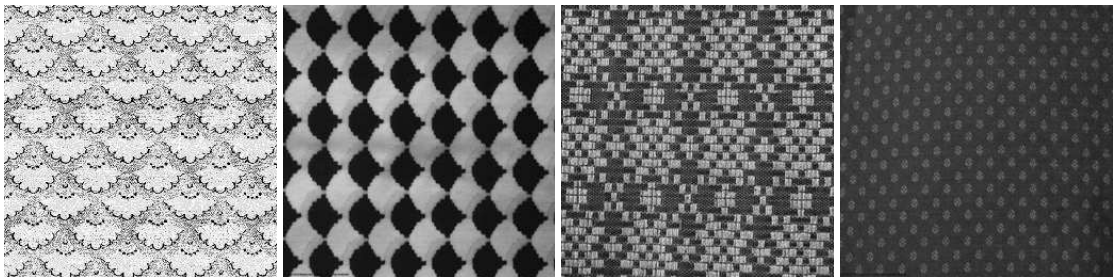


Figura 3.3: Exemplo de texturas regulares.

A Figura 3.4 apresenta exemplos de texturas não regulares, onde as formas e cores presentes na textura não formam padrões repetitivos.



Figura 3.4: Exemplo de texturas não regulares.

3.3 Conceitos Matemáticos

3.3.1 Transformada de Fourier

Imagens normalmente mostram a variação de brilho ou cor no domínio espacial. Diversas técnicas de processamento de imagens trabalham diretamente com esta representação, manipulando as informações através de operações no espaço. Uma outra forma de representação é através do domínio da frequência, onde a imagem é representada pela variação de frequência da cor ou brilho. A transformada de Fourier (TF) é utilizada para converter uma imagem de seu domínio espacial para o domínio de frequência. Em diversas aplicações é desejável e vantajoso, tanto na performance quanto na complexidade dos algoritmos, trabalhar com a imagem no domínio de frequência. Desta forma, a transformada de Fourier oferece uma ampla escala de aplicações em processamento de imagens, que incluem reconhecimento de padrões, descrição de imagens, filtragem, segmentação, compressão de arquivos, etc. A TF decompõe uma imagem dentro de componentes seno e cosseno de diferentes frequências, produzindo uma imagem no domínio de frequência (ou Fourier) no qual cada ponto representa uma frequência particular dentro da imagem. As baixas frequências do espectro são responsáveis pela variação de tom mais suaves, enquanto que as altas frequências são responsáveis pelos detalhes da imagem.

A transformada de Fourier foi uma homenagem ao físico francês Jean Baptiste Joseph Fourier (1768-1830) e é uma ferramenta largamente empregada em processamento de sinais, processamento de sons e em processamento de imagens, sendo que nesta última a TF pode ser utilizada quando queremos conhecer as frequências espaciais de um determinado padrão na imagem.

3.3.1.1 Transformada Discreta de Fourier

A transformada discreta de Fourier (TDF) refere-se à transformada de Fourier quando aplicada a um sinal digital (discreto) ao invés de um sinal analógico (contínuo). A TDF bidimensional de uma imagem quadrada $f(x, y)$, $N \times N$, é definida por:

$$F(u, v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) e^{-j \frac{2\pi}{N} (ux+vy)} \quad (3.1)$$

onde j é a unidade imaginária e e , o número de Euler.

A volta ao domínio espacial à partir do domínio de frequência é possível através da

transformada inversa de Fourier, dada por:

$$f(x, y) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} F(u, v) e^{j \frac{2\pi}{N} (ux+vy)} \quad (3.2)$$

A transformada discreta de Fourier é bastante útil pois ela revela periodicidades existentes na imagem.

3.3.1.2 Algumas Propriedades da Transformada Bidimensional de Fourier

- **Separabilidade:** esta propriedade nos mostra que o par de transformadas discretas de Fourier $F(u, v)$ e $f(x, y)$, pode ser obtido em dois passos separados, considerando-se duas operações sucessivas da TF unidimensional. Em outras palavras, a função bidimensional $F(u, v)$ é obtida pela transformação em cada linha de $f(x, y)$ e o resultado é multiplicado pelo número total das mesmas, N , obtendo-se $F(x, v)$. $F(u, v)$ é obtida ao aplicar uma transformada ao longo de cada coluna de $F(x, v)$.
- **Translação:** esta propriedade nos mostra que a multiplicação de $f(x, y)$ pelo termo exponencial, resulta num deslocamento na frequência para um ponto (u_o, v_o) . De maneira análoga, se multiplicarmos a transformada $F(u, v)$ pelo mesmo termo exponencial e tomarmos a transformada inversa, efetuaremos um deslocamento espacial da origem $(0, 0)$ para o ponto (x_o, y_o) .
- **Periodicidade:** esta propriedade nos mostra que se $f(x, y)$ é periódica, somente um período é necessário para especificar completamente $F(u, v)$ no domínio da frequência. O mesmo se aplica a $f(x, y)$ no domínio espacial.
- **Rotação:** esta propriedade nos mostra que uma rotação em $f(x, y)$ por um ângulo θ , produz a mesma rotação em $F(u, v)$ e vice-versa.
- **Teorema da Convolução:** o teorema da convolução é provavelmente uma das ferramentas mais eficazes na análise em frequência. A importância da convolução no domínio da frequência consiste no fato que se $f(x)$ tem a transformada de Fourier $F(u)$ e $g(x)$ tem sua transformada de Fourier $G(u)$, então $f(x) * g(x)$ tem $F(u)G(u)$ como transformada, ou seja: $f(x) * g(x) \Leftrightarrow F(u)G(u)$, o que mostra que a convolução no domínio espacial pode ser obtida pela transformada inversa do produto $F(u)G(u)$. O resultado pode ser estendido para o domínio da frequência, ou seja: $f(x)g(x) \Leftrightarrow F(u) * G(u)$.

3.3.1.3 Transformada Rápida de Fourier

A transformada rápida de Fourier (TRF) é uma versão mais rápida da TDF. A TRF pode ser aplicada quando a dimensão da amostra é uma potência de 2. O cálculo da TFR realiza aproximadamente $N * \log 2(N)$ operações, enquanto que a TDF realiza aproximadamente N^2 operações. Portanto, a TRF é significativamente mais rápida que a TDF. A TRF utiliza a abordagem *divide and conquer* (dividir para conquistar), que consiste em trabalhar recursivamente, quebrando um problema dentro de dois outros sub-problemas menores do mesmo tipo, até que o problema possa ser resolvido diretamente. As soluções de todos os sub-problemas são então combinados para a solução do problema original.

3.3.1.4 Fase e Magnitude do Espectro

Em geral, a transformada de Fourier é uma função complexa $F(u)$ e pode ser expressa em termos de funções reais como $F(u) = R(u) + jI(u)$, onde $R(u)$ é a parte real e $I(u)$ a parte imaginária. A magnitude de $F(u)$, também conhecida como espectro de Fourier, é definida por $|F(u)|$, enquanto que a fase de $F(u)$ é dada por $\theta(u)$.

No processamento de imagens, normalmente é exibida somente a magnitude da transformada de Fourier, que contém a maioria das informações da estrutura geométrica da imagem no domínio espacial. Porém, para converter a imagem de volta ao domínio espacial à partir do domínio de frequência, precisamos também das informações contidas na fase. A Figura 3.5 mostra uma imagem exemplo, a sua magnitude e a sua fase. Maiores detalhes pode ser encontrado em [15].

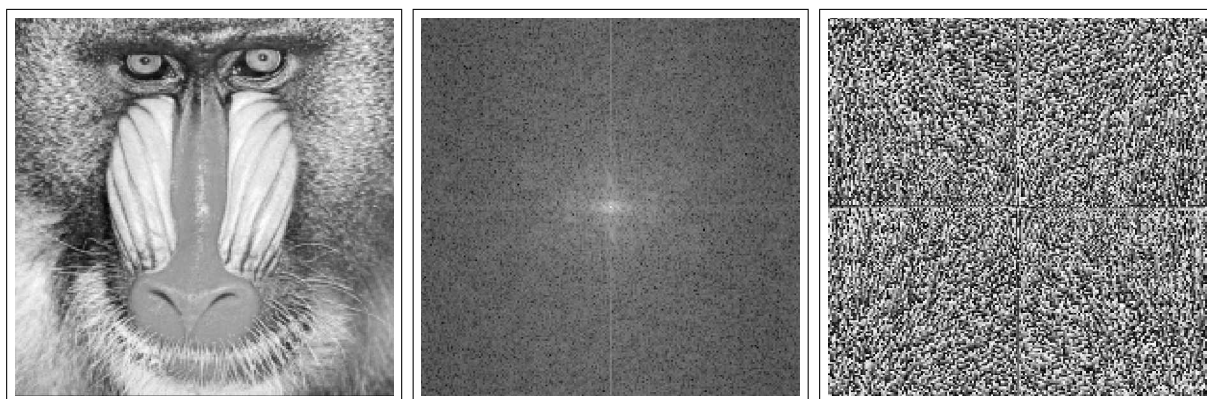


Figura 3.5: Imagem original (à esquerda), a sua magnitude do espectro (centro) e a sua fase (à direita). A transformada inversa de Fourier utiliza ambas, a magnitude e fase para retornar ao domínio espacial da imagem.

3.3.2 Base Ortonormal

Um subconjunto $\{v_1, \dots, v_k\}$ de um espaço vetorial V é chamado ortonormal, se e somente se, o produto escalar $\langle v_i, v_j \rangle = 0$ se $i \neq j$ e $\langle v_i, v_j \rangle \neq 0$ se $i = j$. Além disso $\|v_i\| = 1$. Isto é, os vetores são mutuamente perpendiculares e unitários.

Propriedade: n vetores não-nulos e ortogonais dois a dois em um espaço de dimensão n são linearmente independentes.

Base: Uma base de V é um subconjunto finito $B \subset V$ para o qual as seguintes condições se verificam:

- (a) $[B] = V$, onde $[B]$ é o conjunto gerado por B .
- (b) B é linearmente independente.

3.3.3 Autocorrelação

Chama-se autocorrelação a medida de similaridade de distribuição de amostras, que é calculada pela soma dos produtos entre o conjunto de dados e suas versões deslocadas, ou seja, a autocorrelação é o resultado da correlação de uma amostra com ela mesma. Ela é uma propriedade importante que pode ser utilizada para estimar a quantidade de regularidade presente na imagem.

A função de autocorrelação de uma imagem $f(x, y)$ de dimensão $M \times N$ é dada por:

$$r_f(l, k) = \frac{1}{MN} \sum_{m=-M+1}^{M-1} \sum_{n=-N+1}^{N-1} f^*(x, y) f(x + l, y + k)$$

3.3.4 Espectro de Potência

O espectro de potência de uma imagem é dado pela transformada discreta de Fourier da função de autocorrelação da imagem. O espectro pode ser obtido através do uso de estimadores, como o Periodograma e o estimador de Blackman-Tukey. Ambos são estimadores clássicos baseados na análise de Fourier, sendo que o estimador do Periodograma é considerado um estimador pobre pois tem baixa performance. O estimador de Blackman-Tukey é dado por:

$$P_{BT}(u, v) = \sum_{m=-M+1}^{M-1} \sum_{n=-N+1}^{N-1} r_I(m, n) w(m, n) \exp \left[-j2\pi \left(\frac{mu}{M} + \frac{nv}{N} \right) \right],$$

onde m e n são coordenadas espaciais, u e v são as variáveis de frequência, $r(m, n)$ é a função de autocorrelação da imagem I e $w(m, n)$ é uma função janela.

O estimador BT utiliza uma função janela na função de autocorrelação para reduzir a variância do espectro. Muitas funções janela podem ser utilizadas no estimador de Blackman-Tukey, como a janela retangular, de Bartlett e de Hanning. Maiores detalhes sobre estimação espectral pode ser encontrado em [21].

3.3.5 Autovalor e Autovetor

Existem diversas aplicações práticas tanto na ciência como em engenharia envolvendo o uso de autovalores.

Definição: Seja V um espaço vetorial de dimensão finita sobre um corpo K : dado um operador linear $T : V \rightarrow V$, um vetor $v \in V$ é dito ser um **autovetor** de T se existe $\lambda \in K$ tal que $T(v) = \lambda(v)$. O número real λ é denominado **autovalor** de T associado ao autovetor v .

Para calcularmos os autovalores e autovetores, consideremos como exemplo uma matriz A de dimensão $M \times M$ associada ao operador T e I a matriz identidade. O polinômio de ordem M definido por $\det(A - \lambda I) = 0$, é chamado de polinômio característico ou equação característica de A . As M raízes do polinômio característico são os autovalores de A . Cada autovalor possui um autovetor u correspondente (diferente de 0) que pode ser encontrado resolvendo-se o sistema descrito por $(A - \lambda I)u = 0$.

3.3.6 Norma

Seja V um espaço euclidiano com o produto interno $(u, v) \rightarrow \langle u, v \rangle$. Dado um vetor $u \in V$, chama-se de norma de u a função que associa a cada vetor u , o número real positivo dado por: $\|u\| = \sqrt{\langle u, u \rangle}$.

3.3.7 Similaridade

A similaridade é uma interpretação de semelhança entre uma imagem e outra. Similaridade ou relevância constituem um dos principais problemas em RI. Uma vez que as características de uma imagem tenham sido extraídas e armazenadas no vetor de características, faz-se necessária a interpretação deste conjunto, sendo que a forma mais adotada é através de uma função de similaridade, que normalmente é baseada em al-

guma medida de distância. Diferentes medidas de distância entre dois vetores, x e y de $\mathbb{R}^m$, estão atualmente disponíveis. Algumas das mais utilizadas e conhecidas são dadas à seguir [42]:

Distância Minkowski

$$d(x, y) = \left[\sum_1^m |x_i - y_i|^\lambda \right]^{\frac{1}{\lambda}}$$

Se $\lambda = 2$, a distância Minkowski é equivalente à distância Euclidiana.

Distância Euclidiana

$$d(x, y) = \sqrt{\sum_1^m (x_i - y_i)^2}$$

Distância Manhattan

$$d(x, y) = \sum_1^m |x_i - y_i|$$

Distância Canberra

$$d(x, y) = \sum_1^m \frac{|x_i - y_i|}{|x_i + y_i|}$$

Distância Chebyshev

$$d(x, y) = \max_1^m |x_i - y_i|$$

Distância Mahalanobis

$$d(x, y) = \sqrt{x - y^T S^{-1} (x - y)}$$

onde S é a matriz de covariância

As normas matriciais são definidas de forma análoga.

3.4 Considerações Finais do Capítulo

Neste capítulo foram apresentados alguns conceitos envolvendo o Processamento Digital de Imagens e alguns fundamentos matemáticos que serão úteis na compreensão dos modelos apresentados neste trabalho.

No próximo capítulo serão apresentados alguns trabalhos que serviram de inspiração para o método proposto.

4 Trabalhos Correlatos

4.1 Introdução

A análise, classificação e recuperação de imagens têm sido ao longo dos últimos anos alvo de intensa pesquisa em Visão Computacional e Processamento Digital de Imagens. O processo de encontrar imagens similares em uma coleção pode ser visto como uma tarefa relativamente fácil e simples do ponto do vista do observador humano, porém, torna-se um desafio quando um modelo computacional é utilizado com o mesmo propósito.

Sistemas de recuperação de imagens por conteúdo, CBIR, obtém a representação de uma imagem explorando as suas características visuais cor, textura e forma. Esta representação deve ser a menor possível para que seja eficiente o processo de verificação de similaridade e, preferencialmente invariante às transformações geométricas como rotação, escala e translação, bem como insuscetíveis à presença de ruído, estrago e outras irregularidades ao longo da imagem.

Diversas abordagens em recuperação de imagens foram propostas ao longo dos últimos anos. Gerald e Simon em [32], propuseram um algoritmo de recuperação baseado em dois descritores, um que explora os padrões de cores e outro que explora os padrões de formas. O descritor de cores é baseado na matriz de co-ocorrência de cores, enquanto que a informação sobre a forma é representada através do histograma dos padrões. Combinando estes dois descritores, o algoritmo conseguiu resultados efetivos e eficientes na recuperação de imagens, onde a performance deste método superou alguns dos mais populares algoritmos baseados em *pixels* (*pixel-based*), como o histograma de cores, vetores de conexão de cores e o correlograma de cores.

Zhang [43] também propôs um sistema CBIR que explorava dois padrões da imagem: cor e textura. Combinado estes dois descritores, o método obteve uma boa performance na recuperação de imagens. Dada uma imagem consulta, o algoritmo classificava (*ranking*) as imagens do banco de dados utilizando primeiramente os padrões de cores das imagens.

As imagens do topo deste *ranking* eram novamente classificadas considerando desta vez os padrões da textura das imagens. O descritor de cores construía um histograma de cores baseado no espaço de cores CIEL *u*v*, enquanto que o descritor de texturas era baseado em filtros de Gabor. Basicamente, filtros de Gabor são um grupo de *wavelets*, onde cada *wavelet* captura a energia de uma frequência e direção específica. Expandir um sinal utilizando esta base proporciona uma descrição de frequência localizada, capturando desta forma os padrões e energia local do sinal.

Lin et al., propuseram em [23] um sistema eficiente e robusto de recuperação de imagens baseado em formas (*shape-based*). Os autores utilizaram o método de detecção de bordas *Prompt* [24] e representaram as formas extraídas através de um método de representação denominado de MCS, que era invariante a translação, rotação e escala. Os resultados do método proposto pelos autores mostraram uma boa taxa de recuperação quando da presença de imagens com um nível modesto de deformação.

Shapiro e Brady em [34], propuseram um modelo de correspondência de pontos padrões através da análise modal de formas. Como primeiro passo do algoritmo, m pontos padrões eram escolhidos para representar cada imagem. Então, uma matriz quadrada $m \times m$, denominada matriz proximidade era criada para armazenar as distâncias entre os pontos padrões. Em seguida era aplicada a análise modal na matriz proximidade. O passo final consistia em associar em uma matriz, a relação entre as matrizes modais de duas imagens, verificando a similaridade entre os pontos padrões.

Com base nas idéias propostas por Shapiro e Brady em [34], Carcassoni, Ribeiro e Hancock apresentaram em [9] um método de recuperação de imagens baseado na análise modal dos centros de clusters do espectro de potência das imagens. Em seu trabalho anterior [8], Carcassoni utilizou a análise modal da matriz proximidade de Shapiro e Brady [34], explorando a correspondência de conjuntos de pontos através da análise da representação espectral das imagens, utilizando diferentes métodos para se calcular as probabilidades de correspondência entre os pontos: *i*) através da matriz proximidade de peso Gaussiano, *ii*) da matriz proximidade Sigmoidal, *iii*) da função de peso crescente e, *iv*) da função de peso Euclidiano. Seguindo as idéias apresentadas em [9] e [34], neste trabalho explora-se como a estrutura modal dos padrões, tomados no espaço da frequência das texturas, pode ser utilizada para fins de reconhecimento. Para encontrar os melhores pontos padrões que representem todo o espectro, é utilizada uma variação da técnica de quantização de Lloyd [37].

Este capítulo apresenta a descrição dos modelos correlacionados com o modelo pro-

posto: o modelo de casamento entre pontos padrões realizado por Shapiro e Brady [34]; o modelo de recuperação de texturas através do espectro de potência das imagens, de Carcassoni et al. [9]; o modelo de matriz de co-ocorrência de níveis de cinza de Haralick [17]; e o modelo de quantização para a transmissão de sinais proposto por Lloyd [25].

4.2 Método de Correspondência de Shapiro e Brady

Shapiro e Brady [34] propuseram um novo método para realizar correspondência entre pontos padrões baseado na descrição modal de formas. Após escolhidos os pontos padrões de cada imagem, calcula-se a inter-relação entre esses pontos, sendo esta relação utilizada para avaliar a similaridade entre duas imagens. Shapiro e Brady foram inspirados pelo trabalho de Scott e Longuet-Higgins [33], que propuseram um algoritmo para realizar o casamento (*matching*) entre pontos padrões de duas imagens. Scott e Longuet-Higgins incorporaram uma medida de afinidade entre os padrões baseada na distância entre os elementos, formulada como o princípio de proximidade. O algoritmo recebe como entrada um conjunto de m padrões de uma imagem I_1 e um conjunto de n padrões de uma imagem I_2 . O primeiro passo consiste em enumerar todos os pares possíveis entre os dois conjuntos de padrões e armazená-los numa matriz denominada matriz de proximidade. Baseando-se nestas idéias, Shapiro e Brady propuseram uma abordagem baseada na análise modal da matriz proximidade entre os padrões. Neste modelo, a análise é realizada separadamente para cada uma das imagens, ou seja, numa abordagem intra-imagem, diferentemente do trabalho de Scott e Longuet-Higgins, que utilizaram uma abordagem inter-imagem. O algoritmo oferece robustez à translações e rotações no plano da imagem, bem como tolerância à mudanças de escala da imagem.

Sejam duas imagens I_1 e I_2 com m e n padrões respectivamente. Os m padrões da imagem I_1 definem um conjunto de m eixos atuando como um sistema de coordenadas em um espaço de dimensão m . Cada padrão da imagem é relacionado com um dos eixos, gerando desta forma uma descrição modal de uma imagem baseada na distribuição dos m padrões. O mesmo vale para a imagem I_2 .

Como o modelo utiliza uma abordagem intra-imagem, o cálculo das distâncias entre os pontos padrões é realizada de forma independente para cada imagem, sendo somente na etapa final, ou seja, no cálculo de similaridade, é que há o relacionamento entre as representações de duas imagens, gerando uma matriz de associação na qual será calculada a semelhança entre as imagens.

4.2.1 O Algoritmo de Shapiro e Brady

O algoritmo é dividido em quatro etapas distintas, que podem ser resumidas em:

i) Gerar a matriz proximidade através do cálculo das distâncias entre os m pontos padrões de cada imagem. Este cálculo pode ser realizado através da função de peso Gaussiano;

ii) Calcular a matriz modal através do cálculo dos autovalores e autovetores da matriz proximidade;

iii) Computar a matriz de associação entre as matrizes modais das imagens que estão sendo comparadas;

iv) Verificar a similaridade entre as imagens, utilizando as relações entre padrões presentes nas linhas e colunas da matriz de associação.

Dada uma imagem I_1 com m padrões selecionados com x_i , ($i = 1, \dots, m$) representando cada padrão. Um relacionamento entre estes pontos é construído por uma matriz quadrada H , denominada matriz proximidade, que armazena a distância entre os padrões x_i e x_j dentro da imagem, como se segue:

$$H_{ij} = e^{-r_{ij}^2/2\sigma_x^2} \quad (4.1)$$

onde $r_{ij}^2 = \|x_i - x_j\|^2$ é uma função Gaussiana que visa modelar a probabilidade de relações de adjacências entre os pontos. A interação entre os pontos na imagem é controlada pelo σ_x , onde x reforça a idéia de que a interação ocorre entre padrões da mesma imagem. Quanto maior for o valor de σ , maior será a vizinhança utilizada para o conhecimento de cada padrão. Valores pequenos para σ implicam em iterações mais locais, enquanto que valores maiores permitem iterações mais globais. Podemos ter valores diferentes de σ_x e σ_y para duas imagens, I_1 e I_2 , respectivamente.

A matriz H é simétrica e sua diagonal principal possui valor 1. Após a matriz proximidade ser calculada, a análise modal é então aplicada, criando uma matriz ortogonal V , denominada matriz modal, tal que $V = (E_1 | \dots | E_m)$, onde E_i são os autovetores da matriz proximidade H e cada linha i de V é chamada de vetor de padrões F_i .

$$V = \begin{bmatrix} F_1 \\ \cdot \\ \cdot \\ \cdot \\ F_m \end{bmatrix}$$

O cálculo dos procedimentos descritos anteriormente procede de forma independente e simultânea para as duas imagens, I_1 e I_2 . Desta forma teremos para os m padrões da imagem I_1 , as matrizes proximidade e modal, H_1 e V_1 , respectivamente, enquanto que para os n padrões da imagem I_2 , as matrizes H_2 e V_2 .

O passo final do algoritmo consiste em associar os dois conjuntos de vetores formados pelos pontos padrões de cada imagem, V_1 e V_2 , obtendo como resultado uma matriz de associação $Z_{1,2}$, que representa a semelhança entre os padrões das duas imagens. Podemos ter um número de padrões diferentes para as imagens I_1 e I_2 , resultando em um número de autovetores diferentes para cada matriz modal. Nestes casos, devemos realizar o truncamento dos autovetores da imagem com mais padrões selecionados.

Os valores da matriz de associação Z_{ij} são obtidos pela distância Euclidiana entre as linhas das matrizes modais. Representando a i -ésima linha de V_k por $F_{i,k}$, $k = 1, 2$ temos:

$$Z_{ij} = ||F_{i1} - F_{j2}||^2 \quad (4.2)$$

As linhas de Z representam as características da imagem I_1 , enquanto que as colunas representam as características da imagem I_2 . O casamento é dado por aqueles elementos na matriz Z que possuem os menores valores nas linhas e colunas. Um casamento perfeito é dado pelo valor 0 e um valor maior que 2 indica que não houve casamento.

4.2.2 Exemplo de Correspondência entre Pontos Padrões

Considere o exemplo a seguir onde as matrizes proximidades H_1 e H_2 tenham sido originadas de duas imagens I_1 e I_2 , respectivamente, onde cada qual possui $\sigma_x = \sigma_y = 4$ como o número de padrões selecionados. Desejamos saber se existe uma correspondência entre os padrões das duas imagens através da associação das matrizes modais V_1 e V_2 .

$$H_1 = \begin{bmatrix} 1.00 & 0.10 & 0.07 & 0.05 \\ 0.10 & 1.00 & 0.53 & 0.33 \\ 0.07 & 0.53 & 1.00 & 0.61 \\ 0.05 & 0.33 & 0.61 & 1.00 \end{bmatrix}, \quad V_1 = \begin{bmatrix} 0.01 & 0.12 & 0.98 & 0.12 \\ -0.35 & -0.77 & 0.04 & 0.53 \\ 0.77 & 0.07 & -0.10 & 0.63 \\ -0.53 & 0.62 & -0.14 & 0.56 \end{bmatrix}$$

$$H_2 = \begin{bmatrix} 1.00 & 0.13 & 0.05 & 0.09 \\ 0.13 & 1.00 & 0.45 & 0.29 \\ 0.05 & 0.45 & 1.00 & 0.62 \\ 0.09 & 0.29 & 0.62 & 1.00 \end{bmatrix}, \quad V_2 = \begin{bmatrix} 0.08 & 0.19 & 0.97 & 0.16 \\ -0.26 & -0.82 & 0.10 & 0.50 \\ 0.75 & 0.12 & -0.18 & 0.63 \\ -0.60 & 0.53 & -0.15 & 0.58 \end{bmatrix}$$

A matriz de associação $Z_{1,2}$ indica que o primeiro padrão da imagem I_1 casou com o primeiro padrão da imagem I_2 e que o segundo padrão da imagem I_1 casou com o segundo padrão da imagem I_2 e assim sucessivamente. Os valores em negrito indicam o casamento entre os padrões da imagem I_1 com a imagem I_2 . Como os menores valores das linhas e colunas coincidiram, houve um casamento perfeito entre os padrões das duas imagens, ou seja, para a linha 1, que indexa o primeiro padrão da imagem I_1 , o menor valor corresponde a coluna 1, que indexa o primeiro padrão da imagem I_2 e assim sucessivamente. Os valores próximos ao valor zero reforça ainda mais a idéia de casamento perfeito entre os padrões. Se o menor valor da primeira linha fosse o valor 1.88, indicaria que não houve um casamento perfeito entre os padrões das duas imagens, I_1 e I_2 , pois o primeiro padrão da imagem I_1 teria casado com o segundo padrão da imagem I_2 .

$$Z_{1,2} = \begin{bmatrix} \mathbf{0.01} & 1.88 & 2.15 & 2.03 \\ 2.11 & \mathbf{0.02} & 2.06 & 1.79 \\ 1.86 & 1.91 & \mathbf{0.01} & 2.09 \\ 1.95 & 2.21 & 1.89 & \mathbf{0.01} \end{bmatrix}$$

4.2.3 Resultados Experimentais

Realizamos alguns experimentos computacionais utilizando o algoritmo descrito anteriormente, objetivando verificar o casamento de padrões entre duas imagens. A Figura 4.1 mostra os padrões selecionados da Figura de uma mão (a) e sua forma rotacionada (b) em 45°. A descrição das formas é realizada tomando-se os pontos padrões selecionados ao longo das bordas de cada imagem. Neste exemplo, $m = 6$ pontos padrões foram utilizados para representar cada uma das imagens.

A Figura 4.2 mostra que houve um casamento entre os padrões das imagens da Figura 4.1. Como os autovetores são baseados somente na distância entre os pontos, a descrição da forma não é afetada por transformações que preservam estas distâncias como rotação, translações, escala e reflexões no plano da imagem.

O experimento da Figura 4.3 mostra o resultado obtido da comparação da imagem de uma casa e sua forma escalonada em 100%. A Figura 4.3 (a) mostram as imagens em

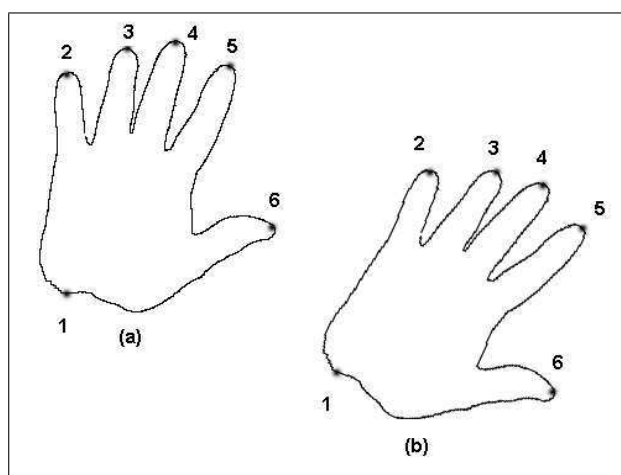


Figura 4.1: Pontos padrões da figura de uma mão e a sua forma rotacionada em 45° .

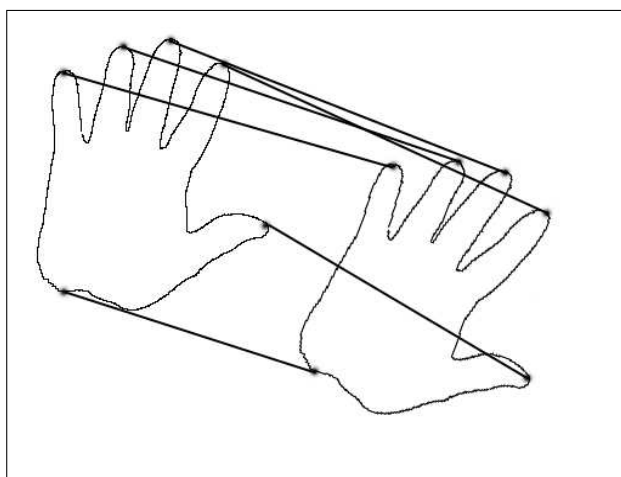


Figura 4.2: Casamento de todos os padrões selecionados ao longo da figura de uma mão e sua forma rotacionada.

tamanho normal e em escala. A Figura 4.3 (b) mostra os pontos padrões marcados para as imagens e a Figura 4.3 (c) mostra o casamento entre os pontos padrões selecionados.

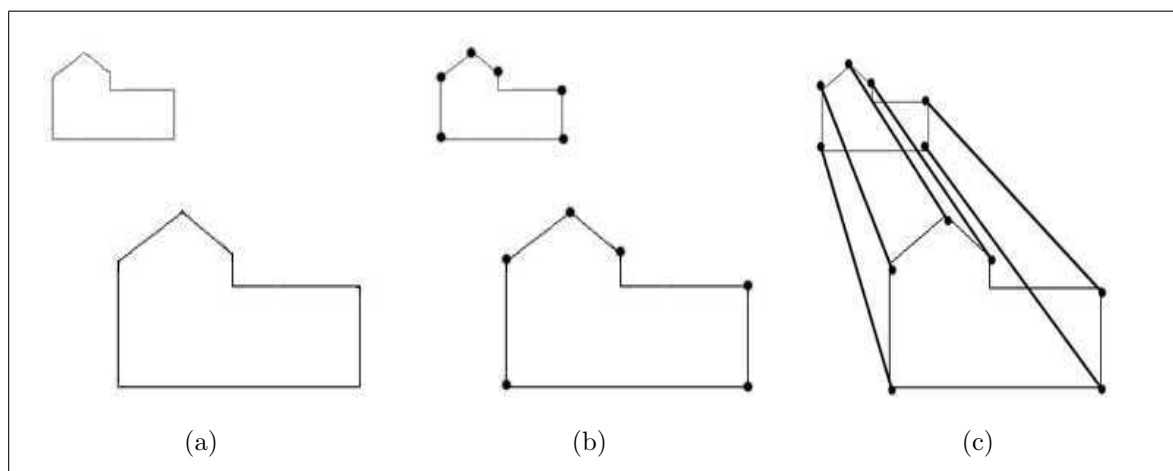


Figura 4.3: Exemplo de casamento entre a figura de uma casa e sua forma em escala.

4.3 Método de Correspondência Espectral de Carcassoni

Diferentemente de Shapiro e Brady [34], o modelo de Carcassoni, Ribeiro e Hancock [9] foi proposto com o intuito de realizar a recuperação de imagens similares em uma coleção de imagens através de uma imagem consulta. Desta forma, Carcassoni et al. [9] propõem um método de indexação de texturas onde a representação da imagem é feita através dos padrões dos picos da função de densidade espectral. Os autores utilizaram o método de correspondência modal entre pontos padrões, de Shapiro e Brady [34], tomando como pontos padrões os centros de clusters do espectro de potência das imagens e comparando a estrutura modal da matriz proximidade desses centros, ou seja, Shapiro e Brady utilizaram pontos padrões selecionados ao longo das bordas das formas como padrões, enquanto que Carcassoni et al. utilizaram os picos espectrais mais significativos como os padrões. Os picos espectrais são obtidos através da transformada de Fourier da autocorrelação da imagem.

4.3.1 O Algoritmo de Carcassoni

O algoritmo do método de Carcassoni et al. pode ser resumido em, primeiramente determinar os m picos mais significativos da imagem e calcular a matriz de proximidade entre esses picos. Em seguida é utilizado um método de clusterização com o intuito de encontrar os centros de atividade espectral. Dos centros dos clusters, para representar o espectro de potência, é extraída a representação modal. A recuperação das imagens do banco de dados que são mais similares à imagem consulta é feita através da comparação entre as matrizes modais das imagens armazenadas e a matriz modal da imagem consulta.

4.3.2 A Representação do Espectro de Potência

Carcassoni et al. utilizaram os picos do espectro de potência e posterior análise modal para fins de reconhecimento de texturas. A razão dada pelos autores para adotarem a abordagem modal é que os autovetores são efetivos na recuperação de imagens e textos em grandes bancos de dados. O passo inicial do método é calcular para as imagens do banco de dados, o espectro de potência das imagens através de uma função de densidade espectral. O espectro de potência de uma imagem bidimensional $f(x, y)$ é definido pela transformada de Fourier da sua autocorrelação. Os autores utilizaram o estimador de Blackman-Tukey, que é a resposta de frequência da função de autocorrelação janelada da imagem, dado por:

$$E_{BT}(u, v) = \sum_{m=-M+1}^{M-1} \sum_{n=-N+1}^{N-1} r_I(m, n) w(m, n) \exp \left[-j2\pi \left(\frac{mu}{M} + \frac{nv}{N} \right) \right], \quad (4.3)$$

onde u e v são as variáveis de frequência, m e n são coordenadas espaciais, $w(m, n)$ é uma função janela e $r(m, n)$ é a função de autocorrelação da imagem $I(x, y)$, que é dada por:

$$r_I(l, k) = \frac{1}{MN} \sum_{m=-M+1}^{M-1} \sum_{n=-N+1}^{N-1} I^*(m, n) I(m+l, n+k), \quad (4.4)$$

onde M e N são o número de pontos utilizados para discretizar a imagem I na direção x e y , respectivamente.

O espectro representa o conteúdo de frequência para diversas direções em um ponto na imagem. Os picos do espectro de potência correspondem as frequências espaciais dominantes na textura.

Os vetores de frequências do i^{th} pico de uma imagem I^α é dado por $U_i^\alpha = (u_i, v_i)$. Os vetores de frequência são concatenados em ordem decrescente, de acordo com a energia dos picos.

4.3.3 Cálculo da Matriz Proximidade dos Picos Dominantes

Determinados os m picos espectrais mais significativos de cada imagem é então realizada a construção da matriz proximidade entre estes picos, através de uma função de peso sigmoideal. Suponha uma imagem indexada α com m picos, os quais vamos representar pelo conjunto: $P = \{U_i^\alpha, i = 1, 2, \dots, m\}$. A matriz proximidade dos picos espectrais é dada pela equação:

$$H_\alpha(i, j) = \frac{2}{\pi \|U_i^\alpha - U_j^\alpha\|} \log \cosh \left(\frac{\pi}{s} \|U_i^\alpha - U_j^\alpha\| \right) \quad (4.5)$$

A matriz proximidade de picos espectrais possui o papel de estabelecer o local dos grupos de pontos mais significantes da imagem. A função $\log(\cosh Z)$ foi escolhida devido ao fato da mesma ser flexível, modelando a probabilidade de adjacência entre pontos, restringindo assim as correspondências individuais dos pontos.

4.3.4 Representação Modal

Para cada uma das matrizes proximidades H_α calcula-se seus autovalores e seus autovetores. Supondo λ_i como o i -ésimo autovalor dessa matriz e ϕ_i sendo o correspondente autovetor, temos:

$$|H_\alpha - \lambda I| = 0$$

e

$$H_\alpha \phi_i^\alpha = \lambda_i^\alpha \phi_i^\alpha$$

Na construção da matriz modal $m \times m$; $\phi^\alpha = (\phi_1^\alpha | \phi_2^\alpha | \dots | \phi_m^\alpha)$, os autovetores são armazenados em ordem decrescente de acordo com a magnitude dos autovalores, isto é, $|\lambda_1^\alpha| > |\lambda_2^\alpha| > \dots > |\lambda_m^\alpha|$. Esses autovetores ϕ_i são ortonormais, ou seja

$$\langle \phi_i, \phi_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

4.3.5 Cálculo da Matriz Centróide

Os m maiores autovetores ϕ^α associados aos m maiores autovalores são utilizados para representar os clusters dos picos do espectro de potência, que é dado pelo centróide dos autovalores através da equação:

$$C_n^\alpha = \frac{\sum_{i=1}^M |\phi_\alpha(i, n)| U_i^\alpha}{\sum_{i=1}^M |\phi_\alpha(i, n)|} \quad (4.6)$$

Esses centróides são usados para calcular a matriz proximidade dos centróides utilizando-se a seguinte equação:

$$H_\alpha^C(i, j) = \frac{2}{\pi \|C_i^\alpha - C_j^\alpha\|} \log \cosh\left(\frac{\pi}{s} \|C_i^\alpha - C_j^\alpha\|\right) \quad (4.7)$$

Após o cálculo da matriz proximidade do centróide H_α^C , os procedimentos 4.3.4 e 4.3.5 são repetidos, dando origem a matriz modal do cluster centro, ψ_α .

4.3.6 Cálculo da Similaridade

A similaridade entre a imagem consulta e as imagens armazenadas no banco de dados é feita através da comparação das matrizes modais $\psi_q = (\psi_1^q \dots \psi_m^q)$ e $\psi_\alpha = (\psi_1^\alpha \dots \psi_m^\alpha)$.

O objetivo do método é encontrar uma imagem na base de dados cujo espectro de potência seja o mais próximo possível do espectro de potência da imagem amostra. O grau de similaridade é medido através da seguinte equação:

$$w_q = \max_{\alpha} \sum_{l=1}^N \sum_{m=1}^N \exp[-k \sum_{n=1}^N (\psi_q(l, n) - \psi_\alpha(m, n))^2] \quad (4.8)$$

onde k é uma constante que tem a função de tornar os termos da soma mais significativos.

Observamos que os elementos das linhas das matrizes modais não contribuem neste somatório caso eles difram muito, e quanto mais próximas estiverem as linhas de duas matrizes comparadas maior será o valor da soma acima.

4.3.7 Resultados Experimentais

Realizamos alguns experimentos computacionais utilizando a técnica de Carcassoni et al. [9], com o objetivo de verificar a eficiência deste método na recuperação de imagens. Os experimentos utilizaram um banco de dados homogêneo com texturas que possuíam padrões regulares, como texturas de tecidos e papéis de presente e parede. Esta categoria de imagens foi a utilizada por Carcassoni et al. em seus experimentos em [9]. A Figura 4.4 exibe uma amostra do banco de dados utilizado nos experimentos.

As Figuras mostradas nos experimentos à seguir estão rotuladas como: (a) para a imagem consulta e (b-e) para os 4 melhores resultados dentre as 10 primeiras imagens retornadas do *ranking*.

O primeiro experimento, mostrado na Figura 4.5, exibe os resultados obtidos pelo método de Carcassoni utilizando a imagem de um tecido como consulta, enquanto que no segundo experimento temos os resultados obtidos utilizando uma imagem de papel de parede como consulta, ambas em um banco de dados de imagens contendo 350 imagens regulares e semi-regulares.

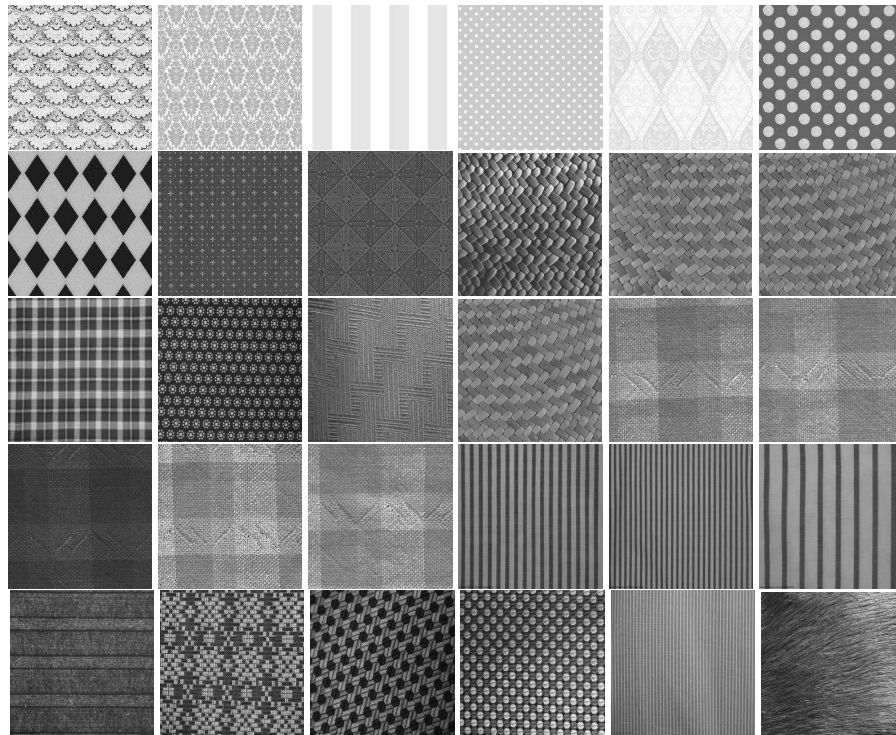


Figura 4.4: Exemplo de texturas regulares utilizadas por Carcassoni.

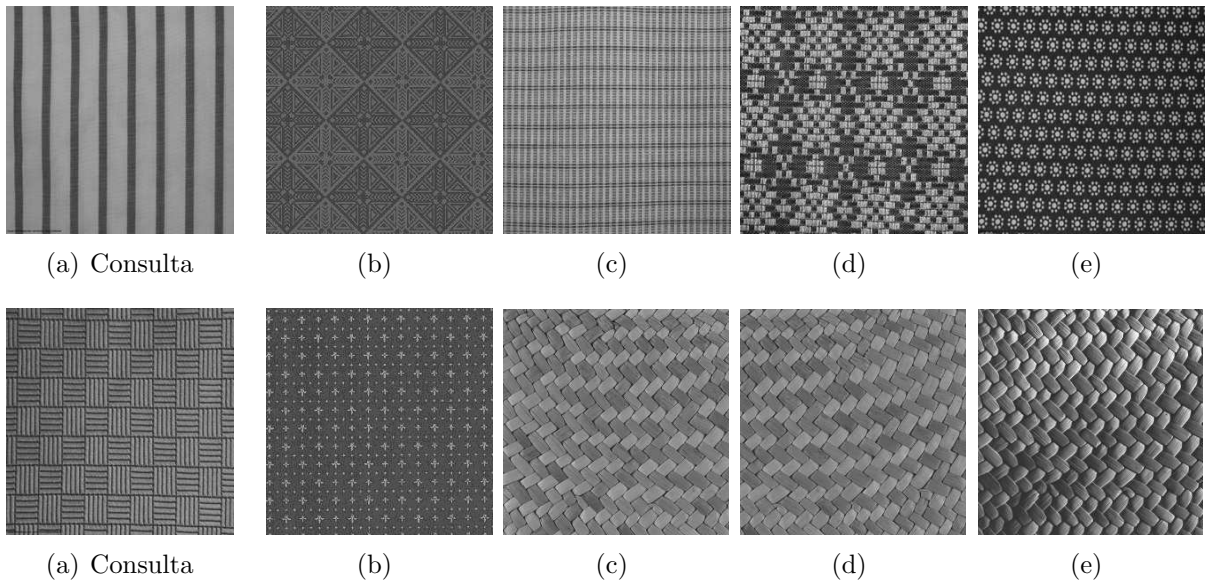


Figura 4.5: Resultados obtidos pelo método de Carcassoni em um banco de dados homogêneo.

4.4 Matrizes de Co-ocorrência de Níveis de Cinza

A matriz de co-ocorrência de níveis de cinza é uma abordagem estatística amplamente conhecida e utilizada como extrator de padrões de textura. Haralick [17] definiu o método GLCM (*Gray Level Co-occurrence Matrices*) como uma matriz de frequências no qual dois

pixels, separados por um certo vetor, ocorrem na imagem. A distribuição da matriz depende do relacionamento angular e de distância entre *pixels*. Variando o vetor usado para se obter a matriz de co-ocorrência permite-se capturar diferentes características da textura. Uma vez criada a matriz GLCM, vários padrões podem ser calculados à partir dela, como a energia, entropia, correlação, contraste, etc.

Uma matriz de co-ocorrência $P_d(i, j)$ contabiliza a co-ocorrência de *pixels* com valores de cinza i e j em uma dada distância d . A distância d é definida em coordenadas polares (d, α) , onde $\alpha = \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ \text{ e } 315^\circ\}$. A matriz de co-ocorrência $P_d(i, j)$ pode ser definida como:

$$P_d(i, j) = Pr(I(p_1) = i \quad \wedge \quad I(p_2) = j \quad \wedge \quad ||p_1 - p_2|| = d)$$

onde Pr é a probabilidade e p_1 e p_2 são posições na escala de cinza da imagem I .

As direções usualmente tratadas são as de $0^\circ, 45^\circ, 90^\circ$ e 135° e as distâncias d são escolhidas de acordo com a granularidade das imagens manipuladas ($d = 1$ para *pixels* vizinhos). Como exemplo, suponhamos a imagem I sendo representada pela matriz de pontos (5×5) abaixo.

$$I = \begin{bmatrix} 0 & 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 & 2 \\ 3 & 2 & 2 & 0 & 1 \\ 0 & 3 & 2 & 2 & 2 \\ 3 & 3 & 3 & 0 & 0 \end{bmatrix}$$

Assuma-se $N_g = 4$ ($I(i, j) \in \{0, 1, 2, 3\}$) como o número de níveis de cinza da imagem I . A matriz de co-ocorrência P para um par (d, θ) é definida pela matriz $N_g \times N_g$

$$P = \frac{1}{R} \begin{bmatrix} \eta(0,0) & \eta(0,1) & \eta(0,2) & \eta(0,3) \\ \eta(1,0) & \eta(1,1) & \eta(1,2) & \eta(1,3) \\ \eta(2,0) & \eta(2,1) & \eta(2,2) & \eta(2,3) \\ \eta(3,0) & \eta(3,1) & \eta(3,2) & \eta(3,3) \end{bmatrix}$$

onde R é o número total de *pixels* pares e $\eta(I_i, I_j)$ é o número de *pixels* pares que tem valores de nível de cinza I_i, I_j na posição (d, θ) .

Para a representação da imagem I dada anteriormente, temos a sua matriz de co-ocorrência para a distância $d = 1$ e ângulo $\theta = 0$ a seguinte matriz:

$$P(d = 1, \theta = 0) = \frac{1}{24} \begin{bmatrix} 4 & 3 & 1 & 2 \\ 3 & 4 & 2 & 0 \\ 1 & 2 & 8 & 2 \\ 2 & 0 & 2 & 4 \end{bmatrix}$$

Uma vez que as informações sobre a textura estão contidas na matriz de co-ocorrência, os padrões da textura podem ser calculados à partir desta matriz. Haralick extraiu 14 padrões da matriz de co-ocorrência, sendo que destes, os mais utilizados na tarefa de classificação e recuperação estão listados na tabela 4.1, onde μ_i e μ_j são a média e σ_i e σ_j são o desvio padrão de P .

Padrão de Textura	Equação
Contraste	$f_1 = \sum_i \sum_j (i - j)^2 P(i, j)$
Energia	$f_2 = \sum_i \sum_j P^2(i, j)$
Entropia	$f_3 = \sum_i \sum_j P(i, j) - \log P(i, j)$
Homogeneidade	$f_4 = \sum_i \sum_j \frac{P(i, j)}{(1 + i - j)}$
Correlação	$f_5 = \frac{1}{\sigma_i \sigma_j} \sum_i \sum_j (i - \mu_i)(j - \mu_j) P(i, j)$

Tabela 4.1: Algumas características que podem ser calculadas da matriz de co-ocorrência.

4.4.1 Algoritmo do Método GLCM

Com o intuito de se calcular a matriz de co-ocorrência para cada imagem armazenada no banco de dados, bem como para as imagens consulta, para posterior análise de similaridade, foram adotados os seguintes procedimentos:

i) Afim de evitar matrizes grandes, normalmente a representação de níveis de cinza das imagens é reduzida para valores entre 4 e 32. Neste trabalho, reduzimos esta representação de 256 para 16 níveis de cinza, onde foi adotado o seguinte procedimento: valores de níveis de cinza entre 1 e 16 na representação de 256 níveis passaram a ser considerados como valor de nível de cinza 1 na representação reduzida e, valores entre 17 e 32 passaram a ser considerados como valor de nível de cinza 2 e assim sucessivamente.

ii) O próximo passo foi gerar as matrizes de co-ocorrência de cada imagem, para as direções $0^\circ, 45^\circ, 90^\circ$ e 135° e distâncias $d = \{1, 2, 3, 4\}$, totalizando um conjunto de 16 matrizes de co-ocorrência para cada imagem.

iii) Em seguida, uma matriz média de co-ocorrência foi obtida através da média da soma das 16 matrizes.

iv) No último passo, foram obtidos valores de cinco descritores: Energia, Entropia, Contraste, Correlação e Homogeneidade. Estes valores foram armazenados em um vetor, um para cada imagem do banco de dados bem como para a imagem consulta.

A similaridade entre duas imagens foi medida através da distância Euclidiana entre os vetores. Quanto menor o valor obtido do cálculo de similaridade, maior a semelhança entre as imagens.

4.4.2 Resultados Experimentais

Realizamos alguns experimentos computacionais utilizando a técnica de matriz de co-ocorrência de níveis de cinza de Haralick [17], com o objetivo de verificar a eficiência deste método na recuperação de imagens. As Figuras 4.6 mostram dois dos experimentos computacionais realizados, onde a imagem de um tecido e a imagem de uma parede de pedras são utilizadas como consulta, em um banco de dados de imagens contendo 800 imagens regulares e não regulares.

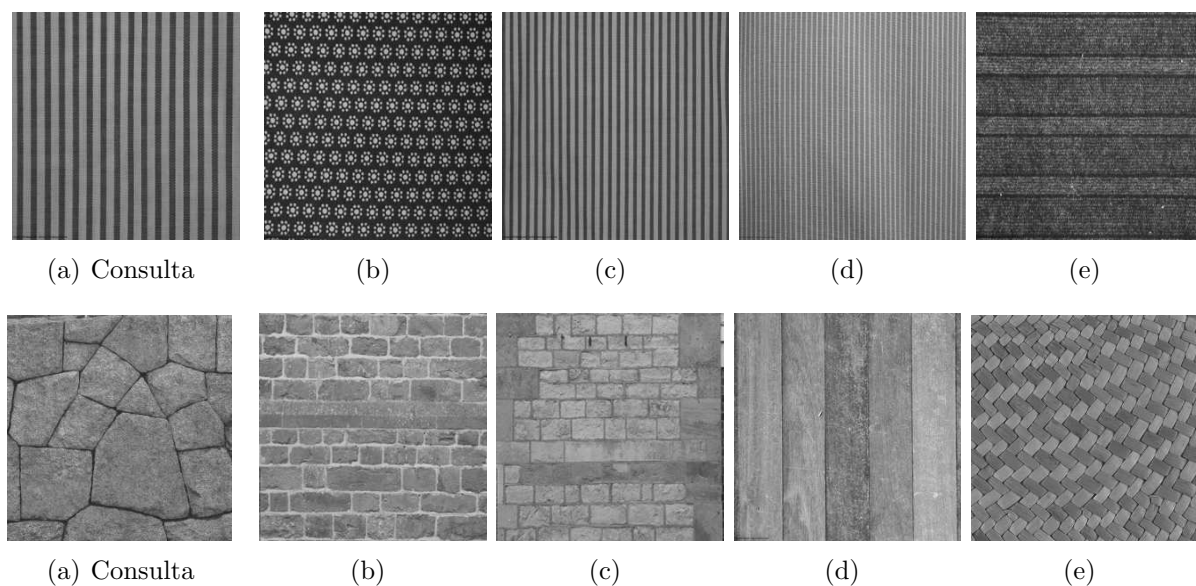


Figura 4.6: Resultados obtidos pelo método da matriz de co-ocorrência de Haralick.

4.5 Técnica de Lloyd

O algoritmo de quantização de Lloyd foi baseado no teorema de amostragem de Shannon-Nyquist. Este teorema afirma que um sinal de voltagem $s(t)$, $-\infty < t < \infty$, contendo somente frequências menores que W ciclos/s podem ser recuperadas de uma sequência de valores de amostragem de acordo com:

$$s(t) = \sum_{j=-\infty}^{\infty} s(t_j)K(t - t_j), \quad -\infty < t < \infty \quad (4.9)$$

onde $s(t_j)$ é o valor de s no j -ésimo instante de amostragem

$$t_j = \frac{j}{2W}, \quad -\infty < j < \infty$$

e

$$K(t) = \frac{\sin 2\pi Wt}{2\pi Wt},$$

onde $\sin t/t$ é um pulso de comprimento apropriado.

O sistema PAM (*Pulse-amplitude modulation*) é um sistema baseado no teorema de amostragem, onde é transmitido sobre o canal apenas a sequência de amostras do sinal $\dots, s(t_{-1}), s(t_0), s(t_1), \dots$ ao invés de todos os valores $s(t)$ do mesmo. O receptor constrói os pulsos $K(t - t_j)$ e os adiciona com as amplitudes recebidas $s(t)$, como em 4.9, produzindo uma reprodução exata do sinal original s .

O modelo proposto por Lloyd é uma variação ao sistema PAM, onde à partir de um conjunto finito de sinais é realizado um processo de amostragem onde os sinais são divididos em subconjuntos finitos, sendo transmitido ao receptor somente a informação do subconjunto onde se encontra cada sinal. Dentro do receptor existe um conjunto de valores fixos de voltagens, denominados **quanta**, um para cada subconjunto. Quando o receptor é informado que uma certa amostra está em determinado subconjunto, ele utiliza o valor do quantum daquele subconjunto como uma aproximação ao valor real do sinal, construindo assim um sinal de banda limitada baseado nos valores de amostragem. Desta forma Lloyd [25] considera o problema de determinar os subconjuntos e quantas que melhor minimizam o ruído do sinal, que é a diferença entre o sinal enviado e o recebido. O algoritmo de Lloyd utiliza uma abordagem probabilística para encontrar os melhores quanta que representarão o sinal.

4.6 Considerações Finais do Capítulo

Neste capítulo foram apresentados os trabalhos correlatos que serviram de base para o modelo proposto. Shapiro e Brady [34] introduziram um modelo intra-imagem baseado em formas, no qual alguns pontos padrões são definidos ao longo da borda do objeto da imagem. Posteriormente, dois objetos podem ser comparados através da associação entre

os seus pontos padrões. O modelo de recuperação de imagens proposto por Carcassoni utiliza a análise modal à partir de clusters dos picos espectrais da imagem, enquanto que o método da matriz de co-ocorrência de níveis de cinza utiliza o inter-relacionamento entre *pixels* vizinhos para extrair as características da imagem, de onde podem ser calculados diversos descritores, como energia, correlação, entropia, etc. Lloyd, em seu método de quantização, propôs que um conjunto finito de sinais, divididos em subconjuntos, podem ser representados através de um número reduzido de valores, ao invés de todo o sinal.

O próximo capítulo apresenta o método proposto.

5 Método Proposto

5.1 Introdução

Seguindo as idéias introduzidas por Carcassoni et al. em [9] e, considerando que duas texturas diferentes podem possuir os N maiores picos do espectro de potência muito próximos, levando a uma similaridade errônea entre as duas imagens, é proposto neste trabalho representar uma imagem através dos valores mais representativos dos picos do espectro de potência da imagem, ao invés dos picos de maior magnitude. Para encontrar os melhores representantes no espaço de frequência da imagem é proposto um modelo de quantização baseado na técnica de quantização de Lloyd [25].

5.2 Fluxograma do Método Proposto

O diagrama da Figura 5.1 mostra as etapas de indexação (processo de representação) e consulta (processo de recuperação) do método proposto. Na etapa de indexação cada uma das imagens armazenadas no banco de dados passa pelo processo de representação (extração de características). Após obtida a representação matricial da imagem, o seu espectro de potência é calculado através da transformada de Fourier da sua autocorrelação. Uma vez obtido o espectro de potência se inicia o processo de quantização da imagem, que consiste em escolher um conjunto de representantes para a mesma. Em seguida é calculada a matriz proximidade entre os representantes escolhidos. A matriz proximidade fornece uma relação de proximidade entre estes representantes. O passo seguinte consiste na análise modal da matriz proximidade, feita através do cálculo dos autovalores e autovetores desta matriz, dando origem a representação final da imagem, denominada de vetor de características, que será utilizado para medir a similaridade entre as imagens.

O processo de recuperação se inicia com a escolha feita pelo usuário de uma imagem consulta. A imagem consulta pode ou não estar armazenada no banco de dados de imagens. Uma vez selecionada a imagem consulta, é obtido o seu vetor de características

através do processo de representação da imagem já descrito anteriormente. Uma vez obtido o vetor de características da imagem consulta, é realizado o cálculo de similaridade entre a imagem consulta e as imagens armazenadas no banco de dados, utilizando os seus vetores características. Calculada a similaridade, um *ranking* é montado em ordem decrescente de acordo com o valor de similaridade da imagem do bando de dados I^a com a imagem consulta I^q , de forma que o topo do *ranking* contenha as imagens mais similares à imagem consulta.

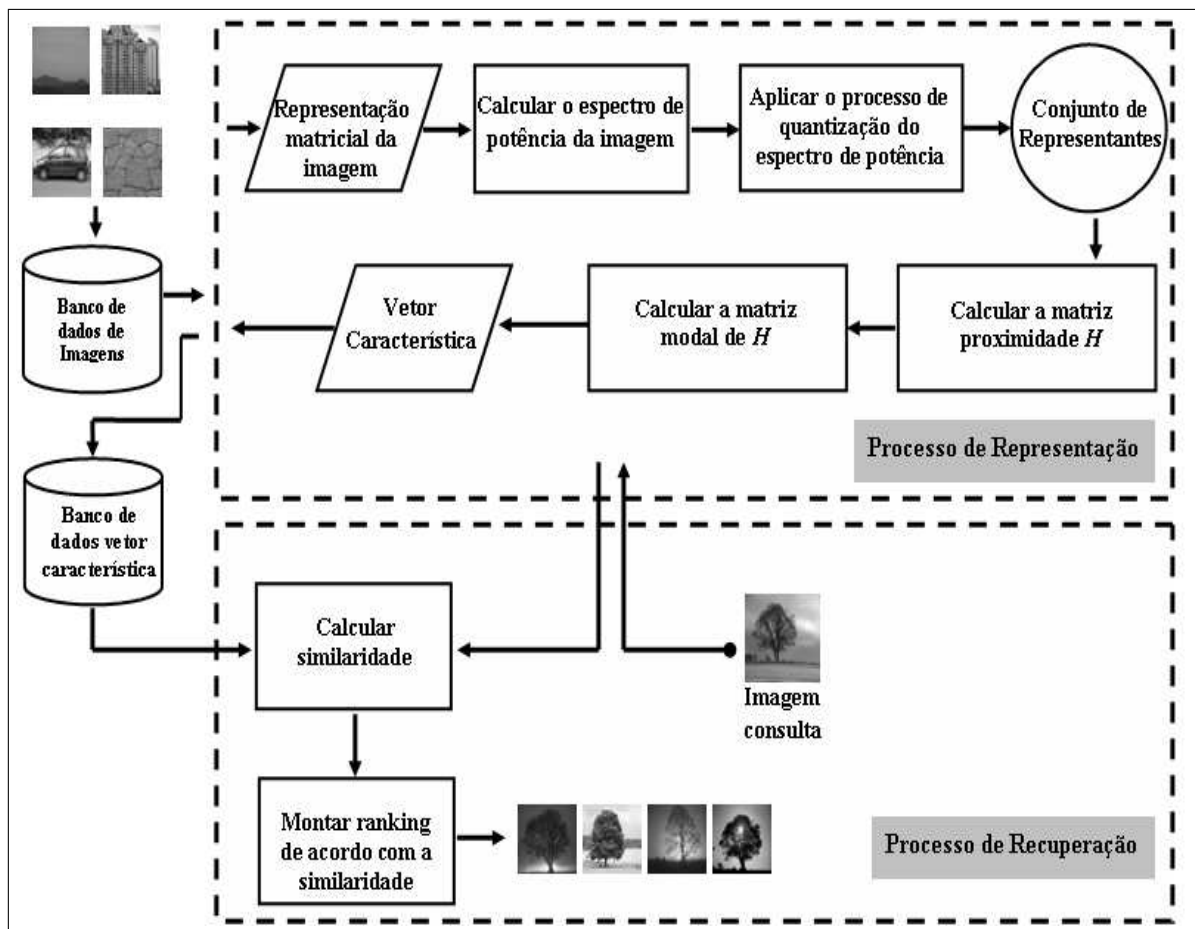


Figura 5.1: Diagrama do processo de representação e consulta de imagens no modelo proposto.

As próximas seções descrevem em detalhes o algoritmo do método proposto.

5.3 A Representação do Espectro

O espectro de potência de uma imagem $I(x, y)$ é definido como sendo a transformada de Fourier da função de autocorrelação dessa imagem. Os pontos $U_i = (u_i, v_i)$ no espectro correspondem às frequências espaciais dominantes de uma imagem. Para se obter o espec-

tro de potência foi utilizado o estimador espectral de Blackman-Tukey, calculado através equação 5.1. Mais detalhes sobre análise espectral bidimensional podem ser encontrados em [21].

$$E_{BT}(u, v) = \sum_{m=-M+1}^{M-1} \sum_{n=-N+1}^{N-1} r_I(m, n) w(m, n) \exp \left[-j2\pi \left(\frac{mu}{M} + \frac{nv}{N} \right) \right], \quad (5.1)$$

onde u e v são as variáveis de frequência, m e n são coordenadas espaciais, $w(m, n)$ é uma função janela e $r(m, n)$ é a função de autocorrelação da imagem $I(x, y)$, que é dada por:

$$r_I(l, k) = \frac{1}{MN} \sum_{m=-M+1}^{M-1} \sum_{n=-N+1}^{N-1} I^*(m, n) I(m+l, n+k), \quad (5.2)$$

onde M e N são o número de pontos utilizados para discretizar a imagem I na direção x e y , respectivamente.

Para fins de reconhecimento de imagens, neste trabalho será utilizada uma quantidade S de valores de cada imagem para representar seu espectro. Os picos representantes q_i serão escolhidos utilizando uma modificação da técnica de Lloyd [4, 13, 37]. O processo de quantização será descrito em detalhes na próxima seção.

5.4 Modelo de Quantização

Um esquema de quantização consiste em se ter um conjunto $Q = \{Q_1, Q_2, \dots, Q_S\}$ e um conjunto de níveis denominado quanta $q = \{q_1, q_2, \dots, q_S\}$. Q é definido de tal forma que dois conjuntos $\{Q_i\}$ e $\{Q_j\}$ possuam intersecção vazia e a união de todos os Q cubra todo o espectro da imagem I .

Além disso, deve ser determinado um conjunto t_k , $k = 1, \dots, S+1$ denominado conjunto de separadores e escolhidos de tal forma que satisfaçam a seguinte condição:

$$t_1 < q_1 < t_2 < q_2 < \dots < q_S < t_{S+1}$$

onde S é o número de níveis (**quanta**) que serão utilizados para representar uma imagem. A Figura 5.2 mostra os elementos do processo de quantização.

Cada quantizador Q produz uma imagem quantizada I , modificando os picos com

valores absolutos pertencentes ao intervalo: $[t_k, t_{k+1})$ por q_k para todo $k = 1, 2, \dots, S$.

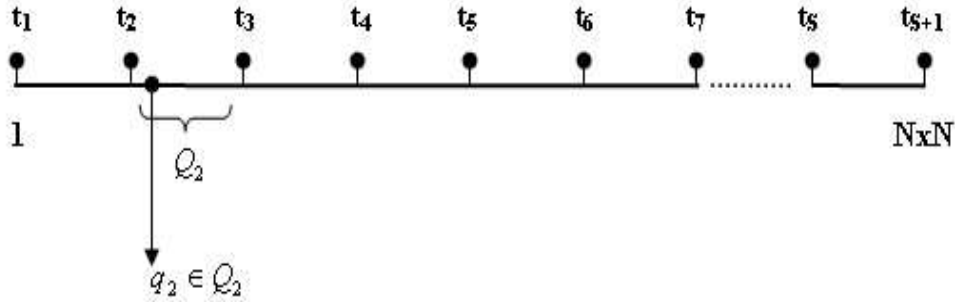


Figura 5.2: Elementos do processo de quantização.

Uma das tarefas mais importantes do procedimento de quantização é encontrar o melhor quanta. Lloyd [25] utiliza uma abordagem probabilística para escolher os quanta q_k , e reduz a abordagem probabilística à um problema variacional de se encontrar a minimização de um Funcional de Energia. O autor minimiza o seguinte funcional:

$$E(Q_s) = \sum_{k=1}^S \int_{Q_k} (q_k - s)^2 dF(s) \quad (5.3)$$

onde $F(s)$ é uma distribuição probabilística associada com a imagem I e dada por:

$$F(s) = P\{|I(u, v)| < s\}$$

Com o intuito de construir os conjuntos de quanta e de separadores, o passo inicial do algoritmo de quantização consiste em ordenar os picos do espectro de potência, em ordem decrescente, de acordo com a magnitude dos picos, i.e. $U_i \leq U_j$, se e somente se, $|U_i| \leq |U_j|$. O primeiro passo do algoritmo criará um conjunto de quanta q_k e um conjunto de separadores t_k , seguindo a técnica de Lloyd.

Como os separadores separam massa, pode ocorrer de dois quanta q_i e q_{i+1} possuírem o mesmo valor, e ainda, dois separadores t_i e t_{i+1} serem muito distantes um do outro. Para evitar estas situações, foi introduzido no algoritmo de Lloyd [25] um quarto passo.

Primeiro passo: O objetivo do primeiro passo é particionar os elementos do espectro de potência de seu conjunto original, dentro de um conjunto finito Q , tal que cada Q_i tenha a mesma massa. Este passo começa com a definição, pelo usuário, de um número de quanta que será usado para agrupar valores dos picos de espectro de frequência das imagens. Considere S_1 este número. Com a finalidade de definir cada $Q_i = [t_{i-1}, t_i)$, $i = 2, \dots, S_1 + 1$,

os separadores t_i são escolhidos de tal forma que:

$$\int_{Q_j} s dF(s) = \int_{Q_r} s dF(s) \quad \forall j, r$$

onde $F(s)$ é a distribuição de probabilidade associada à imagem I e dada por 5.3.

Após definir o número S_1 , os separadores t_i são definidos como: $t_j = t_{j-1} + A, j = 2, 3, \dots, S_1$, onde a quantidade A é dada por:

$$A = \frac{\int_{\Omega} dF(s)}{S_1}$$

onde Ω representa o domínio da imagem.

Os separadores t_1 e t_{S+1} são definidos a priori como $m = \min\{|U_i(u_i, v_i)|, i = 1, \dots, N \times N\}$ and $M = \max\{|U_i(u_i, v_i)|, i = 1, \dots, N \times N\}$, respectivamente. Os quanta q são considerados como números reais compreendidos entre m e M .

Em seguida é realizado o cálculo de cada quantum q_j como o centro de massa de cada correspondente Q_i :

$$q_j = \frac{\int_{Q_j} s dF(s)}{\int_{Q_j} dF(s)}$$

Para evitar que um quantizador possua o mesmo valor de um separador, o valor do quantizador q_j é escolhido como um número inteiro (mais próximo do valor real), pertencente ao intervalo $[m, M]$ e o separador como o número mais próximo de um elemento do conjunto $\{m - 0.5, m + 0.5, m + 1.5, \dots, M + 0.5\}$.

Objetivo do 2º passo: O objetivo do segundo passo é criar um novo conjunto de separadores $\{\bar{t}_i\}$ de tal forma que:

$$\bar{t}_i = t_1 + \frac{i(T - t_1)}{S_2}$$

onde $i = 2, \dots, S_2 - 1$, $T = t_{S+1}$ e S_2 é uma constante inteira definida pelo usuário.

Neste passo temos a mesma quantidade de elementos dentro de cada Q_r ou seja, os separadores são distribuídos em intervalos iguais no vetor de frequências U .

Objetivo do 3º passo: O terceiro passo se inicia com a união dos separadores $\{t_i\}$,

$i = 1, \dots, S_1$, obtidos no primeiro passo com os separadores $\{\overline{t_j}\}$, $j = 1, \dots, S_2$ obtidos no segundo passo. Um novo conjunto ordenado e renomeado de separadores $\{t_k\}$, $k=1, \dots, S$, onde $S \leq S_1 + S_2 + 1$, é criado, obedecendo as regras abaixo:

- Se existirem dois separadores t_α e t_β com o mesmo valor, um dos separadores é descartado.
- Se o número de elementos do espectro de potência, pertencente a cada $Q_r = [t_r, t_{r+1})$ for insignificante, o separador t_r deverá ser descartado do conjunto de separadores.

Satisfeitas essas duas condições, o conjunto de separadores é novamente renomeado como t_k , $k = 1, \dots, S$, $S \leq S_1 + S_2 + 1$.

Após estes passos, os valores de cada quantum são calculados como se segue:

$$q_k = \frac{\int_{Q_k} s dF(s)}{\int_{Q_k} dF(s)}$$

Quarto passo: O quarto passo do algoritmo consiste em um processo de refinamento, onde:

- Se existirem dois quanta q_i e q_{i-1} com praticamente a mesma massa, ou seja, $q_i - q_{i-1} < V$, para alguma constante positiva V , o separador t_i é removido do conjunto de separadores, q_i é removido do conjunto de quanta e o quantum q_{i-1} é recalculado.

Terminado este procedimento, são obtidos os S valores q_j para representar o espectro de potência da imagem.

O próximo passo é estabelecer uma relação de proximidade entre os S representantes escolhidos e analisar a semelhança entre as imagens através da similaridade entre a relação de proximidade destes representantes.

5.5 Construção da Relação entre os Pontos Representantes da Imagem

Neste trabalho, o reconhecimento de texturas é realizado através da análise modal dos vetores de frequência que foram escolhidos à partir do processo de quantização do espectro

de potência. A similaridade é medida pela comparação da matriz modal da imagem consulta com as matrizes modais das imagens armazenadas no banco de dados.

Para medir a proximidade entre o conjunto de pontos que representam a imagem, é calculada como em [9], a matriz peso sigmoidal:

$$H_{\alpha}(i, j) = \frac{2}{\pi ||q_i^{\alpha} - q_j^{\alpha}||} \log \cosh\left(\frac{\pi}{s} ||q_i^{\alpha} - q_j^{\alpha}||\right),$$

onde q_i^{α} é um dos representantes da imagem, calculados como mostrado no processo de quantização. A matriz proximidade estabelece o relacionamento entre os pontos representantes de uma imagem I^{α} .

Neste trabalho, a estrutura modal dos valores representantes da imagem é encontrada através do cálculo dos autovalores e autovetores da matriz proximidade H_{α} . Seja λ_i o i -ésimo autovalor da matriz H_{α} e ϕ_i o seu autovetor correspondente. Neste caso, tem-se:

$$|H_{\alpha} - \lambda I| = 0$$

e

$$H_{\alpha}\phi_i^{\alpha} = \lambda_i^{\alpha}\phi_i^{\alpha}.$$

A matriz modal ϕ^{α} ($N \times N$) é construída tendo os autvetores como seus vetores coluna, ou seja, $\phi^{\alpha} = (\phi_1^{\alpha} | \phi_2^{\alpha} | \dots | \phi_m^{\alpha} |)$, onde os autvetores são ordenados de forma decrescente de acordo com a magnitude de seus autovalores $|\lambda_1^{\alpha}| > |\lambda_2^{\alpha}| > \dots > |\lambda_m^{\alpha}|$. Estes autvetores são mutualmente ortogonais e unitários, formando uma base ortonormal, ou seja:

$$\langle \phi_i, \phi_j \rangle = \begin{cases} 1 & \text{se } i = j, \\ 0 & \text{se } i \neq j \end{cases}$$

5.6 Cálculo da Similaridade

Com o propósito de encontrar a textura na base de dados que possua o espectro de potência mais semelhante ao espectro de potência da textura consulta, a similaridade é medida através da comparação entre a matriz consulta ϕ^q com todas as outras matrizes ϕ^{α} da base de dados, usando uma função de erro Gaussiano definida por:

$$w_q = \max_{\alpha} \sum_{l=1}^M \sum_{m=1}^M \exp[-k \sum_{n=1}^M (\phi^q(l, n) - \phi^{\alpha}(m, n))^2] \quad (5.4)$$

onde k é uma constante e $M \leq S$.

O conjunto obtido do resultado de w_q é ordenado de forma não decrescente, pelo valor w_{α} dado pela equação, formando um *ranking* de imagens similares onde no topo do *ranking* encontram-se as imagens mais similares em relação à imagem consulta.

Cabe salientar que quando os elementos das linhas de ϕ^q e ϕ^{α} são muito diferentes, não há contribuições significantes desses elementos no resultado final e quando as linhas das matrizes são muito semelhantes, o valor da soma na equação acima será maior.

O maior valor de w_{α} representa a imagem I^{α} com maior semelhança com a imagem consulta I^q . A constante k tem a função de tornar os termos da soma mais significativos.

5.7 Resultados Experimentais

A Figura 5.3 mostra três experimentos utilizando o método proposto onde imagens regulares e não regulares foram utilizadas como consulta, em um banco de dados de imagens contendo 800 imagens regulares e não regulares.

5.8 Considerações Finais do Capítulo

Através do processo de quantização é possível representar uma imagem considerando não somente os picos de maior magnitude, mas todos os picos do espectro de potência. Desta forma, acredita-se obter uma representação mais uniforme, considerando as características estruturais contidas nas altas e baixas frequências do espectro de potência da imagem, evitando erros de similaridade causados por imagens diferentes contendo valores similares dos picos do espectro.

O próximo capítulo apresenta os resultados experimentais realizados comparando o método proposto com o método de Carcassoni [9] e com o método de matriz de co-ocorrência de Haralick [17].

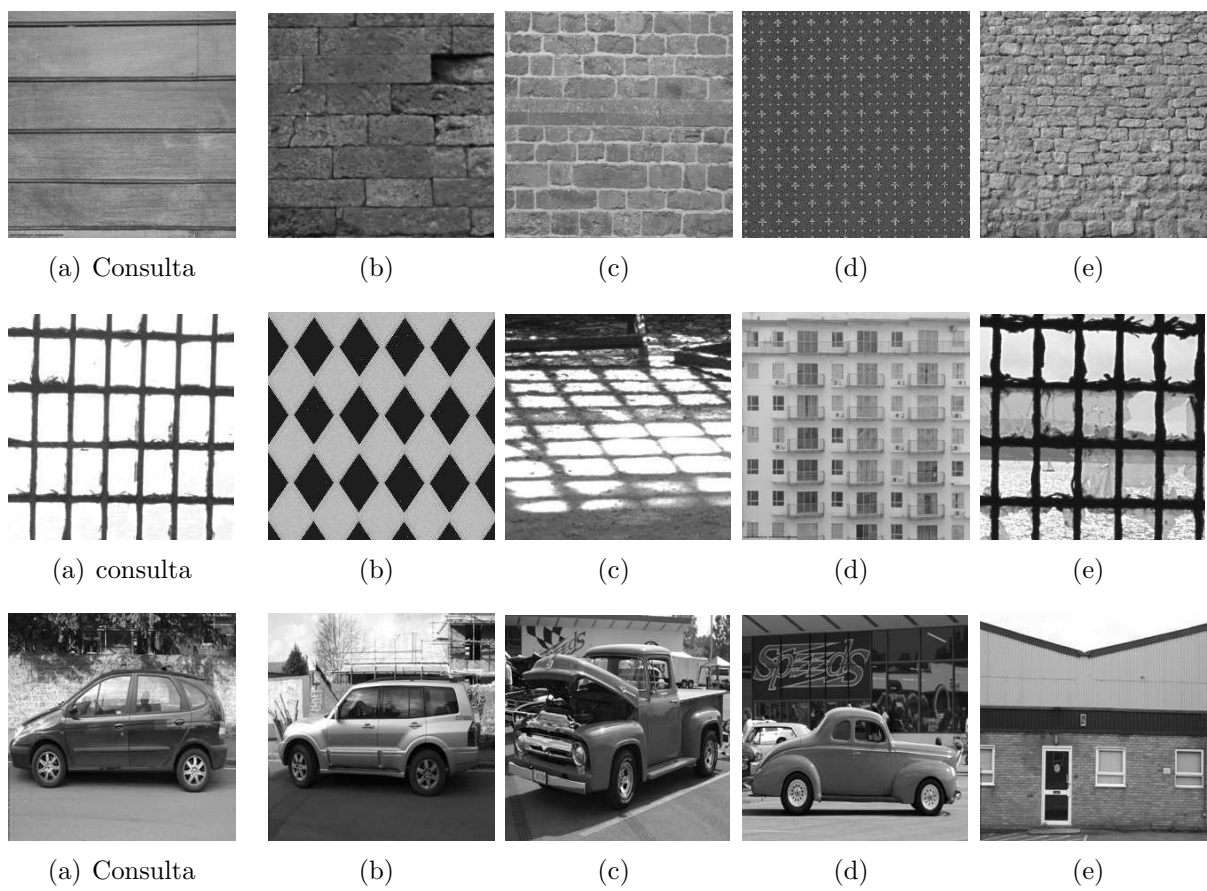


Figura 5.3: Experimentos realizados utilizando o método proposto.

6 Resultados Obtidos

6.1 Introdução

Neste capítulo são apresentados alguns dos experimentos computacionais realizados utilizando o método proposto, o método de Carcassoni et al. [9] e o método da matriz de co-ocorrência de níveis de cinza (GLCM) de Haralick [17]. Os dois primeiros utilizam a análise espectral com propósitos de reconhecimento, enquanto que o modelo GLCM, que é bastante conhecido e utilizado como descritor de padrões de textura, utiliza uma abordagem estatística.

6.2 Banco de Dados de Imagens

Para este trabalho foi utilizada uma coleção de imagens composta por dois bancos de dados (BD) de imagens, sendo um heterogêneo e o outro homogêneo. A coleção contém ao todo 800 imagens, sendo que destas, 350 possuem padrões regulares e semi-regulares (onde os padrões de formas não são totalmente regulares. Exemplo: Figura 6.2(b)). Um banco de dados homogêneo ou de domínio específico, as imagens pertencem a uma mesma categoria, ou seja, possuem uma similaridade implícita umas com as outras. Neste trabalho, as imagens do banco de dados homogêneo são formadas por imagens com padrões regulares ou semi-regulares, como tecidos, papel de parede e de presente. Já em um banco de dados heterogêneo não existe necessariamente uma similaridade entre as imagens que compõem o banco, ou seja, ele é composto por classes de imagens diversificadas e não somente de um domínio específico. Neste tipo de banco de dados os usuários podem realizar consultas com objetivos bastante variados. Enquanto um usuário busca por imagens de paisagens, um outro usuário pode estar interessado em imagens de carros, construções, etc. Neste trabalho, o banco de dados heterogêneo é composto por imagens de árvores, paisagens, carros, construções, nuvens, tijolos, etc.

A tabela 6.1 mostra a divisão da coleção de imagens por banco de dados. Vale ressaltar

que nenhum método de clusterização foi aplicado à coleção de imagens, ou seja, a seleção das imagens de cada BD (homogêneo e heterogêneo) foi realizada por análise visual. O intuito da divisão por categoria de imagens da coleção foi apenas de direcionar e organizar os experimentos. As imagens foram extraídas de diversas fontes, como do álbum digital **Broadatz** [7], dos bancos de dados de imagens **Outex** [28], da Universidade de Oulu, situada na Finlândia e do **Vistex** [1], do Instituto de Tecnologia de Massachusetts, além de outras fontes na *Web*. O tamanho das imagens em estudo é de 256×256 *pixels*, em uma escala de níveis de cinza de 256 cores. A Figura 6.1 exibe uma amostra das classes de imagens contidas no banco de dados utilizado neste trabalho.

BD Heterogêneo	BD Homogêneo
Abstrata	Papel de Parede
Árvores	Papel de Presente
Carros	Tecido
Casas	
Castelos	
Cidades	
Comidas	
Edifícios	
Flores	
Grades	
Montanhas	
Nuvens	
Paisagens	
Papel de Parede	
Papel de Presente	
Pedras	
Plásticos	
Solo	
Tecido	
Tijolos	
Vegetação	

Tabela 6.1: Divisão por classe das imagens dos bancos de dados utilizados nos experimentos.

6.3 Grupos de Experimentos

Afim de direcionar os experimentos e analisar os métodos comparados frente à diferentes tipos de imagens, os experimentos foram divididos em 3 grupos distintos:

- **Primeiro Grupo:** no primeiro grupo de experimentos os três métodos são confrontados utilizando o banco de dados homogêneo (texturas de papel de parede, de

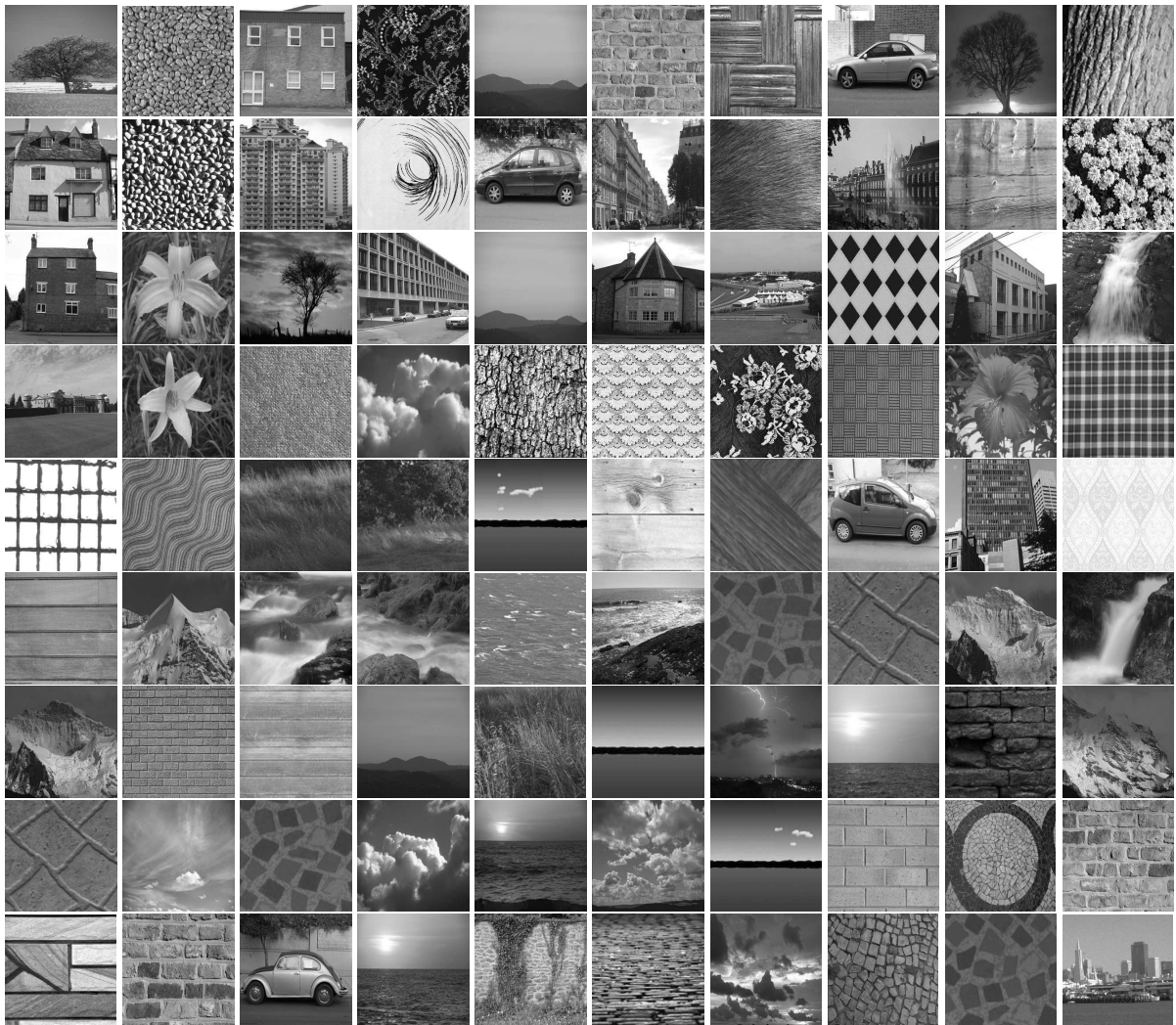


Figura 6.1: Amostra da coleção de imagens utilizada nos experimentos.

presente e tecido).

- **Segundo Grupo:** no segundo grupo de experimentos os três métodos são confrontados utilizando todas as imagens da coleção (bancos de dados homogêneo e heterogêneo).
- **Terceiro Grupo:** no terceiro grupo de experimentos os três métodos são novamente confrontados utilizando todas as imagens da coleção. A diferença entre este e o segundo grupo de experimentos está no domínio das imagens consulta, que neste grupo é formado por imagens com presença de ruído, ampliação (*zoom*) ou com algum tipo de estrago na textura.

As seguintes regras foram estabelecidas em relação aos grupos de experimentos:

1. Assumimos que a imagem consulta não estava na base de dados pois, se estivesse,

ela sempre seria a primeira imagem a ser retornada.

2. Serão analisadas as 10 primeiras imagens do *ranking* recuperado, através da comparação visual das imagens do conjunto de imagens recuperadas e o respectivo conjunto de imagens relevantes da imagem consulta. Serão exibidos os quatro melhores resultados.

Nos resultados exibidos para cada consulta, as imagens estão assim rotuladas: (a-d), (e-g) e (h-l), apresentam os quatro melhores dentre os 10 primeiros do *ranking* obtido pelo método proposto, o de Carcassoni e o GLCM, respectivamente.

6.3.1 Primeiro Grupo de Experimentos

O primeiro grupo de experimentos utiliza somente o banco de dados homogêneo para análise, composto por imagens regulares e semi-regulares de papel de parede, de presente e tecido. Estas imagens possuem em sua maioria padrões repetitivos de formas e cores.

No experimento mostrado na Figura 6.2, uma imagem de tecido com listras verticais uniformemente espaçadas é utilizada como imagem consulta. Notamos que nos resultados dos três métodos comparados estão presentes imagens com padrões similares aos encontrados na imagem consulta, onde a maioria delas apresenta listras na vertical ou horizontal ao longo da textura.

As Figuras 6.3 e 6.4 mostram outros dois experimentos realizados, onde temos como imagens consulta uma imagem de tecido silcado e uma imagem de papel de parede, respectivamente. Em ambos experimentos os métodos comparados também trouxeram como resultado algumas imagens bastante similares às imagens consulta, seja na padronização das formas ou na homogeneidade das cores presentes nas mesmas.

Os resultados destes e de outros experimentos realizados no banco de dados homogêneo nos levaram a concluir que os três métodos comparados são equivalentes quando trabalhando com imagens desta natureza. Em geral, imagens com padrões de formas e cores regulares podem ser mais facilmente interpretadas pelos algoritmos de recuperação, uma vez que na maioria das imagens a distribuição e repetição de formas são constantes ao longo das texturas.

Uma análise de performance entre os modelos comparados utilizando um banco de dados homogêneo é realizada no capítulo 7.

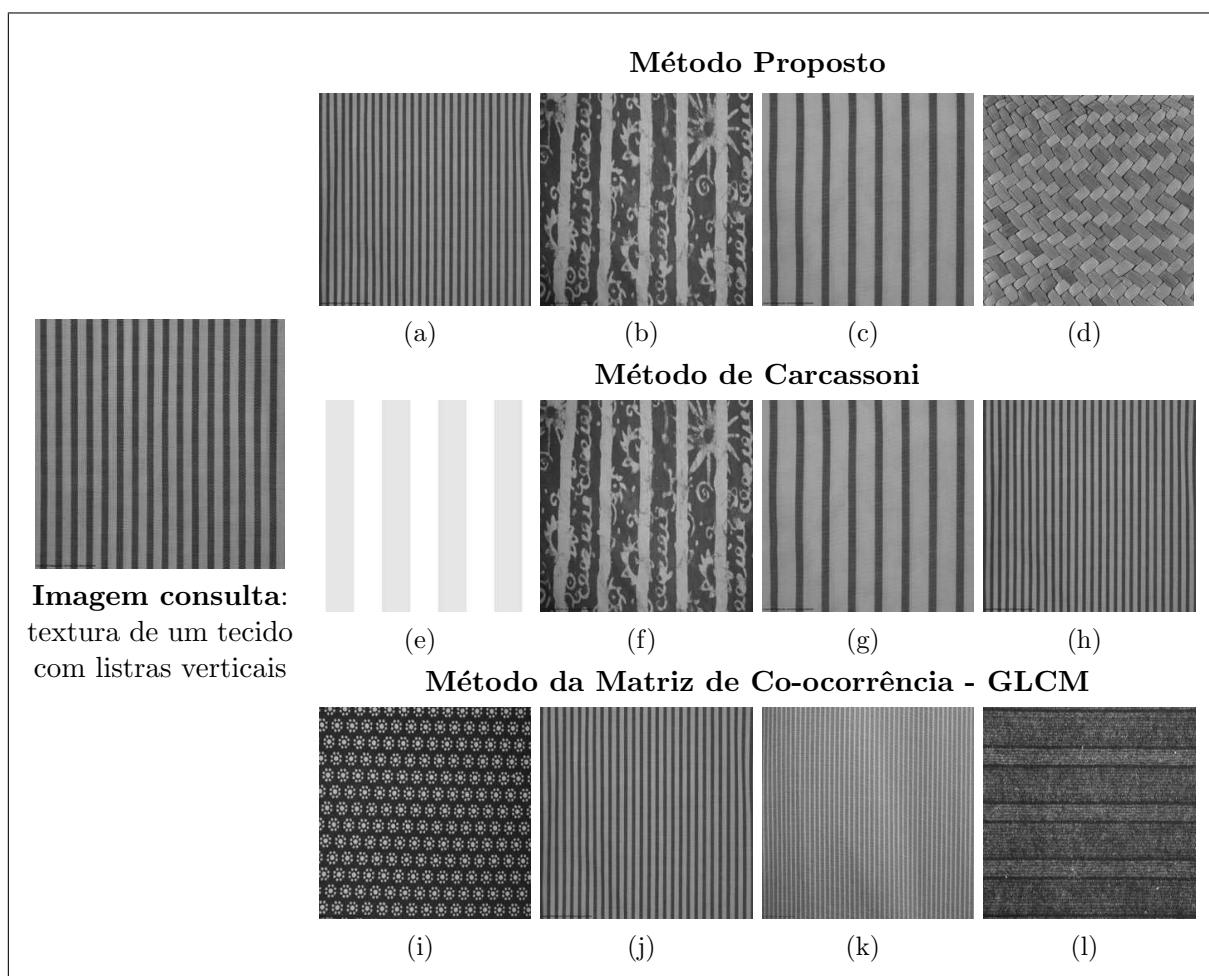


Figura 6.2: Resultados obtidos pelos três métodos comparados: (a-d) mostra os resultados do método proposto, (e-h) os resultados do método de Carcassoni e (i-l) os resultados obtidos pelo método GLCM, de Haralick.

6.3.2 Segundo Grupo de Experimentos

No segundo grupo de experimentos os bancos de dados homogêneo e heterogêneo são utilizados para análise, onde uma imagem consulta submetida aos algoritmos de recuperação dos modelos comparados é confrontada com todas as imagens da coleção. Ampliar o universo das imagens dificulta a obtenção de bons resultados, porém, submeter à prova os modelos de recuperação em tais coleções determinará se os mesmos são capazes de obter bons resultados em uma coleção de imagens não clusterizada, o que constitui uma característica bastante importante à qualquer sistema de recuperação.

No primeiro experimento deste grupo temos como imagem consulta uma imagem de paisagem cujo cenário é formado por montanhas, como mostra a Figura 6.5. Os resultados obtidos pelo método proposto foram bastante expressivos, superando os resultados dos dois outros métodos comparados. Especial atenção pode ser dada às imagens (b) e (c)

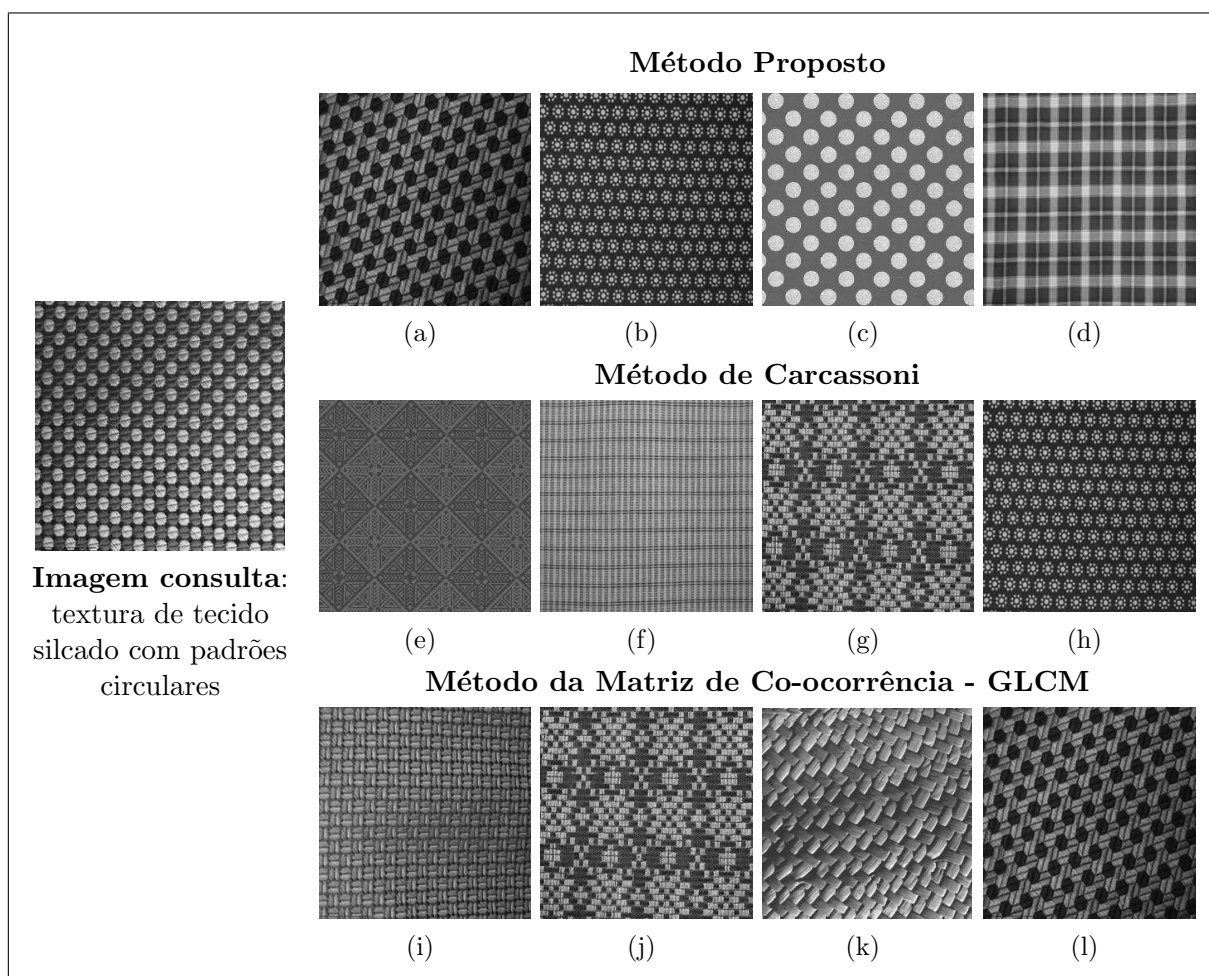


Figura 6.3: Resultados obtidos pelos métodos comparados utilizando a imagem de um tecido silcado como consulta.

deste experimento, recuperadas pelo método proposto, que fazem parte do mesmo cenário da imagem consulta, porém com variações na distância e posicionamento na tomada das imagens.

A Figura 6.6 mostra um outro experimento realizado neste grupo. Uma imagem apresentando uma parede de pedras foi utilizada como consulta. Os três modelos comparados obtiveram bons resultados, uma vez que as imagens recuperadas pertencem à mesma categoria de imagens da imagem consulta, ou seja, paredes de tijolos.

As Figuras 6.7, 6.8 e 6.9 mostram os resultados obtidos pelos modelos comparados utilizando como consulta a imagem de uma paisagem, de uma rua com calçamento de pedras e de uma edificação, respectivamente.

Neste grupo de experimentos, onde toda a coleção de imagens foi utilizada para análise, é perfeitamente normal e aceitável que nem todas as imagens recuperadas apresentem uma similaridade elevada em relação à imagem consulta. Porém, o que pode

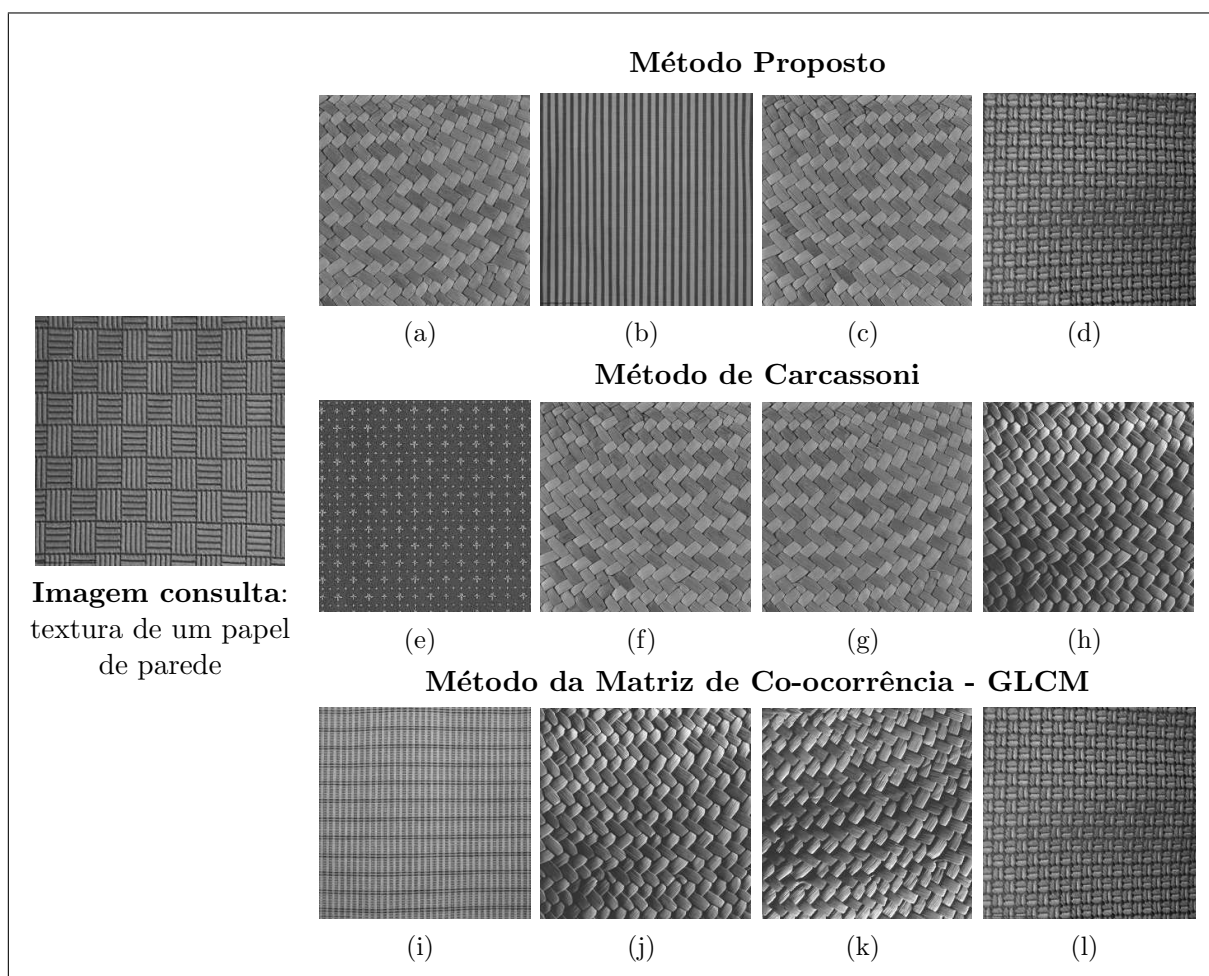


Figura 6.4: Resultados obtidos pela utilização de uma imagem de papel de parede como imagem consulta.

parecer um resultado ruim ou regular para uma consulta, pode ser na verdade um resultado bastante relevante se melhor analisado. Considere como exemplo os resultados obtidos pelo método proposto para o experimento mostrado na Figura 6.7. Neste experimento a imagem consulta é constituída de uma paisagem que possui a figura de uma árvore como destaque na textura. Duas das imagens obtidas pelo método proposto são imagens de outras paisagens, que assim como a imagem consulta, apresentam a figura de uma árvore em seu conteúdo (imagens 6.7 **b** e **d**). Embora as demais imagens retornadas pelo método proposto (imagens 6.7 **a** e **c**) não apresentem a figura de uma árvore em seu conteúdo, existe uma grande similaridade em relação à distribuição das tonalidades de cinza destas imagens em relação à imagem consulta.

Considerando estes e outros detalhes na análise dos resultados dos experimentos aqui apresentados, bem como de outros experimentos realizados, podemos afirmar que o método proposto superou o dois outros métodos em estudo quando todas as classes de imagens foram consideradas pelos algoritmos dos modelos comparados para obtenção dos

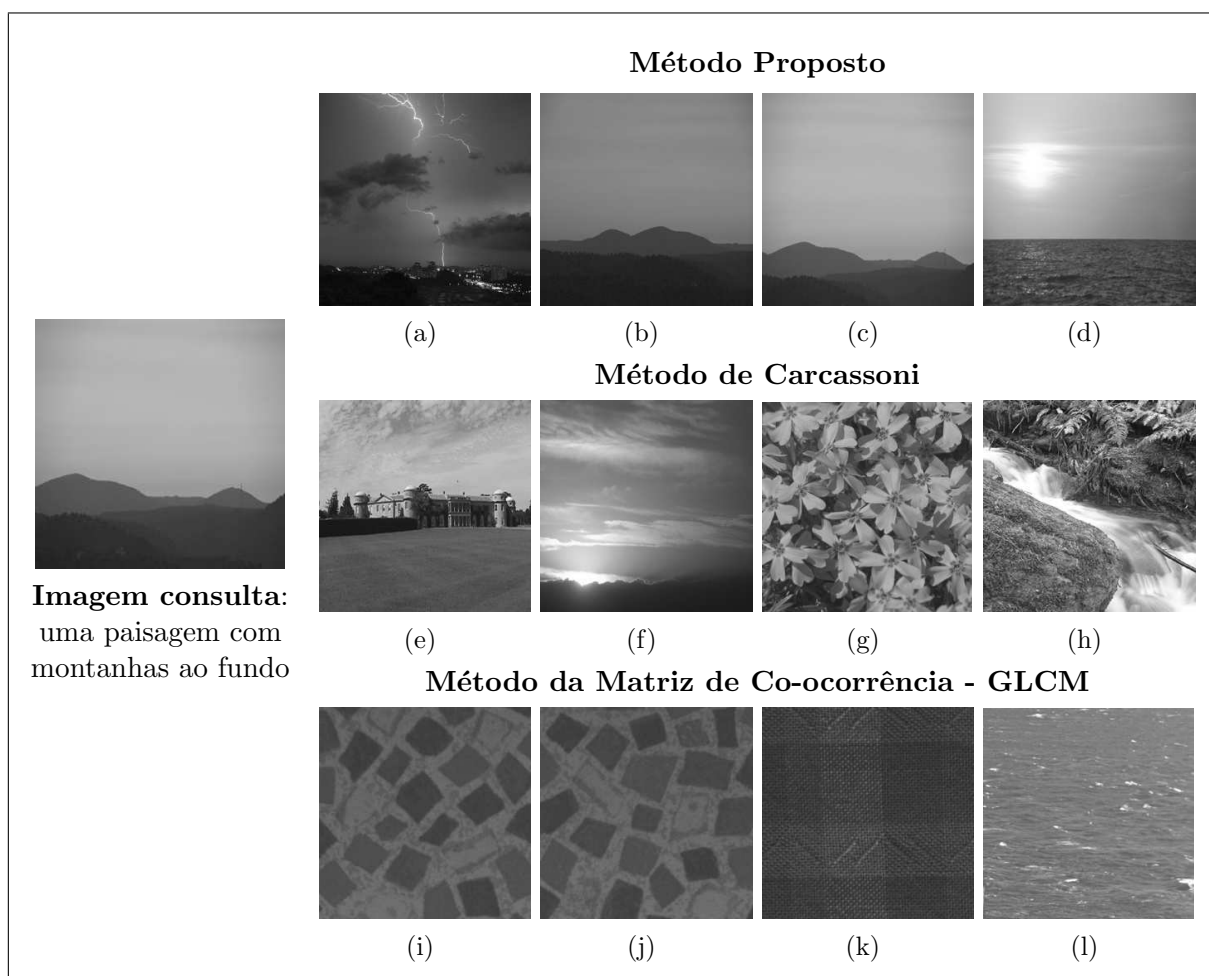


Figura 6.5: Resultados obtidos pelos métodos comparados utilizando toda coleção de imagens, onde a imagem de uma paisagem é utilizada como consulta.

resultados. Uma análise de performance dos modelos neste grupo de experimentos é feita no capítulo 7, através das métricas de *precision* e *recall*.

6.3.3 Terceiro Grupo de Experimentos

No terceiro e último grupo de experimentos o banco de dados de imagens é o mesmo do grupo de experimentos anterior. A diferença está no fato de que o domínio das imagens consulta é formado por imagens modificadas, como imagens com presença de diferentes níveis de ruído Gaussiano, imagens ampliadas (*zoom*) e também por imagens com algum tipo de estrago na textura. Verificar a robustez dos métodos em relação à este tipo de imagem é válido e interessante, uma vez que imagens desta natureza podem ser bastante comuns dependendo de como, de onde e em que circunstâncias elas foram obtidas (exemplo: imagens obtidas por satélites, sondas, robôs, etc).

Nos experimentos à seguir é bom salientar que a imagem consulta se encontra no

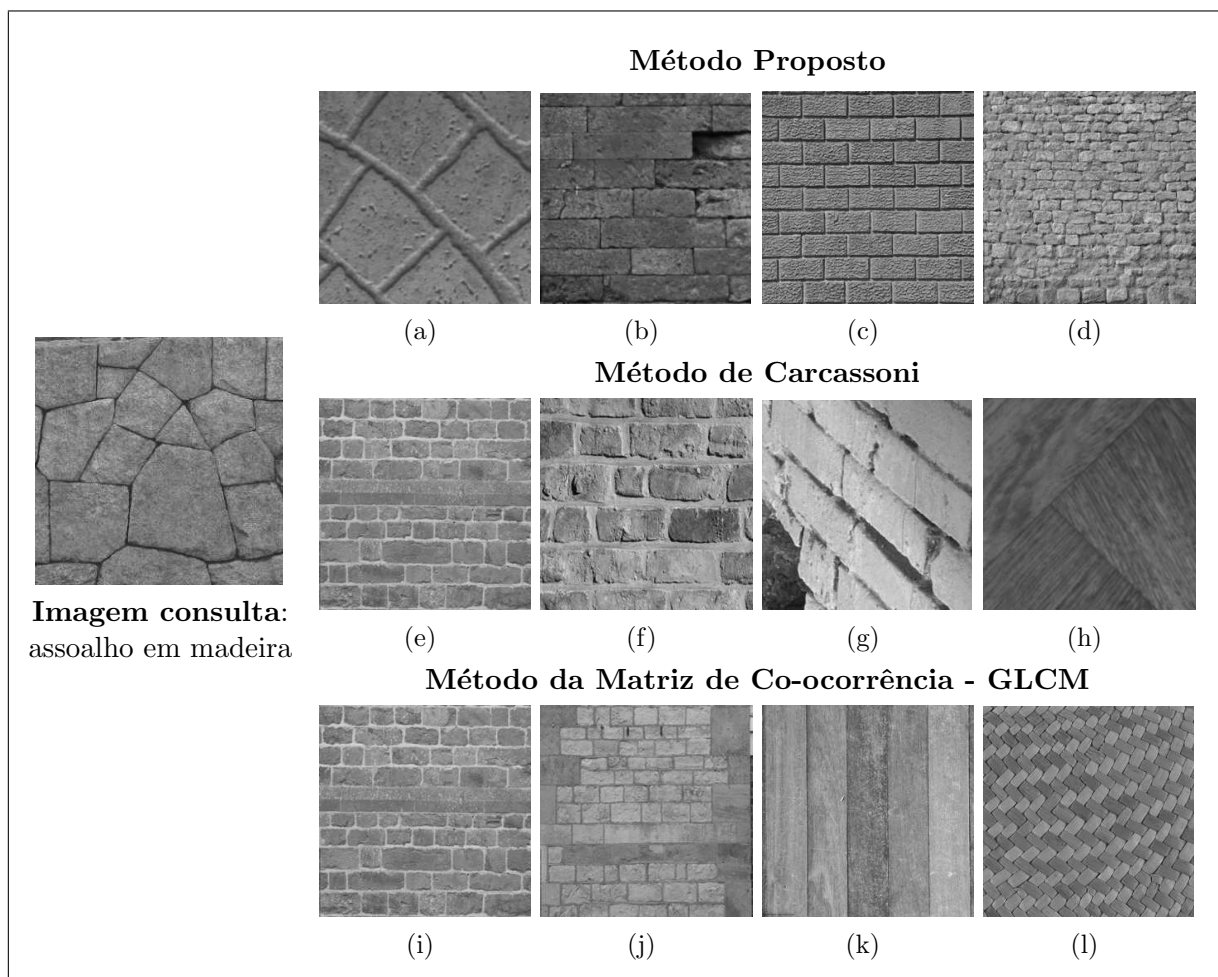


Figura 6.6: Neste experimento foram obtidos resultados expressivos por todos os métodos.

banco de dados na sua versão original, isto é, sem a presença de ruído, estrago ou *zoom*.

No primeiro experimento, mostrado na Figura 6.10, temos como imagem consulta uma imagem de paisagem na qual foi aplicado um ruído Gaussiano de 0 *db*. Os resultados do método proposto mostram que foram recuperadas, além da imagem original, outras imagens que são variações da imagem consulta. O método de Carcassoni também trouxe resultados relevantes, mas menos expressivos do que os do método proposto, enquanto que a presença de ruído afetou bruscamente a performance do método GLCM, onde o ruído e não a imagem foi evidenciado nos resultados obtidos por este modelo, uma vez que as imagens retornadas se parecem com o ruído em si e não somente com a imagem consulta.

Outros experimentos realizados são apresentados nas Figuras 6.11, 6.12 e 6.13. Na Figura 6.11, a consulta é dada por uma imagem com estrago na forma de um risco branco na textura. Na Figura 6.12, a imagem de um tecido com estrago na forma de

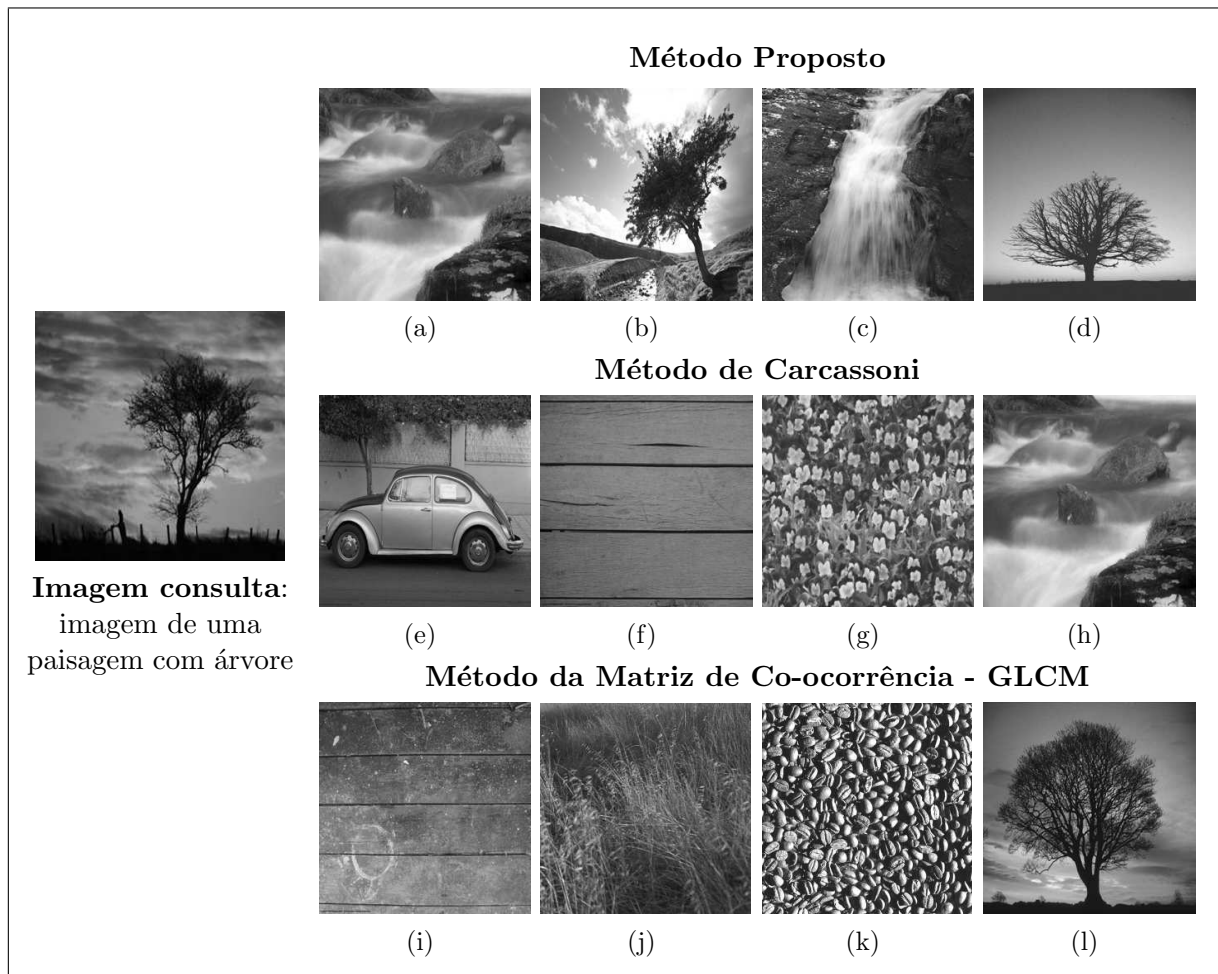


Figura 6.7: Neste experimento uma imagem de paisagem com a figura de uma árvore em seu conteúdo foi utilizada como consulta.

manchas foi utilizado como consulta. A imagem ampliada de uma árvore (ver Figura 6.13) serviu para avaliar os modelos neste domínio de imagens, também bastante comum em algumas áreas como engenharia, geo-processamento, etc. Na maioria dos casos o método proposto superou os dois outros métodos, considerando que os mesmos não foram capazes de recuperar a imagem consulta em sua forma original. Outros experimentos foram realizados com imagens contendo diferentes níveis de ruído Gaussiano, estragos, etc.

No capítulo 7 é feita a análise de performance neste grupo de experimentos, evidenciando através das curvas de *precision* e *recall* uma melhor performance do método proposto.

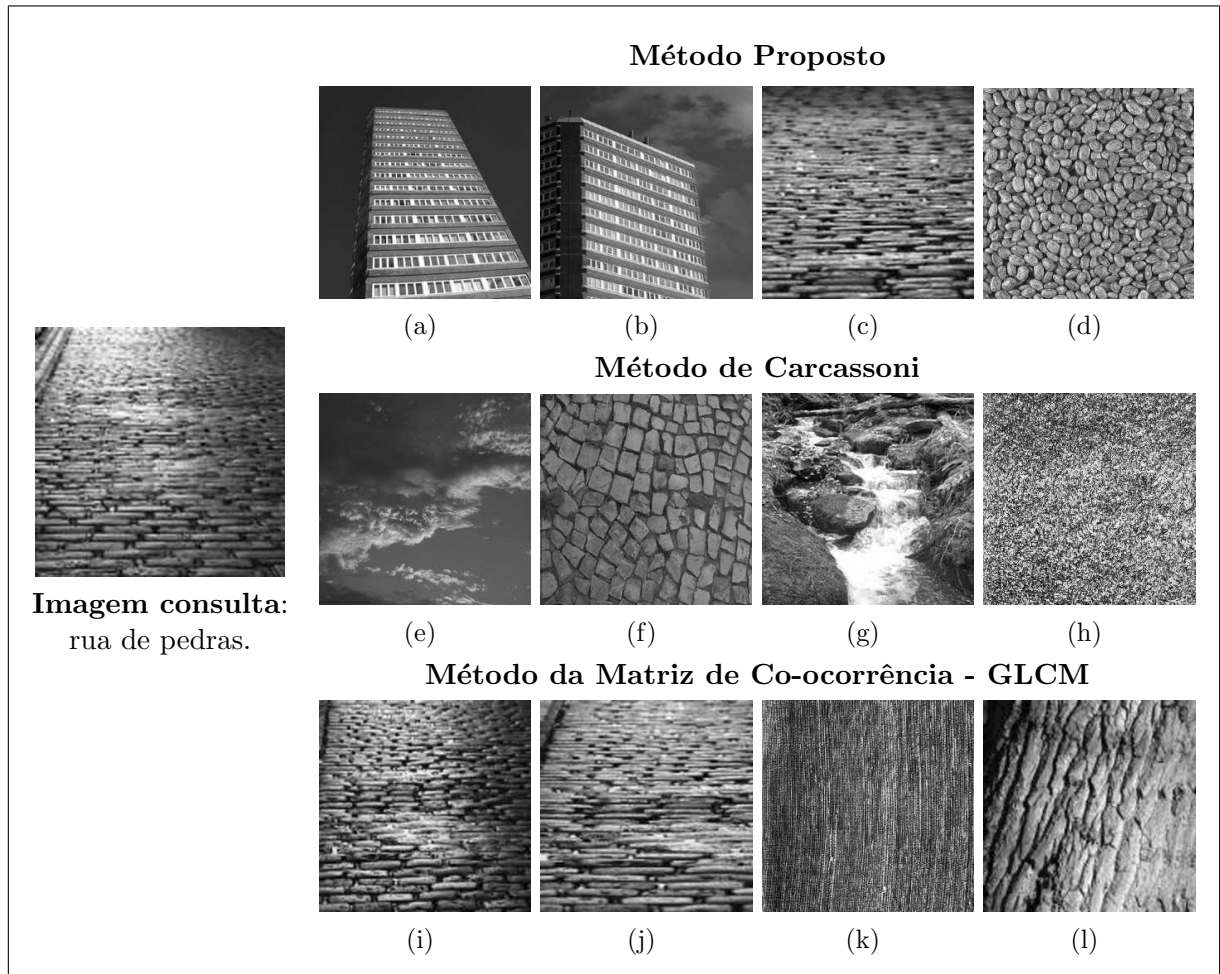


Figura 6.8: A imagem consulta é uma rua com calçamento em pedras.

6.4 Parâmetros do Algoritmo Proposto

O algoritmo proposto requer como entrada alguns parâmetros: S , S_1 , S_2 e k . S é o número de representantes escolhidos para representar cada imagem, S_1 e S_2 representam o número de quanta necessário para agrupar os valores dos picos do espectro de potência nas duas primeiras etapas do processo de quantização e k é a constante utilizada na equação de similaridade. Para o valor de S , valores entre 6 e 8 foram utilizados e como não houveram diferenças significativas nos resultados das consultas, adotamos por questões de performance do algoritmo, $S = 6$. Para o número de quanta foi utilizado $S_1 = S_2 = 10$ enquanto que para k , valores entre 0.1 e 1.5 foram avaliados e $k = 1.2$ foi adotado nos experimentos. A constante k tem a função de tornar os termos da soma mais significativos, não interferindo no somatório.

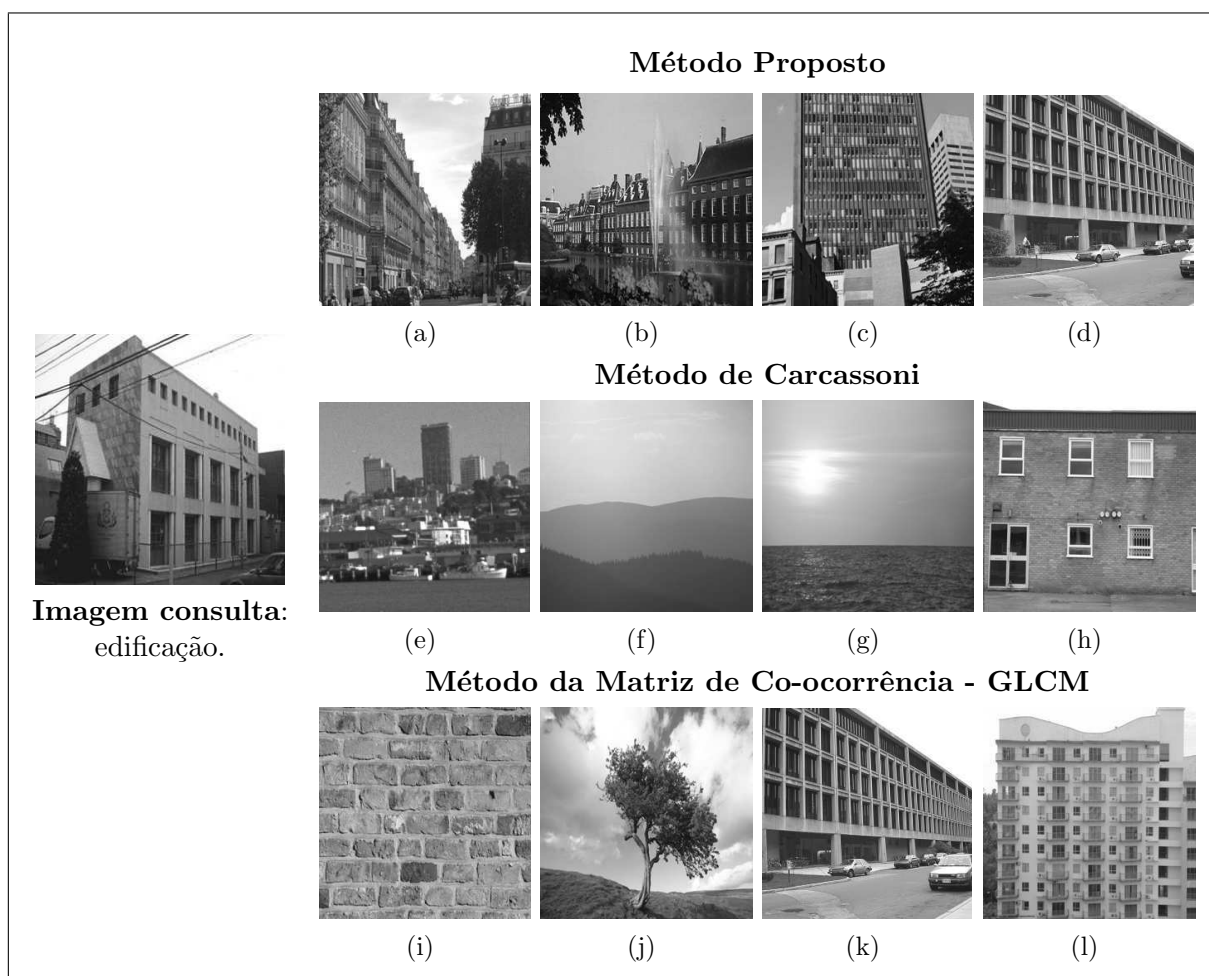


Figura 6.9: Imagens de edificações foram recuperadas por todos os métodos, porém o método proposto trouxe um número maior deste tipo de imagens.

6.5 Tempo de Execução

O tempo de execução utilizando os parâmetros $S = 6$, $S_1 = 10$, $S_2 = 10$ e $k = 1.2$ em um computador com processador Pentium IV de 2Ghz com 512MB de memória principal foi de 0.069 segundos para o procedimento de quantização e para a escolha dos S representantes da imagem consulta (vetor de características) e 0.004ρ segundos para o cálculo de similaridade entre a imagem consulta e todas as imagens do banco de dados de imagens, onde ρ representa o número de imagens no banco de dados.

6.6 Considerações Finais do Capítulo

Alguns dos experimentos computacionais realizados foram apresentados neste capítulo. Os experimentos foram divididos em grupos, avaliando os métodos comparados através de diferentes tipos de imagens, como imagens regulares, semi-regulares, homogêneas, não-

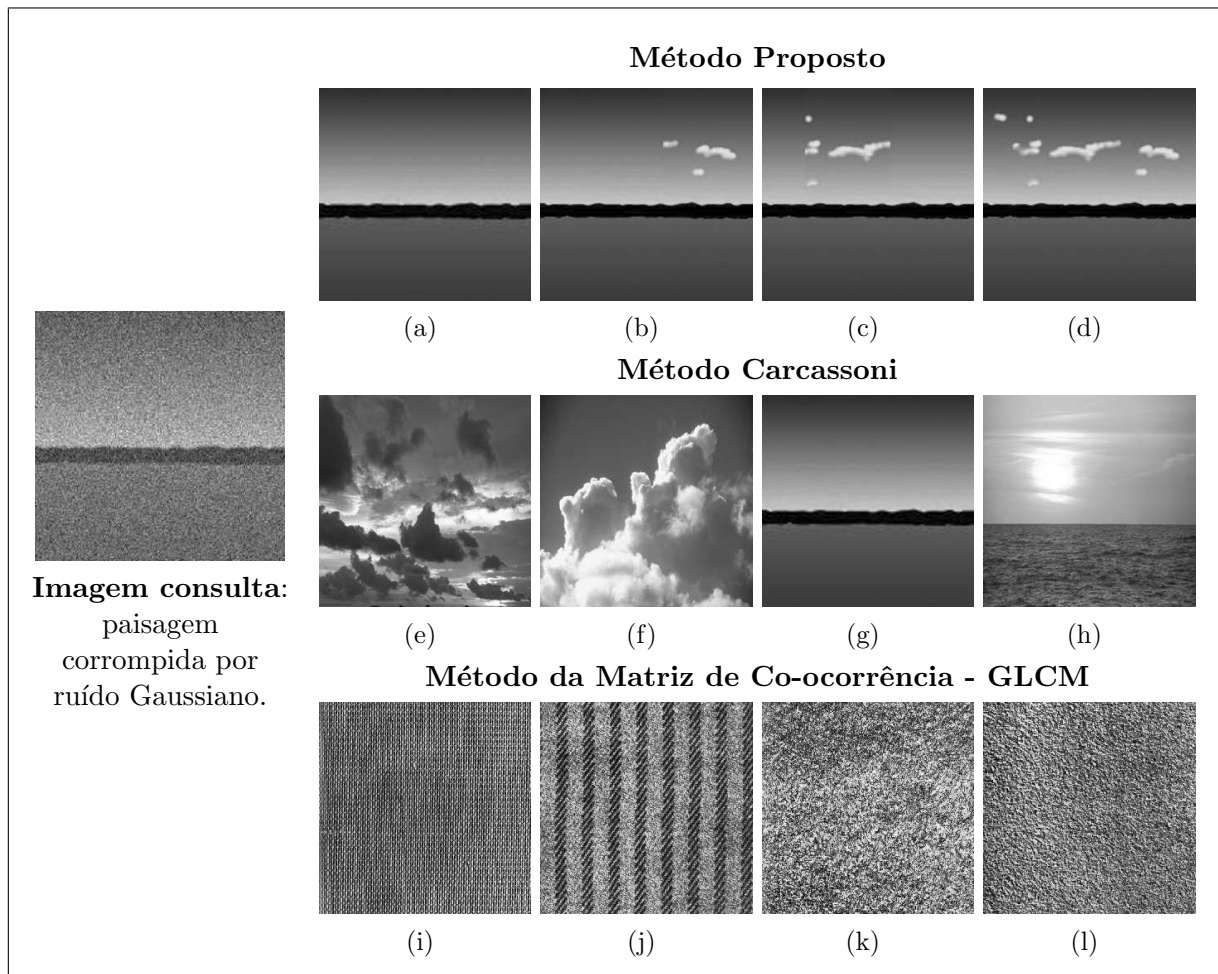


Figura 6.10: Resultados obtidos pelos métodos comparados onde uma imagem de paisagem com ruído Gaussiano de 0 *db* foi utilizada como consulta.

homogêneas, imagens com ruído, ampliadas (*zoom*) e imagens danificadas. Os resultados obtidos pelos métodos comparados no banco de dados homogêneo, contendo imagens regulares e semi-regulares, mostraram uma performance bastante parecida dos três modelos. Porém, adotando toda coleção como domínio das consultas (banco de dados homogêneo e heterogêneo), o método proposto superou os dois outros métodos confrontados.

O próximo capítulo avalia a performance dos modelos comparados nos três grupos de experimentos. São utilizadas as curvas de *precision* e *recall*, as quais constituem um importante instrumento para medir a performance de sistemas de recuperação de informação, sejam eles textuais ou de imagens.

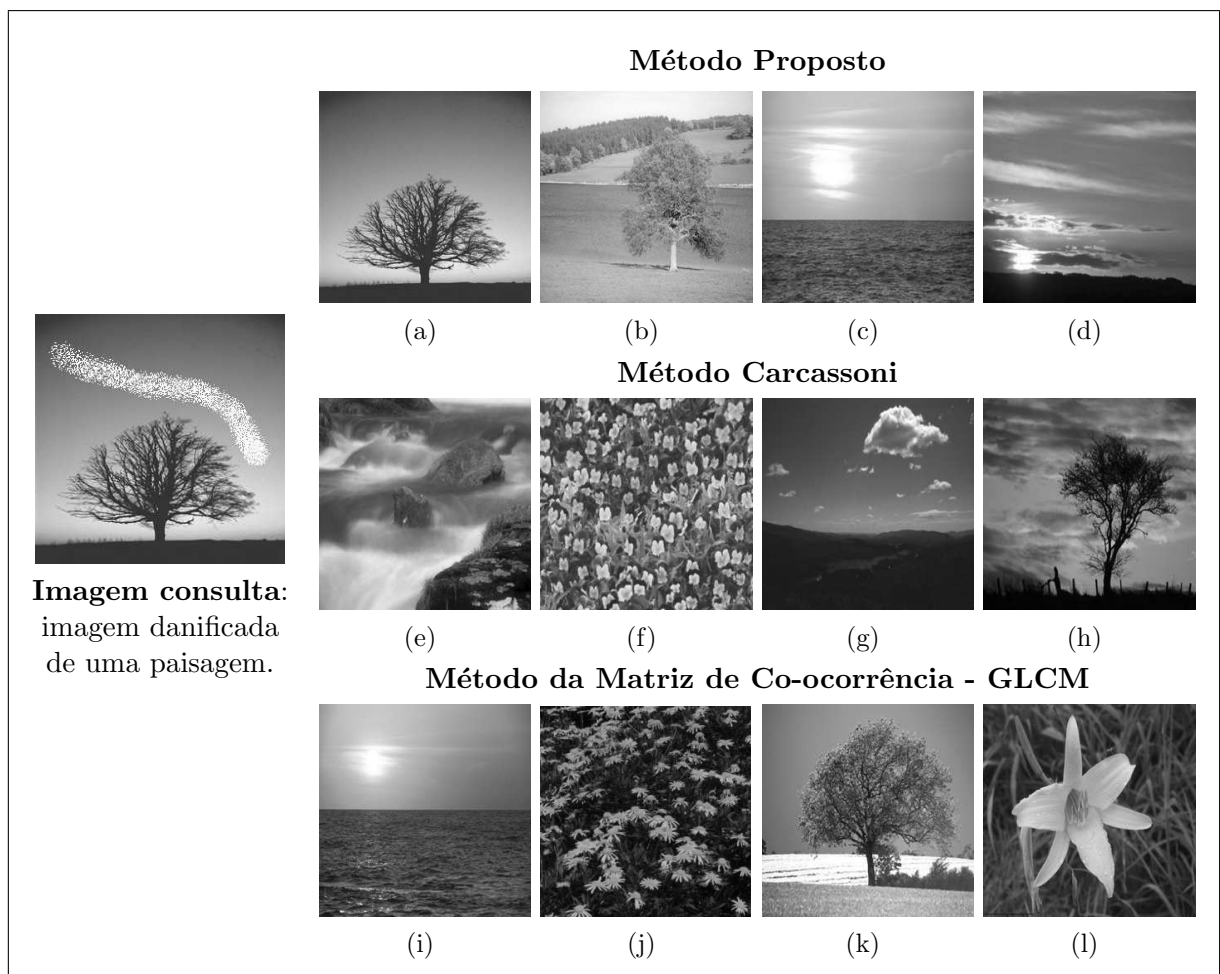


Figura 6.11: Neste experimento temos como imagem consulta uma paisagem danificada por um risco branco ao longo da textura.

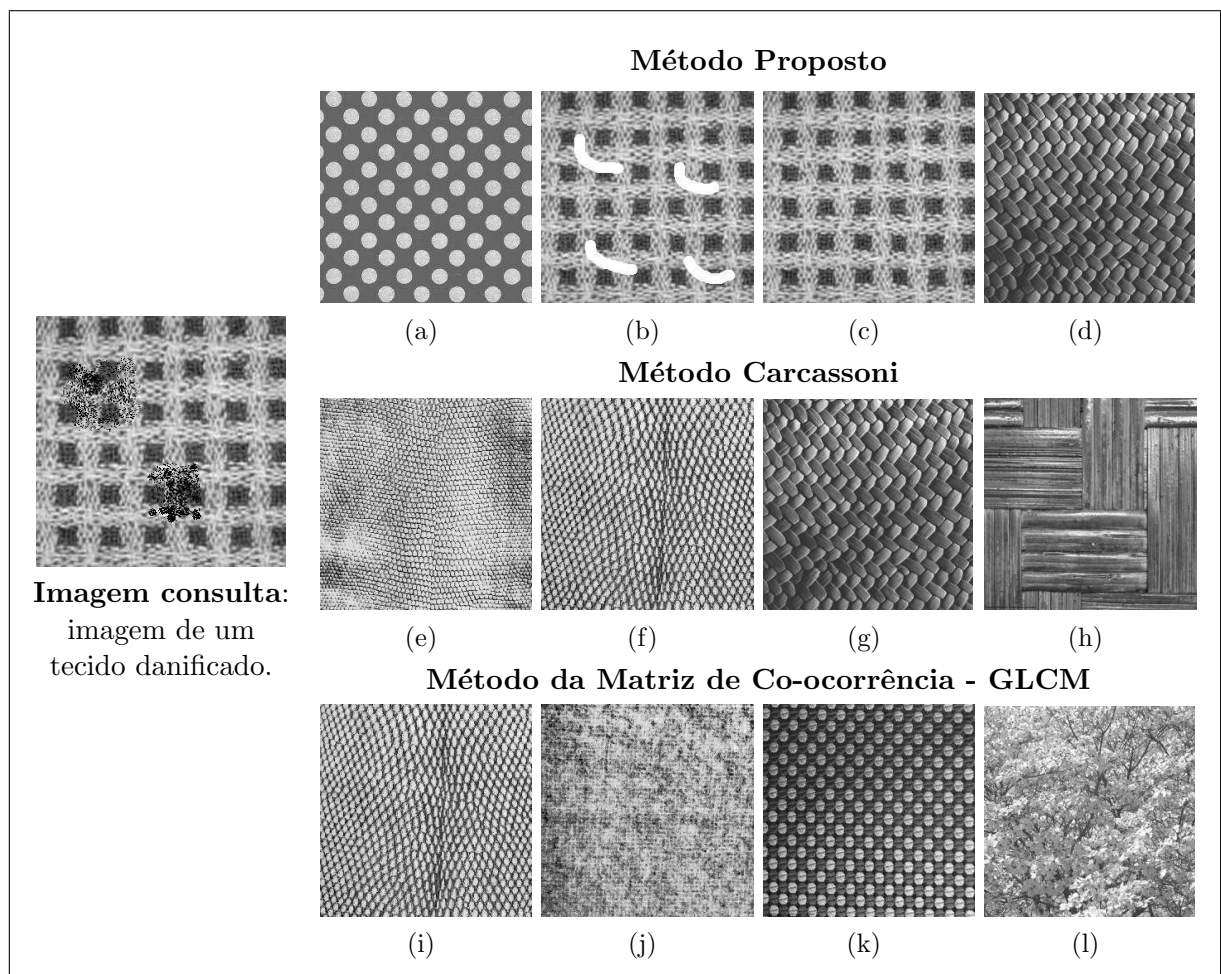


Figura 6.12: Resultados obtidos à partir da consulta feita com a imagem de um tecido danificado com manchas circulares.

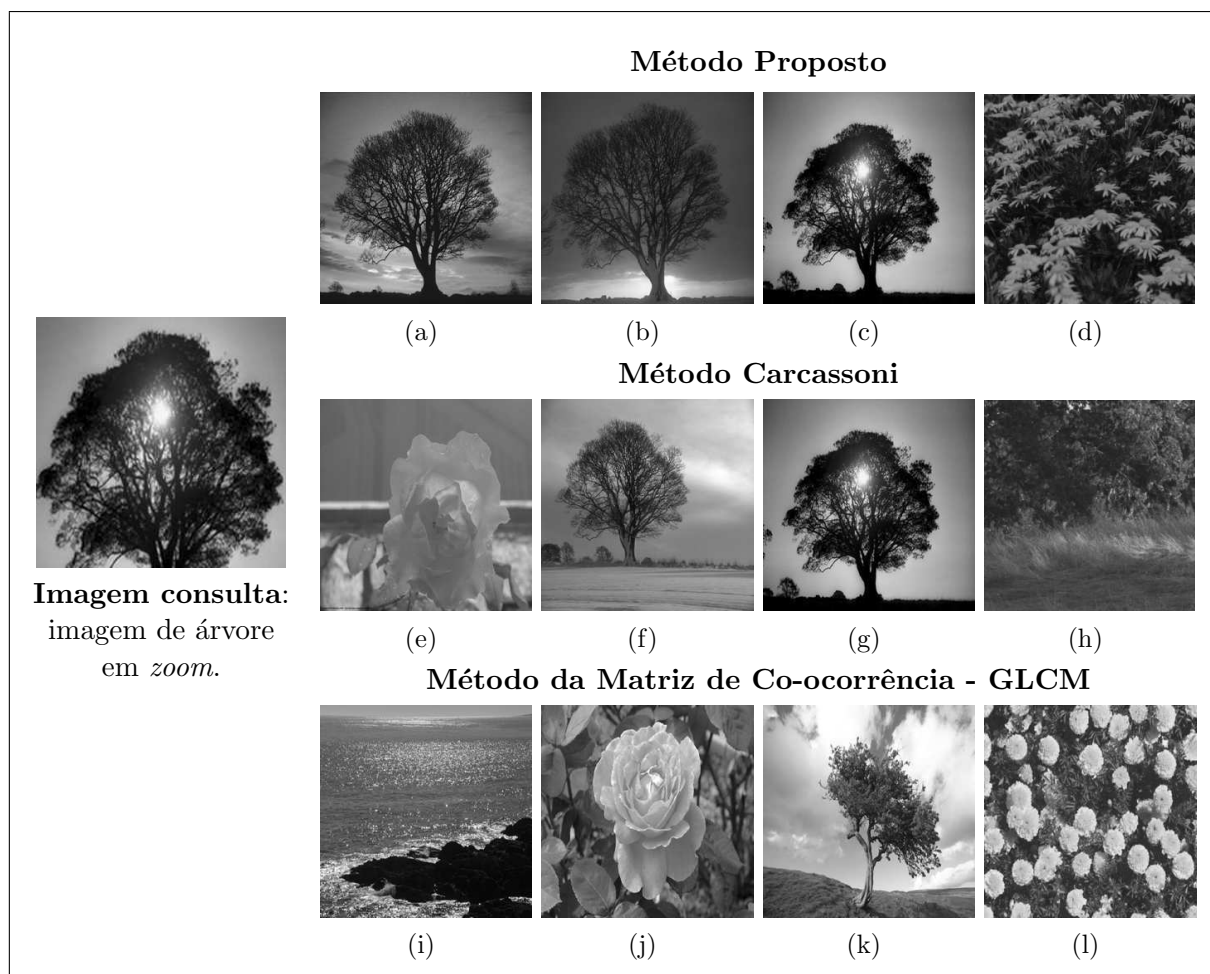


Figura 6.13: Neste experimento a imagem de uma árvore ampliada (*zoom*) foi utilizada como consulta. O método de Carcassoni e o método proposto foram capazes de recuperar a versão normal da imagem (sem *zoom*).

7 Análise dos Resultados

7.1 Introdução

Existem diversas formas de se avaliar um sistema de recuperação de informações, sendo que uma das formas mais comum é através da performance de tempo e espaço. Quanto menor for o tempo de resposta e o espaço utilizado, melhor o sistema será considerado. Porém, existem outras medidas de interesse, como por exemplo, a relevância do conjunto resposta obtido pelo sistema através da submissão de uma consulta. A avaliação da precisão do conjunto resposta é conhecida como avaliação de performance de recuperação, que pode por exemplo ser realizada através das medidas de *precision* e *recall* [3, 5].

7.2 Precision e Recall

Precision é a fração de imagens recuperadas que são relevantes:

$$precision = \frac{|R \cap A|}{|A|}$$

Recall é a fração de imagens relevantes existentes na coleção e que foram recuperadas:

$$recall = \frac{|R \cap A|}{|R|}$$

onde **A** representa um conjunto de imagens recuperadas e **R** o conjunto de imagens relevantes.

7.3 Gráficos de Precision e Recall

Os gráficos de *precision* e *recall* (PR x RE) são normalmente construídos baseando-se em 11 pontos padrões de recall {0%,10%,...100%}. O gráfico PR x RE pode ser realizado

utilizando os resultados de uma única consulta, porém, a forma mais comum é avaliar os algoritmos de recuperação através da execução de várias consultas, gerando para cada uma delas um gráfico de PR x RE e posteriormente, obter um só gráfico que represente todo o conjunto de testes realizados através da precisão média dada pela equação:

$$\bar{P}(r) = \sum_{i=1}^{N_q} \frac{P_i(r)}{N_q}$$

onde $\bar{P}(r)$ é a precisão média em cada nível de *recall* r , N_q é o número de consultas realizadas e $P_i(r)$ é a precisão no nível r para a i -ésima consulta.

7.4 Análise dos Resultados

Nesta seção serão analisados os gráficos de PR x RE referentes à performance dos três modelos comparados nos três grupos de experimentos realizados no capítulo 6.

Foram analisadas cerca de 20 imagens do conjunto resposta (*ranking*) de cada imagem consulta, o que correspondeu ao dobro da média do número de imagens dos conjuntos de relevantes. A análise da performance foi baseada nos resultados obtidos por 30 consultas realizadas utilizando cada um dos três métodos comparados. Cada consulta realizada deu origem a um gráfico de PR x RE, sendo que a performance geral foi dada pela precisão média dos resultados obtidos por cada método em cada grupo de experimentos.

7.4.1 Conjunto de Relevantes

Um conjunto de relevantes pode ser obtido de diversas formas: (i) agrupar as imagens em grupos pré-definidos, baseando-se em características visuais relacionadas; (ii) agrupar as imagens por domínio, por exemplo, imagens médicas, faces, impressões digitais, animais, etc; (iii) adotar o julgamento humano para definir relevância, ou seja, o que o usuário espera como resultado de uma consulta.

Neste trabalho, o conjunto de relevantes de cada imagem consulta foi obtido através do julgamento humano. Para evitar problemas referentes à subjetividade humana, a análise visual das imagens relevantes à cada imagem consulta foi realizada por um grupo de diferentes avaliadores (de áreas não relacionadas à imagens, evitando assim análise de relevância tendenciosa), onde cada qual escolheu para cada imagem consulta um conjunto de imagens relevantes, analisando para tal toda a coleção de imagens. A intersecção entre

os conjuntos de relevantes definidos pelos avaliadores gerou para cada imagem consulta o seu conjunto final de relevantes, o qual foi considerado nos experimentos para julgamento dos resultados obtidos pelos métodos avaliados. A Figura 7.1 mostra algumas das imagens consultas utilizadas nos conjuntos de experimentos e seus respectivos conjuntos de relevantes.

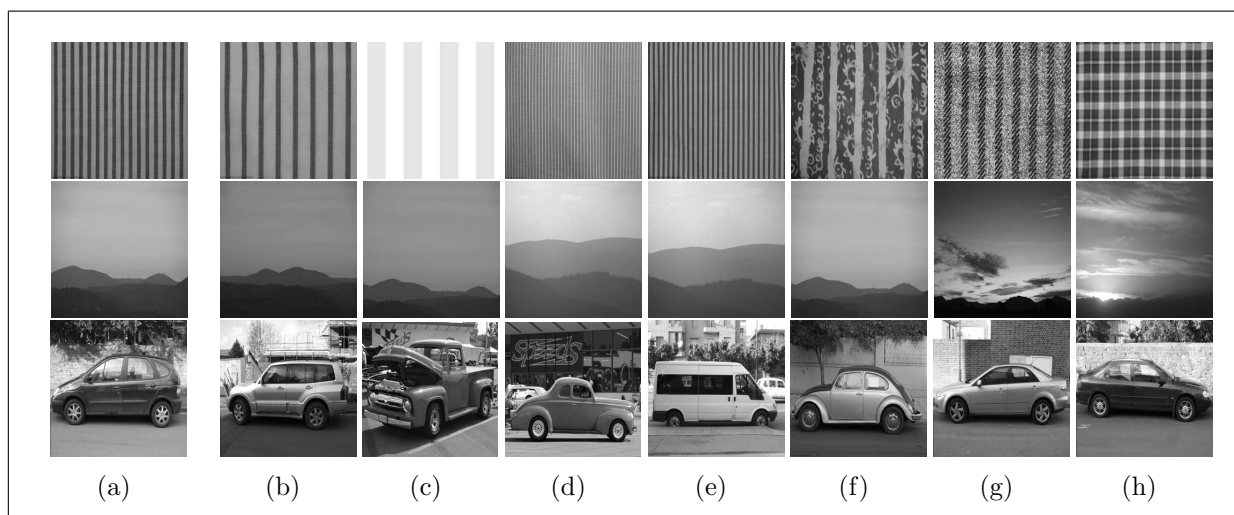


Figura 7.1: Exemplo de imagens relevantes definidas para alguns experimentos realizados. Na coluna (a) temos as imagens consulta.

7.4.2 Performance no Primeiro Grupo de Experimentos

No primeiro grupo de experimentos o método proposto, o de Carcassoni et al. [9] e o método da matriz de co-ocorrência (GLCM) de Haralick [17] foram comparados utilizando um banco de dados homogêneo, contendo imagens regulares e semi-regulares de papéis de parede, de presente e de tecidos. O gráfico da Figura 7.2 mostra as curvas de PR x RE para os três métodos analisados.

Analisando as curvas ao longo deste gráfico podemos considerar que os três modelos tiveram praticamente o mesmo desempenho.

Vale salientar que o objetivo principal de qualquer sistema de recuperação é obter os maiores valores (%) de *precision* e *recall* tanto quanto possível.

7.4.3 Performance no Segundo Grupo de Experimentos

No segundo grupo de experimentos os três métodos são confrontados utilizando os dois bancos de dados (homogêneo e heterogêneo), o que significa considerar imagens regulares,

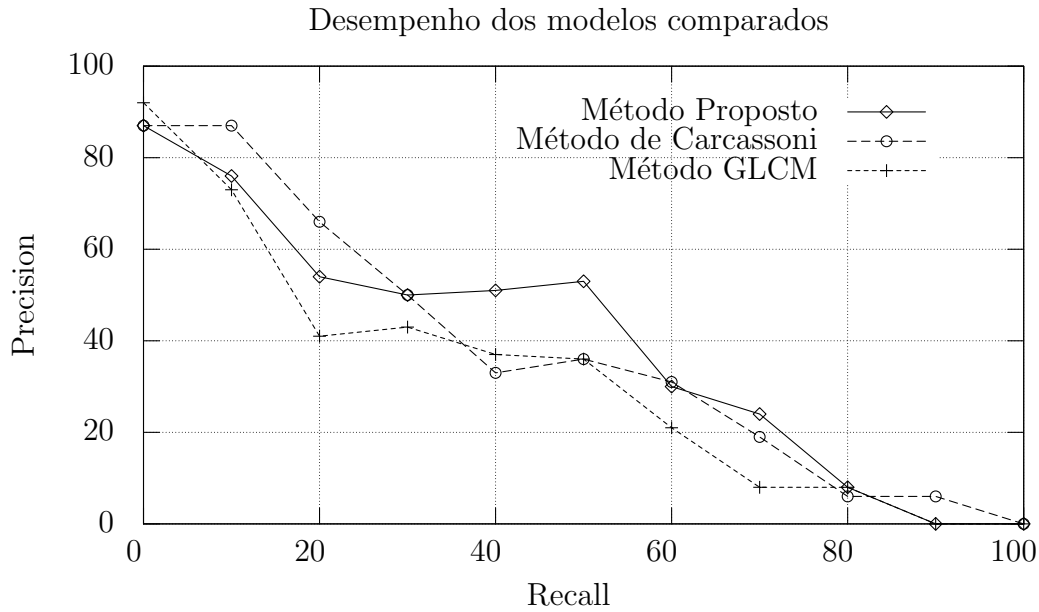


Figura 7.2: Curvas de PR x RE para os três métodos comparados utilizando um banco de dados homogêneo.

não-regulares, homogêneas e não-homogêneas. O gráfico da Figura 7.3 plota as curvas de PR x RE para os três métodos analisados.

Conforme mostrado no gráfico, o método proposto foi praticamente superior ao longo de toda a curva de PR x RE. Na área de alta precisão do gráfico (onde o *recall* $\leq 20\%$) o método de Carcassoni teve um desempenho similar ao do método proposto, porém este não se manteve ao longo da curva.

7.4.4 Performance no Terceiro Grupo de Experimentos

No terceiro grupo de experimentos as imagens consulta possuíam alguma modificação, como a presença de ruído, estrago ou ampliação da imagem. O gráfico da Figura 7.4 mostra as curvas para os três métodos comparados.

Neste grupo de experimentos, o gráfico PR x RE mostra o método proposto superior aos dois outros métodos durante toda a curva de desempenho. A abrangência em relação à recuperação de imagens relevantes, medida pelo *recall*, nos mostra que cerca de 60% delas foram recuperadas pelo método proposto, enquanto que apenas 20% e 30% foram conseguidos pelo método GLCM e o de Carcassoni, respectivamente.

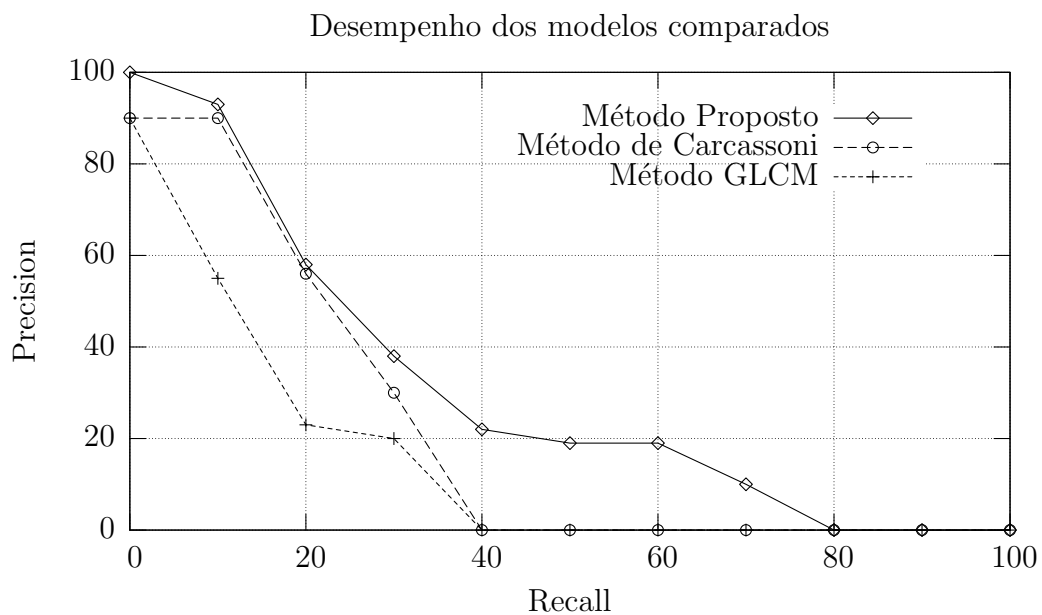


Figura 7.3: Curvas de PR x RE para os três métodos comparados utilizando todas as imagens da coleção.

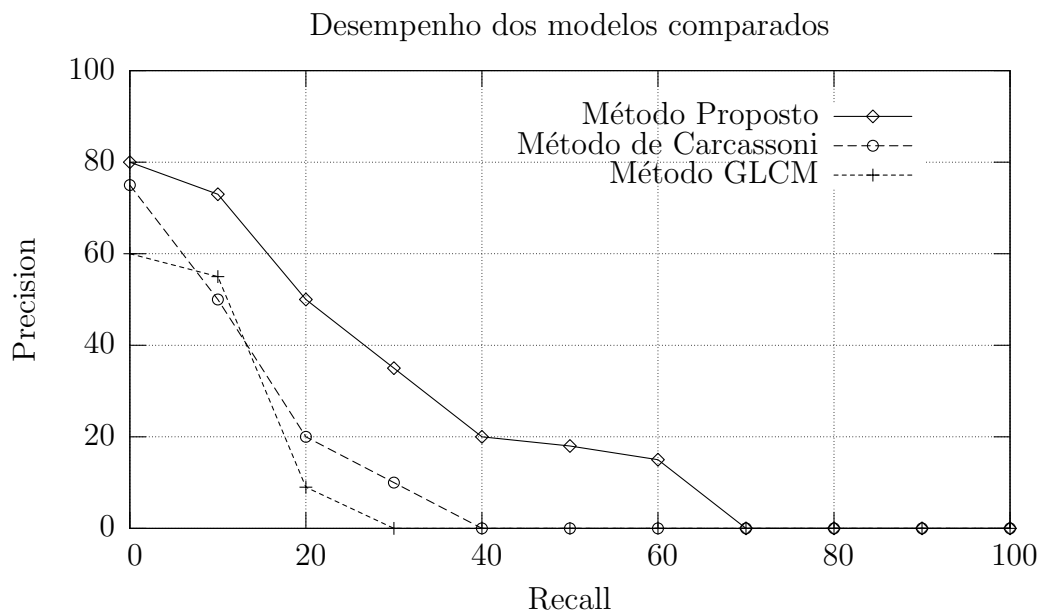


Figura 7.4: Curvas de PR x RE para os três métodos comparados no terceiro grupo de experimentos. As imagens consulta apresentam diferentes níveis de ruído Gaussiano, estragos, etc.

7.5 Considerações Finais do Capítulo

A avaliação de um sistema de recuperação pode ser medida através das curvas de *precision* e *recall*. Tais curvas constituem um instrumento importante na RI textual e vêm ao longo dos últimos anos sendo aplicada também na análise de performance de sistemas CBIR. A análise das curvas de PR x RE nos proporciona uma visão geral sobre o desempenho do sistema, onde é possível verificar a precisão e abrangência em relação às imagens relevantes que foram recuperadas. Dos três métodos avaliados (método proposto, de Carcassoni e da matriz de co-ocorrência (GLCM) de Haralick) nos três grupos de experimentos realizados, verificamos através das curvas PR x RE que no primeiro grupo de experimentos, onde somente imagens regulares e semi-regulares (onde os padrões de formas não são totalmente regulares. Exemplo: Figura 6.2(b)) foram utilizadas, que os três métodos tiveram resultados bastante parecidos. Analisando o gráfico PR x RE do segundo grupo de experimentos, verificamos que o método proposto foi superior aos dois outros métodos comparados, porém com uma vantagem não tão significativa quanto àquela encontrada no terceiro grupo de experimentos, cuja análise do gráfico de PR x RE nos mostra a superioridade do método proposto em relação aos dois outros métodos, quando utilizando como consulta imagens com presença de ruído, estrago e outras alterações na textura.

8 Conclusão, Contribuições e Trabalhos Futuros

Neste trabalho foi proposta uma nova técnica de recuperação de imagens baseada em conteúdo, onde padrões de textura são extraídos através do espectro de potência da imagem e quantizados de forma a se obter uma representação sucinta da mesma. O casamento (*matching*) é feito através da análise da similaridade entre a relação dos pontos representantes do espectro de potência da imagem consulta e das imagens armazenadas no banco de dados. O trabalho foi inspirado no modelo de recuperação de texturas de Carcassoni et al. [9]. Para se obter os pontos mais significativos do espectro de potência foi utilizada uma variação da técnica de quantização de Lloyd [37].

Diversos experimentos computacionais foram realizados utilizando além do método proposto, o método de Carcassoni et al. [9] e o método de extração de padrões de textura de Haralick, que utiliza a matriz de co-ocorrência de níveis de cinza como descritor de textura. Os experimentos foram divididos em três grupos distintos, sendo um com imagens homogêneas, um outro contendo toda a coleção de imagens (homogêneas e heterogêneas) e um terceiro, onde imagens especiais (ruído, estrago, etc) foram utilizadas como imagens consulta, considerando para análise toda a coleção de imagens. Analisando a performance dos três métodos comparados nos três grupos de experimentos através de gráficos de *precision* e *recall*, temos que o método proposto mostrou-se mais robusto que os dois outros métodos, no segundo e terceiro grupo de experimentos, enquanto que uma performance similar foi obtida no primeiro grupo pelos três modelos comparados.

Embora nos experimentos envolvendo texturas uniformes e não uniformes o método proposto tenha alcançado bons resultados, está sendo objeto de corrente estudo como a clusterização da coleção de imagens poderia fornecer melhor performance de recuperação, possibilitando desta forma aumentar a coleção de imagens a qualquer tamanho desejado.

Outro ponto sob investigação mais aprofundada é sobre a escolha do número S de pontos para representar uma imagem. É interessante conhecer o comportamento do método

proposto quando esse número aumenta ou diminui, de forma a encontrar um S que melhor represente toda a coleção de imagens.

Estudos em relação à aplicação do método proposto em imagens coloridas estão em andamento, bem como a aplicação do método em um domínio de imagens mais especializado, como imagens médicas, faces, etc.

APÊNDICE A - Experimentos Adicionais

Número de Representantes

O método proposto neste trabalho utiliza os S maiores representantes extraídos do processo de quantização para representar uma imagem. Uma faixa de valores de S foram avaliados durante os experimentos, $S = \{6, 8, 10\}$. Destes, $S = 6$ foi o que melhor representou a imagem, considerando que este retornou um *ranking* constituído de mais imagens relevantes em um conjunto de experimentos. O experimento mostrado na Figura A.1 mostra os resultados obtidos para uma consulta, utilizando diferentes valores para S .

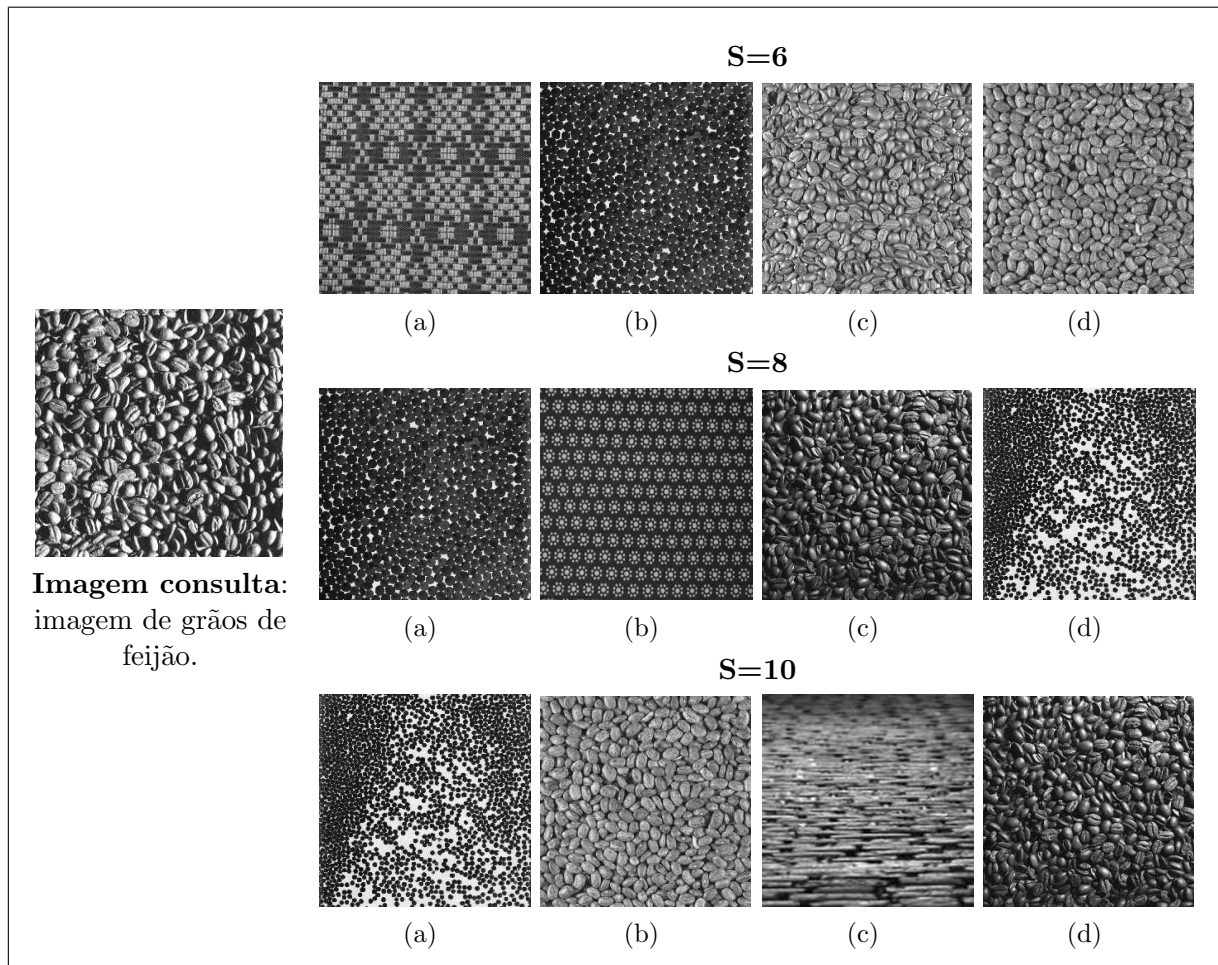


Figura A.1 : Resultados obtidos utilizando diferentes valores para S .

APÊNDICE B - Sistema CBIR Proposto

Interface do Usuário

A Figura B.1 mostra a interface do usuário do sistema de recuperação de imagens proposto. A aplicação foi desenvolvida na ferramenta Microsoft Visual C++. Na Figura B.2 temos o *ranking* retornado como resultado de uma consulta.

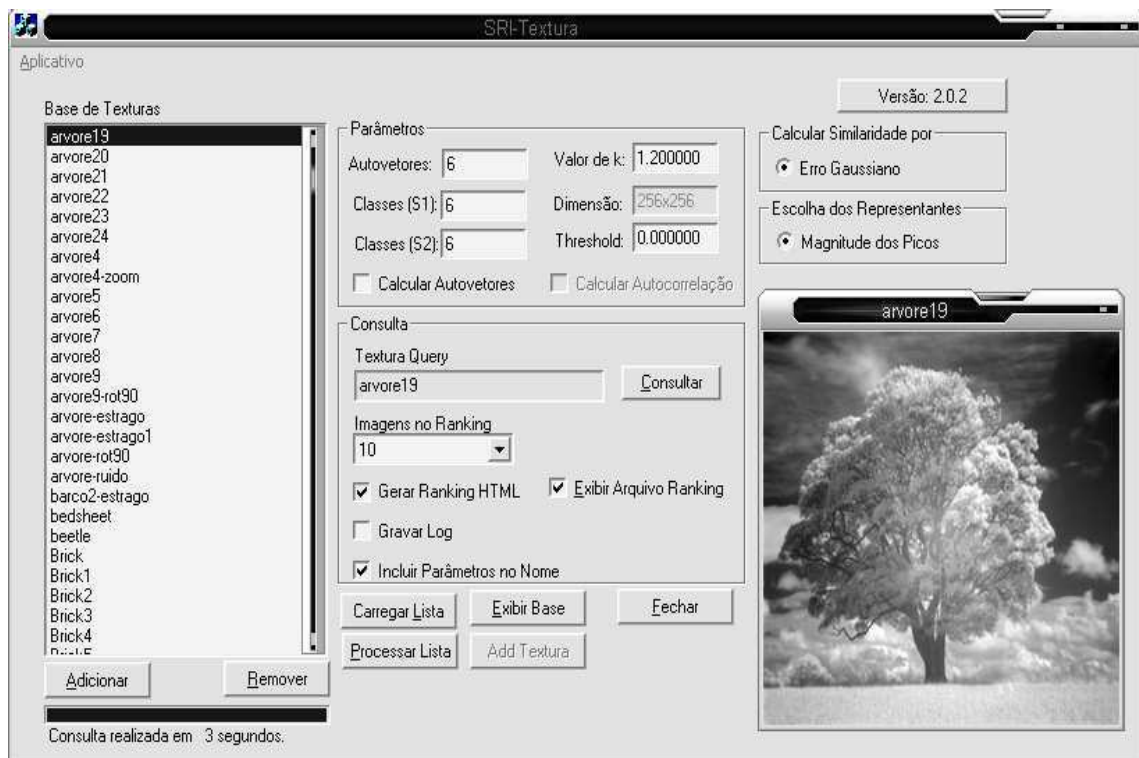
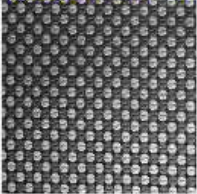


Figura B.1 : Interface do usuário do sistema CBIR proposto.

Nro de imagens	352	Dimensão	256x256	Data consulta	26-Sep-2005 13:19	Threshold	0.000000
Nro de AV	6	Valor de K	1.200000	Nro de classes	6	Mínimo	8.721537
				Eq.Calc.Sim.	Erro Gaussiano	M.E.Rep.	Mg.Picos

Query: silk tie 2



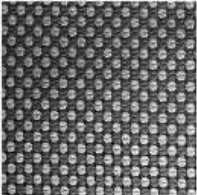
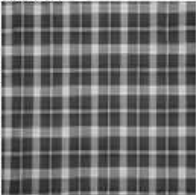
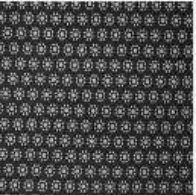
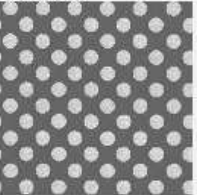
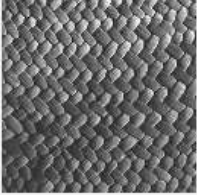
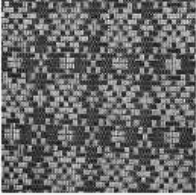
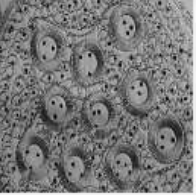
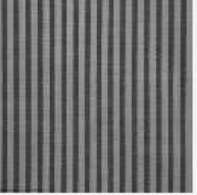
(1) 8.721537 silk_tie 2	(2) 8.694851 tecido	(3) 7.750692 silk_3	(4) 7.680401 tie	(5) 7.233994 Fabric06
				
(6) 7.221292 Fabric17	(7) 7.196329 silk_4	(8) 7.144196 silk_2	(9) 7.130188 fibra	(10) 7.119628 listras
				

Figura B.2 : *Ranking* do método proposto retornado pela submissão de uma imagem consulta.

Referências

- [1] Vistex: Vision texture database. Maintained by the Vision and Modeling group at the MIT Media Lab. <http://whitechapel.media.mit.edu/vismod/>, 1995.
- [2] E. Albuz, E. Kocalar, and A. A. Khokhar. Scalable color image indexing and retrieval using vector wavelets. *IEEE Transactions on Knowledge and Data Engineering*, 13(5):851–861, 2001.
- [3] R. A. Baeza-Yates and B. A. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.
- [4] C. A. Z. Barcelos, M. J. R. Ferreira, and M. L. Rodrigues. Textured image retrieval by the modal analysis of quantized spectral point patterns. In *ICTAI 2004 - The 16th IEEE International Conference on Tools with Artificial Intelligence*, pages 771–773, 2004.
- [5] C. A. Z. Barcelos, M. J. R. Ferreira, and M. L. Rodrigues. Texture image retrieval: A feature-based correspondence method in fourier spectrum. In *3rd ICAPR - International Conference on Advances in Pattern Recognition*, pages 424–433, 2005.
- [6] F. Le Bourgeois. Content based image retrieval using gradient color fields. In *ICPR*, pages 5027–5030, 2000.
- [7] P. Brodatz. *A Photographic Album for Artists and Designers*. Dover, New York, 1966.
- [8] M. Carcassoni and E. R. Hancock. Point pattern matching with robust spectral correspondence. In *CVPR*, pages 1649–1655. IEEE Computer Society, 2000.
- [9] M. Carcassoni, E. Ribeiro, and E. R. Hancock. Texture recognition through modal analysis of spectral peak patterns. In *ICPR (1)*, pages 243–246, 2002.
- [10] Ma. Das, E. M. Riseman, and B. A. Draper. Focus: Searching for multi-colored objects in a diverse image database. In *CVPR*, pages 756–761, 1997.
- [11] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.
- [12] S. A. Dudani, K. J. Breeding, and R. B. McGhee. Aircraft identification by moment invariants. *IEEE Trans. Computers*, 26(1):39–46, 1977.
- [13] M. J. R. Ferreira, C. A. Z. Barcelos, and M. L. Rodrigues. Recuperação de texturas através da análise modal de pontos quantizados como padrões espectrais. In *WTDCGPI - Sibgrapi 2004 - XVII Brazilian Symposium on Computer Graphics and Image Processing*, Curitiba-Brasil, 2004.

- [14] M. Flickner, H. S. Sawhney, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The qbic system. *IEEE Computer*, 28(9):23–32, 1995.
- [15] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.
- [16] A. Haar. Zur theorie der orthogonalen funktionen-systeme. *Math. Ann.*, 69, pages 331–371, 1910.
- [17] R. M. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. *IEEE Trans. on SMC*, 3(11):610–622, 1973.
- [18] M. K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. Inform. Theory*, pages 179–187, 1962.
- [19] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, and R. Zabih. Image indexing using color correlograms. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, page 762, Washington, DC, USA, 1997. IEEE Computer Society.
- [20] L. M. Kaplan, R. Murenzi, and K. R. Namuduri. Fast texture database retrieval using extended fractal features. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 162–175, 1998.
- [21] S.M. Kay. *Modern Spectral Estimation, Theory and Application*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.
- [22] M. Kokare, P. K. Biswas, and B. N. Chatterji. Rotated complex wavelet based texture features for content based image retrieval. In *ICPR (1)*, pages 652–655, 2004.
- [23] H. Lin, Y. Kao, S. Yen, and C. Wang. A study of shape-based image retrieval. In *ICDCS Workshops*, pages 118–123, 2004.
- [24] H. J. Lin and Y. T. Kao. A prompt contour detection method. *International Conference On the Distributed Multimedia Systems*, 2001.
- [25] S. P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–136, 1982.
- [26] B. S. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(8):837–842, 1996.
- [27] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. H. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The qbic project: Querying images by content, using color, texture, and shape. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 173–187, 1993.
- [28] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinene. Outex - a new framework for empirical evaluation of texture analysis algorithms. In *Proc. 16th Intl. Conf. Pattern Recognition 2002*, 2002.
- [29] M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra, and T. S. Huang. Supporting similarity queries in MARS. In *ACM Multimedia*, pages 403–413, 1997.

- [30] Y. Rui, T. S. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. In *ICIP (2)*, pages 815–818, 1997.
- [31] E. Sangineto. Shape similarity image retrieval by hypothesis and test. In *ICPR (2)*, pages 1017–1020, 2004.
- [32] G. Schaefer, S. Lieutaud, and G. Qiu. Cvpic image retrieval based on block colour co-occurrence matrix and pattern histogram. In *ICIP*, pages 413–416, 2004.
- [33] G. Scott and H. L. Higgins. An algorithm for associating the features of two patterns. volume B244, pages 21–26, 1991.
- [34] L. S. Shapiro and M. Brady. Feature-based correspondence: an eigenvector approach. *Image Vision Comput.*, 10(5):283–288, 1992.
- [35] J. R. Smith and S. Chang. Visualeek: A fully automated content-based image query system. In *ACM Multimedia*, pages 87–98, 1996.
- [36] J. R. Smith and S. Chang. Image and video search engine for the world wide web. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 84–95, 1997.
- [37] A. M. R. Souza and C. A. Z. Barcelos. Sobre eliminação de ruídos e quantização no processo de restauração de imagens. *Revista Horizonte Científico - <http://www.propp.ufu.br/revistaelectronica/index.html>*, pages 1–16, 2002.
- [38] M. J. Swain. Interactive indexing into image databases. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 95–103, 1993.
- [39] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 1(7):11–32, 1991.
- [40] H. Tamura and N. Yokoya. Image database systems: A survey. *Pattern Recognition*, 17(1):29–43, 1984.
- [41] A. Vailaya. *Semantic Classification in Image Databases*. PhD thesis, Michigan State University, 2000.
- [42] D. R. Wilson and T. R. Martinez. Improved heterogeneous distance functions. *J. Artif. Intell. Res. (JAIR)*, 6:1–34, 1997.
- [43] D. Zhang. Improving image retrieval performance by using both color and texture features. In *Third International Conference on Image and Graphics*, pages 172–175, 2004.