Um modelo para recuperação por conteúdo
de Imagens de Sensoriamento Remoto

Luis Mariano del Val Cura

Tese de Doutorado

Instituto de Computação
Universidade Estadual de Campinas

Um modelo para recuperação por conteúdo de Imagens de Sensoriamento Remoto

Luis Mariano del Val Cura[1]

Novembro 2002

Banca Examinadora:

- Prof. Dr. Neucimar Jerônimo Leite (Orientador)
  Instituto de Computação - UNICAMP
- Prof. Dr. Jurandir Zello Junior
  Centro de Ensino e Pesquisas em Agricultura(CEPAGRI)
- Prof. Dr. Jansle Viera Rocha
  Faculdade de Engenharia Agrícola - UNICAMP
- Prof. Dr. Alexandre Xavier Falcão
  Instituto de Computação - UNICAMP
- Profa. Dra. Maria Beatriz Felgar de Toledo
  Instituto de Computação - UNICAMP
- Prof Dr. Geovanne Cayres Magalhães (Suplente)
  Instituto de Computação - UNICAMP
- Prof Dr. Paulo Lício de Geus (Suplente)
  Instituto de Computação - UNICAMP

[1]**Esta dissertação foi financiada parcialmente pela FAPESP (Processo 96/12380-5) e desenvolvida dentro do projeto do CNPq PRONEX-SAI (Sistemas Avançados de Informação)**

Um modelo para recuperação por conteúdo de Imagens de Sensoriamento Remoto

Este exemplar corresponde à redação final da Tese devidamente corrigida e defendida por Luis Mariano del Val Cura e aprovada pela Banca Examinadora.

Campinas, 3 de setembro de 2003.

Prof. Dr. Neucimar Jerônimo Leite
(Orientador)
Instituto de Computação - UNICAMP ()

Profa. Dra Claudia M. Bauzer Medeiros
(Co-Orientadora)
Instituto de Computação - UNICAMP

Tese apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação.

TERMO DE APROVAÇÃO

Tese defendida e aprovada em 20 de dezembro de 2002, pela Banca examinadora composta pelos Professores Doutores:

Prof. Dr. Jurandir Zullo Júnior
CEPAGRI - UNICAMP

Prof. Dr. Jansle Veira Rocha
FEAGRI - UNICAMP

Profa. Dra. Maria Beatriz Felgar de Toledo
IC – UNICAMP

Prof. Dr. Alexandre Xavier Falcão
IC – UNICAMP

Prof. Dr. Neucimar Jerônimo Leite
IC – UNICAMP

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP

Val Cura, Luis Mariano del

v23m Um modelo para recuperação por conteúdo de imagens de sensoriamento remoto/Luis Mariano del Val Cura - Campinas, [S.P.:s,n], 2002

Orientador: Neucimar Jerônomo Leite

Co-orientadora: Claudia Maria Bauzer Medeiros

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Computação.

1. Processamento de imagens. 2. Bancos de dados. 3. Sistemas de recuperação de informação. I. Leite, Neucimar Jerônimo. II. Medeiros, Claudia Maria Bauzer. III. Universidade Estadual de Campinas. Instituto de Computação. IV. Título.

A mi hija Mônica

Prefácio

O problema da recuperação de imagens por conteúdo tem sido uma área de muito interesse nos últimos anos, com múltiplas aplicações em diferentes domínios de geração de imagens. Uma classe de imagem onde este problema não tem sido resolvido satisfatoriamente refere-se à classe de Sensoriamento Remoto. Imagens de Sensoriamento Remoto (ISR) são obtidas como combinação do sensoramento da Terra em múltiplas bandas espectrais.

Esta tese aborda o problema da recuperação por em conteúdo das ISR . Este tipo de recuperação parte da caracterização do conteúdo de uma imagem e uma das suas principais abordagens considera modelos matemáticos da área de Processamento de Imagens a ser abodada nesta tese.

Neste trabalho, abordamos o processo de recuperação de ISR que utilizando três recursos principais: padrões de textura e cor como elemento básico da consulta, uso de múltiplos modelos matemáticos de representação e caracterização do conteúdo e um mecanismo de retroalimentação para o processo de consulta.

As principais contribuições da tese são: (1) uma análise dos problemas da recuperação por conteúdo para ISR; (2) a proposta de um modelo para esta recuperação; (3) um modelo e métrica de similaridade baseado no modelo proposto; (4) proposta de implementação do processamento das consultas que mostra a viabilidade do modelo.

Abstract

Content-based retrieval of images is a topic of growing interest given us multiple applications. One kind of images that have not yet been dealt with satisfactorily are the so-called Remote Sensing Images. Remote Sensing Images (RSI) are a especial type of image, created by combination of sensoring on different spectral bands .

This work deals with the problem of content-based retrieval of Remote Sensiong Images(RSI). It uses the image retrieval approach based on content representation models from image processing area.

This work presents a content-based image retrieval model for RSI, based on three main features: patterns of color and texture as basic query concept, use of multiple content representation models and a feedback relevance machanism.

The main contributions os these work are: (1) an analysis of content-based RSI problems; (2) a proposal of a model for RSI retrieval; (3) a proposal of model and metric for similarity measure; (4) a proposal of algorithm for processing of content-based queries

Agradecimentos

Aos meus orientadores, os professores Neucimar Jerônimo Leite e Claudia Bauzer Medeiros pelas discusões frutiferas, a dedicação na orientação e pela amizade.

A Marta pelo carinho, a paciência e a força constantes.

À minha família, pelo apoio e confiança desde o princípio.

Ao professor Jansle Viera Rocha, pela disposição permanente e pelas imagens utilizadas no trabalho.

Aos colegas dos grupos de bancos de daoos e processamento de imagens em todos estes anos que tanto tem ajudado no desenvolvimento da pesquisa. Em especial ao Daniel pelas magnificas discussões..

A minha familia cubana em Brasil

A todos os amigos que tem acompanhado e ajudado nestes anos no Brasil.

Aos professores e funcionários do Instituto de Computação, especialmente ao Ricardo Dahab pela amizade de sempre.

À FAPESP pelo apoio financeiro recebido que permitiu a realização deste projeto.

Ao Brasil, ao povo brasileiro, pela generosidade e braços sempre abertos.

Sumário

Lista de Figuras

Capítulo 1

Introdução

1.1 Visão Geral do problema

Atualmente, estima-se que 97% de todos os arquivos na *Web* contêm imagens. Com isto, a recuperação de imagens baseada em conteúdo é um tópico de crescente interesse. A possibilidade de recuperar imagens armazenadas em um Banco de Dados, a partir do seu conteúdo abre espaços a múltiplas aplicações. Este princípio de recuperação pode ser utilizado em sistemas de processamento de imagens médicas (tomográficas, ecográficas,etc) [57, 72], aplicações multimídia [27], Sistemas de Informações Geográficas[3, 5, 53], dentre outros.

O problema da recuperação de imagens por conteúdo apresenta vários desafios em diversas áreas, o que implica na necessidade de uma abordagem integradora. A área de Bancos de Dados aborda o armazenamento e a indexação das imagens, assim como a busca de linguagens que facilitem as consultas dos usuários e a expressão do conteúdo procurado. Pesquisas realizadas na área de Processamento de Imagens e Reconhecimento de Padrões visam definir modelos que permitam extrair a informação de padrões e objetos das imagens e a definição de descritores que representem o conteúdo. Associado a estes descritores, um outro desafio refere-se à busca de funções de similaridade que modelem o critério humano de percepção, em que participam também pesquisadores da área de Inteligência Artificial. Este processo inclui os especialistas dos diversos domínios de aplicação (biólogos, agrônomos, geólogos, médicos, etc) para a formalização das características das imagens que melhor descrevam seu conteúdo.

Existem três abordagens principais em Bancos de Dados, no que tange recuperação de imagens pelo conteúdo [28, 43]:

- *Baseada em atributos:* As imagens são descritas através de atributos alfanuméricos estruturados, que servem de base às consultas. Por exemplo, um campo com uma

cor predominante se a imagem contiver um quadro, o nome da pessoa se a imagem corresponde a um rosto. Esta abordagem utiliza mecanismos tradicionais de organização e indexação de dados (*B-trees*, listas invertidas, etc). Neste caso, são permitidas apenas consultas a partir dos atributos alfanuméricos associados às imagens

- *Baseada em textos ou "anotada":* O conteúdo das imagens é descrito através de documentos textuais. Uma consulta, neste caso, é processada com técnicas de recuperação de informação. A limitação desta abordagem está na ambigüidade e incompletitude da descrição de imagens através da linguagem natural (por exemplo, dificuldade em descrever texturas) e também nos diferentes vocabulários utilizados por diferentes usuários.

- *Baseada em características de conteúdo:* As imagens são processadas buscando extração e indexação de suas *características*, utilizando técnicas da área de Processamento de Imagens e Reconhecimento de Padrões. As características incluem cor, textura,forma, informações específicas aos objetos extraídos das imagens, etc. Esta abordagem permite resolver limitações das anteriores mas, em geral, não consegue extrair informações semânticas de alto nível sobre as imagens, sendo recomendável sua integração com as outras abordagens anteriores.

As duas primeiras abordagens descritas têm sido desenvolvidas fundamentalmente por pesquisadores das áreas de Bancos de Dados e Recuperação de Informação,e a última, por especialistas da área de Processamento de Imagens e Reconhecimento de Padrões [28]. Nesta tese, o enfoque fundamental concerne a busca de um modelo de recuperação por conteúdo que incremente a efetividade da abordagem baseada em características de conteúdo.

Quando as abordagens de recuperação por conteúdo consideram as características específicas do domínio das imagens processadas, aumenta-se a efetividade dos resultados obtidos. Esta especialização permite aplicar técnicas e mecanismos de consulta orientados ao domínio dessas imagens. Estas técnicas exploram, por exemplo, o tipo de características, a organização e disposição dos objetos relevantes ao seu conteúdo, os padrões de cor, forma e textura, dentre outros. Por exemplo, imagens de células sempre possuem um objeto central com bordas bem definidas, um núcleo celular dentro desse objeto e uma textura do corpo celular altamente relevante para os especialistas. No caso de rostos humanos as imagens apresentam, geralmente no seu centro, sempre aparece no centro, o objeto correspondente ao rosto, do qual é sempre possível extrair parâmetros de forma.

No caso específico de Imagens de Sensoriamento Remoto (ISR), não existem propostas de grande sucesso para recuperação de imagens por conteúdo que contemplem alguns dos

aspectos abordados neste trabalho. Estas imagens são obtidas a partir da detecção da emissão de ondas de diferentes freqüências pela superfície da Terra ou da resposta da superfície da Terra à emissão de ondas sobre ela.

Um dos grandes desafios da área de sensoriamento remoto, para os próximos anos, refere-se ao gerenciamento, armazenamento, processamento e recuperação de volumes importantes de informação de imagens. Vários projetos associados à observação e monitoramento terrestre visam um incremento significativo da geração de imagens digitalizadas. Um exemplo disto é o projeto NASA EOS (*Earth Observation System*) [51]. Este crescente volume de informação sugere a necessidade de criar mecanismos automatizados que permitam processamento e análise de ISR.

ISR são manipuladas dentro de Bancos de Dados Geográficos (BDG) que se caracterizam por gerenciar dados georeferenciados, isto é, relativos à superfície terrestre [50]. As funções de um BDG estão centradas no aspecto espacial (localização e geometria), mas oferecendo poucas facilidades para o gerenciamento efetivo das imagens. Na verdade, BDG tratam imagens como *bitmaps* sem semântica conhecida e cabe ao usuários tratar estas imagens à parte. Em geral, pouco tem sido feito em BDG para facilitar a recuperação das imagens a partir do seu conteúdo. Sistemas comerciais oferecem recursos na abordagem por atributos e certas facilidades para consultas, a partir dos valores dos pixels no domínio espacial [7]. O processamento adequado de ISR por um BDG deve considerar a introdução de facilidades de processamento característico de Bancos de Dados de Imagens, permitindo ao BDG uma recuperação por conteúdo.

Um sistema de recuperação por conteúdo de imagens, neste contexto, deve ser entendido como uma ferramenta auxiliar no processo de análise visual ou foto-interpretação realizado pelos especialistas em sensoriamento remoto, na procura de padrões ou comportamentos espaciais similares em diferentes imagens. Neste sentido, tal sistema pode contribuir à solução do processamento do grande volume de ISR.

A recuperação por conteúdo de ISR, que são freqüentemente descritas em termos de textura e cor, apresenta vários problemas ainda sem solução satisfatória.

Em primeiro lugar, ISR e em particular as imagens LANDSAT, são o resultado de um processo de composição artificial de bandas definida pelo usuário e que produz falsas cores, isto é, cores que não correspondem à visualização real da superfície da Terra. Estas falsas cores compõem objetos e padrões muito bem definidos que podem ser processados manual ou automaticamente. Diferentes especialistas podem criar composições diversas que geram distintas imagens associadas exatamente à mesma informação. Assim sendo, o que se vê é uma percepção particular de quem pré-processou a imagem e não seu conteúdo "real".

Em segundo lugar, uma propriedade fundamental das ISR é o fato de que seu conteúdo é sensível às características da sua localização geográfica, às condições meteorológicas do

momento da sua coleta, assim como à qualidade diferenciada dos dispositivos de sensoriamento. Este fato faz com que as características de conteúdo de textura e cor de um mesmo fenômeno ou objeto possam aparecer diferentes em imagens produzidas em locais, momentos e condições diversas.

Um dos problemas mais complexos na recuperação por conteúdo é a definição de um modelo de similaridade que simule a noção de similaridade humana. Em terceiro lugar, no caso das ISR, esta noção de similaridade depende de vários fatores como tipo de composição de cores utilizada, qualidade das imagens mas, sobretudo, da experiência do especialista. A capacidade de enxergar e discriminar objetos e padrões dentro de uma ISR é um processo gradual e demorado. A noção de similaridade entre padrões e textura de um especialista pode variar no tempo e durante o processo de foto-interpretação de uma imagem específica.

A redução do universo do tipo de imagem sendo processada agiliza os resultados do processo de recuperação. No caso das ISR esta redução do universo pode permitir o emprego de técnicas mais específicas a este tipo de imagens. No entanto, o conteúdo das ISR é altamente complexo, apresentando múltiplas combinações de padrões de textura e cor. Esta complexidade dificulta seu processamento automatizado e a busca de métodos efetivos de classificação, segmentação e interpretação. Assim, os métodos aplicados no processamento de uma imagem, em geral, são decididos a partir da experiência do especialista e do conteúdo apresentado nessa imagem em particular.

1.2 Abordagem da tese

As discussões apresentadas neste trabalho partem de três premissas básicas: (a) não existe um modelo matemático universal de descrição de conteúdo de ISR (da mesma forma que não existe modelo geral do senso de similaridade humano); (b) a ocorrência de um mesmo fenômeno em imagens de diferentes locais e tempos, ainda com as alterações esperadas, conserva similaridade em textura e cor que permite ao especialista relacioná-las; (c) as propriedades das ISR e do seu processamento fazem com que a similaridade entre seus padrões e objetos seja dependente de cada especialista e mutável no tempo.

Dadas estas premissas, a principal contribuição da tese consiste da apresentação de um modelo para recuperação por conteúdo. Este modelo é baseado em três conceitos:

- O uso de padrões como parâmetro de conteúdo para as consultas.
- A utilização simultânea de vários modelos de representação para descrever e recuperar o conteúdo de uma imagem.
- A possibilidade de retroalimentação (*feedback*) pelo usuário para refinar padrões e similaridades.

O **padrão** é o conceito fundamental para uma consulta. Um padrão é definido como um conjunto de amostras de diferentes imagens associado a uma mesma entidade ou fenômeno do mundo real. O objetivo de uma consulta então é recuperar as imagens que contenham regiões similares a esta definição de padrão.

O segundo elemento da proposta é o uso de **múltiplos modelos de representação do conteúdo**. Um modelo de representação é definido como um modelo matemático que permite descrever uma determinada característica de conteúdo e sobre o qual se define algum critério de similaridade. A proposta da tese considera o uso simultâneo de vários modelos, de maneira tal que o resultado final do sistema considere a contribuição colaborativa de todos eles. O objetivo desta abordagem é tentar explorar as capacidades de caracterização de textura e cor de diferentes modelos, efetivos em determinados padrões, mas que por outro lado não conseguem ser universais.

O terceiro elemento apresentado pelo modelo é um **mecanismo de retroalimentação**. No modelo proposto, uma consulta é entendida como um processo iterativo de refinamento da resposta. A cada iteração, o usuário avalia a resposta que o sistema oferece para sua consulta (baseada em padrões). O mecanismo de retroalimentação proposto permite ao usuário refinar sua definição sobre os padrões, refinando também a noção de similaridade. Este mecanismo deve conduzir a um processo de convergência entre o especialista e o sistema na busca de um resultado adequado da consulta.

1.3 Contribuições e organização do texto

Considerando os problemas apresentados anteriormente e a abordagem seguida, as principais contribuições desta tese são:

- Estudo dos problemas e particularidades das ISR visando sua recuperação por conteúdo.
- Definição de um modelo de recuperação por conteúdo para ISR que considere estas características, assim como as particularidades do processamento destas imagens pelos especialistas.
- Definição de um modelo e métrica de similaridade baseadas no uso de múltiplas representações de conteúdo, cujos resultados são refinados iterativamente a partir da retroalimentação (*feedback*) de cada especialista.
- Uma proposta de implementação do modelo que mostre a viabilidade de sua utilização.

O texto está organizado da seguinte forma:

- O capítulo 2 apresenta a base teórica necessária ao entendimento do texto - modelos, descritores de conteúdo e propostas de sistemas.
- O capítulo 3 analisa problemas relativos à descrição de ISR e a busca por conteúdo deste tipo de imagens.
- O capítulo 4 descreve a solução dos problemas levantados no capítulo 3 apresentando o modelo proposto de recuperação por conteúdo.
- O capítulo 5 apresenta uma arquitetura de implementação do modelo, apresentado no capítulo 4, e inclui diferentes algoritmos de processamento de uma consulta baseado em múltiplos modelos, incluindo sua otimização.
- O capítulo 6 descreve brevemente uma implementação de parte da proposta do capítulo 5.
- O capítulo 7 contém conclusões e extensões.

Capítulo 2

A recuperação de imagens baseada em conteúdo

2.1 Introdução

A introdução de facilidades em um SGBD para recuperação de imagens por conteúdo introduz mudanças e impactos nos seus diferentes módulos. Existem duas etapas fundamentais em um Sistema de Bancos de Dados de Imagens que o tornam diferente de um SGBD padrão: a inserção de imagens e a realização das consultas. Este processo é ilustrado na figura 2.1.

No momento da inserção de cada imagem, diferentes processamentos devem ser realizados para extrair as características de conteúdo consideradas relevantes. Estas características são representações compactas, geralmente vetores de valores, que são utilizados para sua indexação. As estruturas de indexação visam em geral organizar as representações de forma a suportar facilmente o cálculo da distância ou similaridade entre estes vetores.

A consulta consiste em estabelecer uma descrição, a partir de diferentes recursos oferecidos, das propriedades desejadas na imagem alvo. Em seguida, é realizado um processamento dessa descrição, que pode incluir técnicas de processamento de imagens, gerando uma representação do tipo de imagem procurada. Esta representação é utilizada na etapa de processamento da consulta que utiliza as estruturas de índices já criadas.

Neste processo, os principais problemas a serem considerados, discutidos nas próximas seções, são:

- Definição de características de conteúdo relevantes e estruturas de representação (seção 2.2)

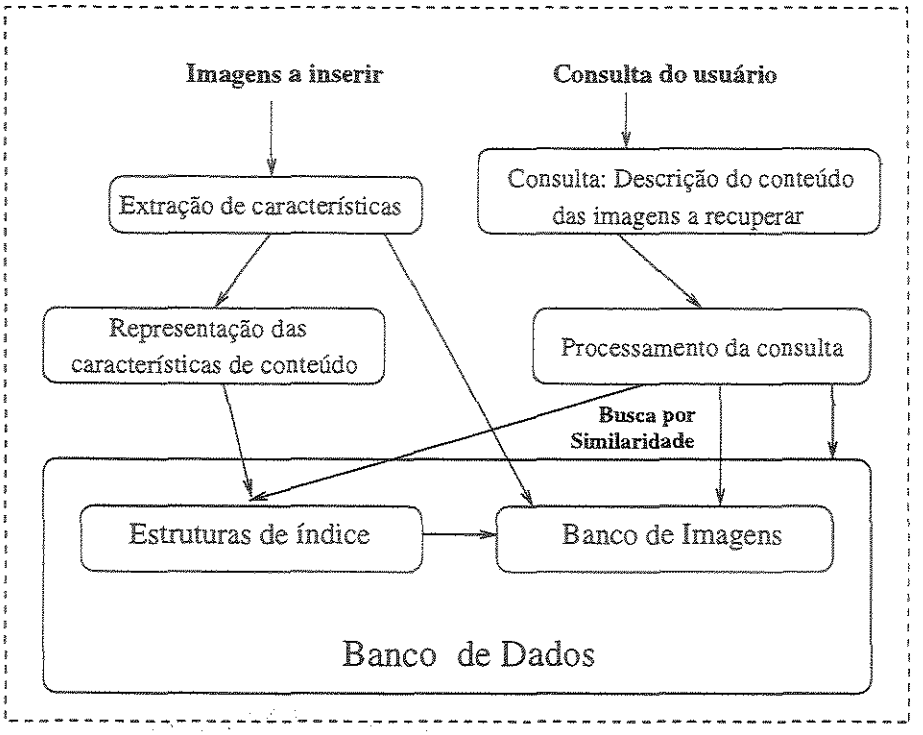


Figura 2.1: Processo de recuperação por conteúdo em um Banco de Dados de Imagens.

- Definição de critérios de similaridade entre imagens (seção 2.3)
- Modelos de consulta com suporte para características de conteúdo (seção 2.4)
- Estruturas de índice para características de conteúdo (seção 2.5)

2.2 Definição de características de conteúdo e estruturas de representação

O primeiro problema em um sistema de busca por conteúdo é definir quais as características de conteúdo relevantes para a aplicação ou para o tipo de imagem que se deseja recuperar. Em geral, quanto mais abrangentes forem consideradas as imagens e seu conteúdo, menos precisos são os resultados obtidos. Diferentes tipos de imagens, assim como diferentes aplicações, determinam as características do conteúdo relevantes na interpretação.

Cada "família"de características de conteúdo está associada a um grupo de algoritmos de processamento de imagens para extrair valores que simplifiquem e descrevam essa característica de conteúdo. Estes **descritores**, em geral, são representados por veto-

res definidos em um espaço multidimensional denominado **Espaço de Características**. Muitos dos algoritmos de processamento baseiam-se em modelos matemáticos não exatos das características de conteúdo (por exemplo, textura). Ao mesmo tempo, em uma consulta, a descrição da característica procurada é também imprecisa. Desta forma, o mecanismo tradicional de casamento exato não pode ser aplicado, o que exige a definição de medidas de similaridade nestes espaços.

Algoritmos de processamento ou de extração de um descritor devem ser associados a **critérios de similaridade**, que são em geral dependendes de métricas de distância. Estes algoritmos não devem ser computacionalmente caros pois serão utilizados na etapa de consulta, no processo de extração (idealmente em tempo real) da informação relativa à descrição do conteúdo das imagens procuradas.

Um descritor associado a uma característica é **global** quando descreve toda a imagem e **local** quando considera a descrição da característica em uma região ou objetos da imagem. Neste último caso, múltiplos descritores são associados a uma mesma imagem descrevendo as características dos seus objetos ou regiões. Para a descrição local é necessário aplicar técnicas de segmentação e/ou classificação que identifiquem as diferentes regiões de interesse.

As seções 2.6 e 2.7 apresentam algumas técnicas para a análise e caracterização de imagens associadas à textura e cor.

2.3 Definição de critérios de similaridade

Como mencionado anteriormente, a relação entre descritores de uma mesma característica de conteúdo não pode se basear em um casamento exato de similaridade e, portanto, é preciso introduzir medidas de similaridade alternativas. Um dos desafios, neste sentido, refere-se à busca de descritores de conteúdo que modelem a percepção humana de similaridade. De fato, o sucesso de uma consulta baseia-se em aspectos perceptuais e às vezes subjetivos sobre o que é relevante como resposta.

Seguindo estas idéias, propostas de modelos matemáticos para medidas de similaridade têm sido apresentadas, algumas das quais surgiram de estudos psicológicos. Uma taxonomia das medidas de similaridade em [35] é apresentada a seguir.

2.3.1 Medidas baseadas em métricas

A maioria das abordagens modelam a similaridade a partir de métricas. embora muitas vezes sem justificar sua correspondência com um modelo perceptual e considerando apenas critérios estatísticos. Exemplos clássicos das medidas baseadas em métricas são as r-métricas de Minkowski que têm como forma geral:

$$L_r(x,y) = \left[\sum_{i=1}^{n} |x_i - y_i|^r\right]^{1/r}, r \geq 1 \tag{2.1}$$

$$L_\infty(x,y) = max_i|x_i - y_i| \tag{2.2}$$

sendo x, y pontos em um Espaço de Característica n-dimensional. Para $r = 1$ e $r = 2$ temos as métricas *city-block* e euclidiana, respectivamente. Estas métricas satisfazem um conjunto de axiomas bem conhecidos:

$$\begin{array}{ll} \text{Constância ou Auto-similaridade:} & d(S_a, S_a) = d(S_b, S_b) \\ \text{Minimalidade:} & d(S_a, S_a) \leq d(S_a, S_b) \\ \text{Simetria:} & d(S_a, S_b) = d(S_b, S_a) \\ \text{Desigualdade triangular:} & d(S_a, S_b) + d(S_b, S_c) \geq d(S_a, S_c) \end{array} \tag{2.3}$$

sendo S_a, S_b, neste caso, vetores de características de duas imagens.

Existem características onde medidas baseadas em métricas satisfazem o conceito humano de similaridade. Por exemplo, a métrica euclidiana aplicada sobre o sistema de cores HSI modela adequadamente a percepção humana de similaridade em relação às cores [25]. Experimentos em [87] validam o uso de uma métrica para comparar descritores de textura associados com rugosidade, contraste e direcionalidade.

O uso de medidas baseadas em métricas para medir similaridade é muito útil, em particular pela geometria que gera, facilitando o processo de busca dentro do Espaço de Característica. Arkin argumenta em [4] que uma medida de similaridade deve ser uma métrica para garantir buscas eficientes, a partir do axioma da desigualdade triangular. Sabendo que a distância entre A e B é grande e que a distância entre B e C é pequena, é possível inferir que a distância entre A e C é grande sem precisar calculá-la. Igualmente, se é sabido que A está muito próximo a B e B muito próximo a C é possível inferir que A está próximo a C. Estes axiomas são essenciais em algumas estruturas de indexação como o caso do *M-tree* [13].

Em [19] Fagin concorda com o argumento, mas afirma que é possível relaxar a desigualdade triangular, e simultaneamente inferir as distâncias entre diferentes pontos, de forma similar a como antes descrito nos casos de A, B e C. No seu trabalho é proposta uma função para medir similaridade, a partir da forma de objetos, que não satisfaz exatamente a desigualdade triangular, mas satisfaz uma expressão deste axioma de forma relaxada:

$$k(d(S_a, S_b) + d(S_b, S_c)) \geq d(S_a, S_c) \tag{2.4}$$

sendo k uma constante em geral pequena, e portanto permitindo fazer inferências sobre as distâncias entre pontos.

2.3.2 Medidas baseadas em teoria de conjuntos

Enquanto vários autores julgam medidas métricas adequadas, outros trabalhos questionam a validade dos diferentes axiomas das métricas do ponto de vista perceptual [68, 71, 88, 89]. Experimentos com diferentes estímulos e padrões mostram comportamentos do ser humano que variam em relação à desigualdade triangular [89] e ainda em relação à propriedade de simetria [88]. Nesta direção, Tversky [89] propõe um **Modelo de contraste** com uma medida de similaridade baseada em teoria de conjuntos. Esta medida considera a existência a priori de um conjunto de possíveis características que um conceito ou objeto (neste caso imagens) podem ter e que servem para a sua descrição. A partir destas características, se define como similaridade entre duas imagens uma combinação linear de medidas sobre as características que são comuns e características que estão ou não presentes em cada uma delas. Se consideramos A e B como o conjunto de características presentes nas imagens a e b, respectivamente, a similaridade entre elas é expressa por:

$$S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A),\ para\ \alpha, \beta, \theta \geq 0 \tag{2.5}$$

onde f corresponde a uma função que associa valores à presença de características. Os parâmetros α, β e θ permitem associar maior relevância às características da imagem A, da imagem B ou as que estão presentes nas duas respectivamente. Um maior valor para α implica em atribuir uma maior relevância às características da imagem A da imagem A, o mesmo ocorrendo para β e a imagem B.

Note-se que os parâmetros α, β e θ podem ser ajustados e valores diferentes para eles podem gerar medidas que não satisfazem os axiomas métricos.

Este modelo é limitado pois exige que as características de conteúdo sejam representadas por predicados booleanos, o que não é comum em descrições de propriedades em imagens, onde em geral a descrição é realizada com valores numéricos. Como solução, Santini [68] propõe uma extensão apresentando um **Modelo de contraste fuzzy**. Este modelo parte do pressuposto de que existem n medições na imagem

$$s = \{s^1, \cdots, s^n\} \tag{2.6}$$

a partir das quais pode ser construído um vetor *fuzzy* de m predicados:

$$\mu(s) = \{\mu^1(s), \cdots, \mu^m(s)\} \tag{2.7}$$

que caracterizem o conteúdo. Introduzindo os operadores lógicos *fuzzy* de interseção e diferença:

$$\mu_\cap(s_1, s_2) = \{min\{\mu^1(s_1), \mu^1(s_2)\}, \cdots, min\{\mu^m(s_1), \mu^m(s_2)\}\}, \tag{2.8}$$

$$\mu_{-}(s_1, s_2) = \{max\{\mu^1(s_1) - \mu^1(s_2), 0\}, \cdots, max\{\mu^m(s_1) - \mu^m(s_2), 0\}\}, \tag{2.9}$$

a expressão de similaridade na equação (2.5) pode ser escrita como:

$$\begin{aligned} S(s_1, s_2) = \; & \theta \sum_{\lambda=1}^{m} min\{\mu^\lambda(s_1), \mu^\lambda(s_2)\} - \\ & \alpha \sum_{\lambda=1}^{m} max\{\mu^\lambda(s_1) - \mu^\lambda(s_2), 0\} - \\ & \beta \sum_{\lambda=1}^{m} max\{\mu^\lambda(s_2) - \mu^\lambda(s_1), 0\} \end{aligned} \tag{2.10}$$

sendo $S(s_1, s_2)$ a medida de similaridade.

Embora mais flexível, esta nova formulação do modelo requer a transformação de descritores de conteúdo numéricos em predicados *fuzzy*, o que em geral não é simples. No entanto, corresponde a uma abordagem de grande interesse quando se deseja modelar similaridade fugindo das restrições dos axiomas métricos. O problema principal deste modelo está em definir e implementar índices espaciais adequados. A ausência dos axiomas métricos limita a criação de estruturas de índice espacial e por tanto compromete a eficiência da busca por conteúdo.

2.4 Modelos de consulta com suporte para características de conteúdo

Modelos de consulta baseados em características de conteúdo devem fornecer recursos que permitam especificar tais características. Como a maioria destas não pode ser definida de maneira precisa, estes recursos devem possibilitar a descrição aproximada dessas características nas imagens a serem procuradas.

Os modelos de consulta podem ser analisados sob diferentes perspectivas: modelos de dados, linguagem de consulta e formas de interação com o resultado.

Do ponto de vista do **modelo de dados**, alguns sistemas consideram uma abordagem de **recuperação de dados**, estendendo um modelo de banco de dados para incluir imagens como novos tipos [53, 1]. Os modelos orientados a objetos ou relacional-estendido permitem esta extensão de maneira natural, incluindo a definição de novos operadores sobre tais tipos. Estes operadores podem, inclusive, ser de similaridade e seu uso nas linguagens de consulta é imediato. Outros trabalhos usam a abordagem de **recuperação de informação**, onde não existe um modelo de banco de dados estruturado, e consideram um repositório de imagens com anotações textuais associadas sobre o qual se define algum mecanismo para expressar consultas.

Do ponto de vista da **linguagem de consulta**, existem propostas em que **consultas textuais** podem ser realizadas com linguagens textuais que permitem descrever as

imagens procuradas usando recursos lingüísticos. Por exemplo, em [53], para se recuperar imagens com determinados padrões de cores, é utilizado um conjunto de descrições imprecisas predefinidas ou metadados (por exemplo: (*MostlyRed, SomeOrange, WhiteStuff*), ou expressões predefinidas, denominadas conceitos, que envolvem combinações destas descrições e de atributos alfanuméricos (por exemplo: *SunSet* pode ser definido como *MostlyRed* OR *MostlyYellow* OR *MostlyPurple*. Em [58] a linguagem de consulta textual permite descrever as relações topológicas e a posição relativa de objetos complexos procurados nas imagens.

A abordagem textual é limitada e, em geral, os sistemas introduzem **consultas gráficas** com recursos para a descrição do conteúdo da imagem procurada. A forma mais simples desta descrição corresponde a apresentar ao sistema uma imagem como modelo daquelas que se deseja recuperar. Esta abordagem é chamada de *Query by pictorial example* pela similaridade com o modelo de consulta do mesmo nome em bancos de dados convencionais. Outras abordagens permitem a descrição em separado de diferentes características. Por exemplo, em [21] é possível desenhar formas de objetos procurados nas imagens em um ambiente gráfico e selecionar as cores procuradas de forma visual em outro ambiente, ambos formando parte da mesma consulta.

Quando as linguagens oferecem recursos para considerar simultaneamente diferentes aspectos do conteúdo, dizemos que permite **consultas compostas**. Estas consultas podem ser traduzidas em predicados, em que são combinados múltiplos termos com operadores lógicos de primeira ordem (and, or, etc). Cada termo expressa uma condição sobre o conteúdo utilizando, em geral, as medidas de similaridade. Por exemplo, o predicado $SimilarColor(A, B) \geq 0,7$ and $SimilarTexture(A, B) \geq 0,8$ corresponde a uma consulta composta que envolve simultaneamente similaridades por cor e textura.

Outro ponto de vista considera as **formas e interação com o resultado**. Quando os predicados associados às consultas são booleanos, temos casamento exato ou *matching*. Ao ser executada uma consulta por este critério, o predicado booleano é avaliado em todas as imagens. O resultado da consulta corresponde àquelas imagens onde é obtido um resultado verdadeiro para o predicado. Desta forma, uma consulta divide as imagens, no Banco de Dados, em dois subconjuntos: as que não casam com o predicado e as que casam, que são devolvidas como resultado. No caso de consultas compostas, nesta abordagem é necessário decidir os patamares de similaridade aceitáveis em cada termo (por exemplo, $0,7$ e $0,8$ no predicado $SimilarColor(A, B) \geq 0,7$ and $SimilarTexture(A, B) \geq 0,8$. Santini [67] chama esta abordagem de **busca exata** (*matching search*)

A **busca por similaridade** (*similarity search*) é apresentada em contraposição à abordagem anterior. Aqui a resposta a uma consulta é formalmente "todo"o banco de dados, isto é, não existe uma diferenciação, como no caso anterior, entre imagens que satisfazem ou não a consulta. Para cada imagem é calculado um valor que define sua similaridade com

a especificação da consulta. Como resultado, a resposta a uma consulta é todo o banco de dados, ordenado segundo um critério de similaridade adotado. Na prática, o resultado é reduzido a uma quantidade de imagens ou a um patamar de similaridade. Como forma de implementar este modelo, vários autores [18, 54, 67] propõem lógica *fuzzy*. Neste caso, é possível construir consultas compostas com predicados e operadores lógicos (and, not, or) *fuzzy* que retornam valores de similaridade no intervalo $[0, 1]$. Este valor final corresponde a uma avaliação global da similaridade, sobre a qual é estabelecida uma ordenação ou *ranking*. Por exemplo, o predicado *fuzzy* $SimilarColor(A, B)$ and $SimilarTexture(A, B)$ ao ser avaliado gera um valor de similaridade a partir da combinação das medidas de similaridade e operadores lógicos *fuzzy*.

Em uma abordagem ao mesmo problema, Ortega [54] considera um modelo probabilístico, no qual o resultado de similaridade obtido representa uma estimativa da probabilidade de que a imagem corresponda à consulta especificada pelo usuário.

A abordagem de busca por similaridade apresenta várias modalidades, como em [54, 20, 17], que incluem pesos nas diferentes condições dos predicados *fuzzy*, isto é, pesos associados às similaridades pelas diferentes características de conteúdo. Estes pesos permitem ao usuário manipular a relevância das diferentes similaridades no cálculo da similaridade final.

Embora a introdução de pesos aumente a flexibilidade das consultas, em [69, 65] se argumenta que o usuário não deve ser responsável pela definição destes pesos, que nem sempre têm uma interpretação semântica clara. Estes autores propõem analisar uma consulta como um processo de aproximação sucessiva ao resultado. Santini [69] faz uma distinção entre uma consulta em que o usuário sabe o que está procurando, embora a descrição do seu modelo de consulta não seja exata, e a consulta em que o usuário não sabe exatamente o que procura, e a sua própria medida de similaridade vai evoluindo. Adicionalmente, como a maioria das características de conteúdo não possuem ainda informação semântica de alto nível, é usual que, nos sistemas e propostas existentes, os resultados de consultas não correspondam às expectativas do usuário, mesmo com mecanismos de similaridade precisos.

Nestas novas abordagens, chamadas de **exploratória** [69], **human in the loop** [64] ou de **feedback** [65], uma recuperação é vista como um processo de interação com os resultados de sucessivas consultas, de tal forma que as avaliações do usuário sobre a qualidade do resultado em um passo permitam ao sistema redefinir e refinar uma próxima consulta, e assim sucessivamente, até alcançar um resultado aceitável pelo usuário.

2.5 Indexação dos elementos de conteúdo

Dado um Espaço de Característica e uma medida de similaridade, um ponto neste espaço representa o conteúdo de uma imagem ou de um objeto dentro da imagem. Buscar aquelas imagens ou objetos com maior similaridade é equivalente a procurar os pontos representados no espaço, cuja distância seja a menor possível, sendo a distância calculada a partir da medida de similaridade usada. Quando esta medida é uma métrica, a busca pode ser facilitada com a introdução de estruturas de indexação espacial existentes na área de Bancos de Dados Espaciais. Estas estruturas permitem uma solução eficiente nas buscas dos pontos mais próximos ou os pontos dentro de um patamar de distância.

As mais populares destas estruturas são a família das *R-tree* [29, 75, 8, 37], as quais estendem ao espaço as idéias de balanceamento das tradicionais *B-trees*. Estas árvores em geral utilizam a distância euclidiana. A árvore M [13], por sua vez, permite a indexação com uma medida de distância qualquer que satisfaça os axiomas métricos.

Estas estruturas têm sido utilizadas para indexar objetos em espaços 2-3 dimensionais. No entanto, quando o número de dimensões aumenta - como no caso de Espaços de Características - o desempenho destas estruturas começa a diminuir consideravelmente, podendo em alguns casos, ser inferior às buscas seqüenciais. Uma forma de aliviar este problema é diminuir a dimensionalidade dos espaços de características, o que nem sempre é possível. Novas estruturas estão sendo propostas [23] para espaços multidimensionais que prometem facilitar o acesso eficiente em Bancos de Dados de Imagens.

Pela importância da textura e cor na análise e reconhecimento em Imagens de Sensoriamento Remoto serão apresentadas, a seguir, algumas técnicas para seu processamento e caracterização.

2.6 Textura em Recuperação por Conteúdo

Não existe uma definição exata para descrever textura. Várias definições informais têm sido apresentadas baseadas em alguns modelos ou hipóteses sobre o conceito. Haralick [31, 30] define textura como uma propriedade associada à distribuição estatística espacial de tons de cinza. Por sua vez Huet et. al. [32] considera textura como a organização dos pixels com variações de alta freqüência que apresentam caráter pseudoperiódico. Para Jain et. al. [33], a textura é caracterizada pela invariância de certas medidas locais em sub-regiões da imagem.

Embora não exista um modelo definitivo, alguns fatores podem ser destacados: (1) seu caráter local, isto é, a caracterização de cada pixel em relação à textura depende das informações em uma vizinhança e (2) a importância da escala ou resolução que influencia a percepção sobre uma textura.

Existem diferentes modelos para o processamento de texturas. Em [14] Claussi apresenta uma taxonomia dividindo os métodos em:

- *Métodos estruturais*: Consideram que a textura possui duas componentes básicas: uma primitiva de textura (*textons*) e uma organização espacial dessas primitivas [36]. A primitiva corresponde a um padrão fixo que é movimentado com variações estatísticas compondo as regiões texturizadas.

- *Métodos estatísticos*: Consideram uma caracterização probabilística da textura de uma região, tentando caracterizar a textura por propriedades e magnitudes estatísticas e seu relacionamento com elementos perceptuais tais como suavidade, direcionalidade, granulosidade, etc. Estes métodos combinam análise espectral e espacial.

- *Métodos baseados em modelos*: Tentam caracterizar a textura a partir de uma função analítica, de tal forma que os parâmetros associados a estas funções descrevam a textura. As abordagens típicas são os Campos Markovianos e modelos fractais.

Existem duas formas de incorporar textura como característica de conteúdo em recuperação de imagens: caracterizando toda a imagem com um descritor global ou caracterizando regiões homogêneas com descritores locais. Ambas as técnicas estão muito relacionadas com a análise para classificação e segmentação de imagens por textura.

Um critério que consegue discriminar regiões de uma imagem por textura poderia ser considerado como uma medida de similaridade entre texturas diferentes. No entanto, no caso da classificação ou segmentação, estas operações em geral consideram apenas a informação contida na imagem sendo processada. Neste caso, a população a classificar ou segmentar é limitada aos pixels da imagem. Como no caso de recuperação por conteúdo a função de similaridade tem que considerar todo o possível Espaço de Características, nem sempre um bom critério de segmentação representa um critério robusto para modelar similaridade.

Os descritores locais de uma imagem para recuperação por Conteúdo requerem uma operação de segmentação das imagens. Neste caso, consideramos que a segmentação tem as seguintes particularidades:

- *Não precisa ser exata:* O objetivo da segmentação é identificar que uma determinada textura está presente na imagem. Estabelecer exatamente os limites exatos da região de textura homogênea, embora desejável, não é imprescindível como em outras aplicações.

- *Pode ser realizada somente em regiões de interesse:* O processo de segmentação não precisa cobrir toda a imagem. Regiões não homogêneas em textura ou com área muito pequena não precisam ser consideradas.

Existem inúmeros trabalhos que abordam o problema de caracterização e segmentação de texturas. Muitos dos modelos baseiam-se na análise em uma vizinhança de cada pixel como matrizes de co-ocorrência e campos randômicos de Markov. Outras abordagens recentes como a transformada Wavelet, funções de Gabor e granulometrias morfológicas incluem uma análise multiresolução. A seguir são apresentadas algumas das abordagens mais importantes relativas à textura.

2.6.1 Tamura et. al.

Tamura et. al [87] desenvolveram experimentos psicológicos sobre um conjunto de características para textura: rugosidade, contraste, direcionalidade, semelhança com linhas, regularidade e aspereza. As três primeiras são consideradas as mais significativas do ponto de vista perceptual. Para estas características são propostas medidas baseadas em relacionamentos entre os níveis de cinza na vizinhança dos pixels.

Estas medidas são as seguintes:

- *Rugosidade:* Inicialmente, para cada pixel (x, y) é calculado o valor médio $A_k(x, y)$ dos níveis de cinza em uma vizinhança 2^k. Posteriormente, para cada pixel é calculada a diferença entre estas médias em vizinhanças sem sobreposição em direções opostas. Para a direção horizontal a diferença seria:

$$E_k(x, y) = |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)| \tag{2.11}$$

  A seguir é calculado para cada pixel qual o tamanho da vizinhança que maximiza o valor de E, isto é:

$$S_{best}(x, y) = 2^k \text{ tal que } E_k = max(E_1, E_2, \cdots, E_L) \tag{2.12}$$

  Finalmente, o descritor de rugosidade da imagem é definido como o valor médio em toda a imagem dos valores S_{best}.

$$F_{Rug} = \frac{1}{mn} \sum_{i}^{m} \sum_{j}^{n} S_{best}(i, j) \tag{2.13}$$

- *Contraste:* O contraste pode ser entendido como uma propriedade da textura. Seu valor é calculado por:

$$F_{con} = \frac{\sigma}{{\alpha_4}^n} \tag{2.14}$$

  onde $\alpha_4 = \frac{\mu_4}{\sigma_4}$ é a curtose, sendo σ e μ o desvio quadrático médio e a média do histograma da imagem, respectivamente.

- *Direcionalidade:* A idéia deste descritor é construir um histograma de probabilidade das bordas locais em relação à orientação. É usado o fato de que o gradiente possui uma magnitude e uma direção. Para cada pixel é calculada a diferença entre os valores de dois filtros com orientação vertical e horizontal e sua direção, isto é:

$$\begin{aligned} |\Delta G| &= (|\Delta H| + |\Delta V|)/2 \\ \theta &= tan^{-1}(\Delta V/\Delta H) + \pi/2 \end{aligned} \tag{2.15}$$

sendo ΔH e ΔV o resultado da aplicação dos filtros orientados.

A seguir é construído um histograma sobre os valores de θ, considerando significativos aqueles valores de magnitude $|\Delta G|$ acima de um limiar t :

$$H_D(k) = N_\theta(k) / \sum_{i=0}^{n-1} N_\theta(i) \tag{2.16}$$

sendo N_θ o número de pixels com $(2k-1)\pi/2\pi \leq \theta \leq (2k+1)\pi/2\pi$ e tais que $|\Delta G| \geq t$

A direcionalidade da textura na imagem é calculada a partir das alturas dos picos do histograma construído.

2.6.2 Matrizes de co-ocorrência

A matriz de co-ocorrência é a abordagem mais tradicional para o processamento de textura e foi introduzida por Haralick[31, 30]. Este método baseia-se na caracterização de uma região com textura homogênea, a partir de dados estatísticos associados à ocorrência simultânea de níveis de cinza dos pixels na região.

A matriz contém a probabilidade conjunta de aparição de todas as combinações de níveis de cinza, dois a dois, dados dois parâmetros: uma distância entre os pixels (δ) e uma orientação entre pixels (θ). Mais exatamente, é definida uma matriz de probabilidade [14]:

$$Pr(X) = \{C_{ij}(\delta, \theta)\} \tag{2.17}$$

onde $C_{ij}(\delta, \theta)$, é a **matriz de co-ocorrência** definida como:

$$C_{ij}(\delta, \theta) = \frac{P_{ij}(\delta, \theta)}{\sum_{i,j=1}^{G} P_{ij}(\delta, \theta)} \tag{2.18}$$

em que $P_{ij}(\delta, \theta)$ representa o número de ocorrências dois níveis de cinza g_i e g_j entre pixels a uma distância (δ) e orientação (θ). O denominador representa o total de ocorrências de níveis de cinza, sendo G o número de níveis de cinza diferentes. Várias matrizes podem ser

calculadas para considerar múltiplas distâncias e orientações entre pixels. Dado o custo resultante do cálculo para quaisquer pares de níveis de cinza, geralmente é realizada uma quantização destes níveis, diminuindo o tamanho da matriz.

Haralick [31, 30] propõe um grupo de 14 estatísticas obtidas da matriz para caracterizar a textura. Gotlieb et. al [26] estudaram experimentalmente estas estatísticas concluindo que as de maior capacidade de discriminação para textura são as seguintes:

- Contraste: $f_2 = \sum_{n=0}^{G-1} n^2 \{\sum_{|i-j|=n}^{G-1} C_{ij}\}$
- Momento de diferença inversa: $f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} C_{ij}$
- Entropia: $f_9 = -\sum_i \sum_j C_{ij} \log(C_{ij})$

A matriz de co-ocorrência tem apresentado boa capacidade de discriminação de textura. A principal desvantagem é o seu alto custo computacional.

2.6.3 Campos Randômicos de Markov

Nesta abordagem, uma imagem é modelada como um vetor aleatório $A = \{A_s, s \in S\}$ sendo S o conjunto de pixels da imagem. Um vetor $a = \{a_s, s \in S\}$ é chamado de uma *configuração* de A. Seja $P(a) = P(A_s = a_s, s \in S)$ a probabilidade da configuração a e seja $V = (V_s, s \in S)$ um sistema de vizinhos para o pixel s, tal que $s \notin V_s$ e $s \in V_t \Leftrightarrow t \in V_s$.

Um campo aleatório A é um *Campo Markoviano* se obedece à propriedade: $p(a) > 0$ e $p(a_s|a_r, r \in S - \{s\}) = p(a_s|a_r, r \in V_s)$, isto é, a probabilidade da ocorrência dos valores em cada pixel depende apenas dos valores dos seus vizinhos. Pode-se provar que um campo aleatório é um campo markoviano se e somente se $p(A = a)$ é uma medida de *Gibbs* definida por $P(a_s|V_s) = \frac{1}{Z} e^{-H(a_s, V_s)}$, sendo $H(a_s, V_s)$ uma função de energia e Z uma função de partição com o objetivo de normalização da distribuição. A função de energia H descreve a contribuição dos pixels vizinhos em cada pixel.

Modelos de textura podem ser definidos considerando diferentes funções de energia. Estas funções, em geral, possuem parâmetros Θ que descrevem a forma com que os vizinhos contribuem para o pixel. Considerando a priori uma função de energia, a estimação dos seus parâmetros para cada pixel é o problema fundamental a resolver. Estes parâmetros estimados podem servir como descritor das características do pixel. Em [73] é utilizada uma função de energia binomial enquanto os autores em [55] utiliza uma função gaussiana.

A abordagem com Campos Markovianos tem apresentado bons resultados, embora possua problemas relacionados a mudanças de escala e ao alto custo computacional.

Outra abordagem com Campos Randômicos foi proposta em [42]. Aqui é utilizada a teoria de decomposição de Campos aleatórios de *Wold*. Segundo esta teoria, um campo

aleatório y pode ser representado de forma única pela decomposição de três componentes ortogonais:

$$y(m,n) = w(m,n) + p(m,n) + g(m,n) \tag{2.19}$$

em que p corresponde a um campo harmônico, g corresponde a um campo evanescente e w corresponde a um campo não determinístico.

Um estudo experimental em [61] descreve que o ser humano agrupa padrões segundo os critérios de repetitividade, direcionalidade e complexidade do padrão. Os autores associam o campo p à componente de repetitividade, o campo g à direcionalidade e o campo não determinístico w à complexidade. Em um primeiro passo, é calculada a periodicidade da componente harmônica. Se considerada periódica, então é usada como critério de comparação. Caso contrário, então são avaliadas as outras componentes.

2.6.4 Transformada Wavelet

Define-se *Wavelet* como uma função $\psi(t)$ tal que:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{2.20}$$

ou seja ondas bem localizadas. Estas funções podem ser dilatadas segundo um parâmetro σ e transladadas pelo parâmetro u, formando a chamada "Mother Wavelet":

$$\psi_{u,\sigma} = \frac{1}{\sqrt{\sigma}}\psi(\frac{t-u}{\sigma}) \tag{2.21}$$

a partir da qual se define a Transformada:

$$Wf(u,\sigma) = \int_{-\infty}^{+\infty} f(t)\psi_{u,\sigma}(t)dt \tag{2.22}$$

O parâmetro u possibilita a localização espacial e a escala σ realiza a expansão da função centrada em u. Quando a Transformada de Fourier de $\psi_{u,\sigma}$ é calculada, também existe um centro de freqüência dependente de σ e com intervalo de freqüência inversamente proporcional a σ. A flexibilidade introduzida por *Wavelets* está em que a variação do parâmetro σ permite variar tanto a localização do centro de freqüência como do intervalo dessa freqüência.

Mallat [45] define uma *Multiresolução* como sendo uma família de espaços de aproximação de funções $\{V_j\}_{j\in Z}$, tal que $V_j \subset V_{j-1} \subset \cdots \subset L^2(R)$, um espaço de Hilbert, onde, intuitivamente, as funções pertencentes a V_j são as que pertencem a V_{j-1} dilatadas por um fator de 2, isto é, com resolução duas vezes menor, estabelecendo a chamada seqüência *diádica*. Neste sentido, j define a resolução do espaço.

Paralelamente, é definida outra família $\{W_j\}_{j\in Z}$ dos complementos de cada espaço da *Multiresolução* V_j no espaço de maior resolução V_{j-1}, isto é, $V_{j-1} = V_j \cup W_j$.

Em [45] é definida uma função ϕ *escalar* tal que para qualquer resolução j, a família de funções $\{\phi_{j,n}\}_{n\in Z} = \frac{1}{\sqrt{2^j}}\phi(\frac{t-n}{2^j})$ é uma base ortonormal de $\{V_j\}$, isto é, qualquer função $f(x) \in V_j$ pode ser construída como combinação linear de $\{\phi_{j,n}\}_{n\in Z}$.

Similarmente, é demonstrado que a partir dessa função ψ pode ser construída uma função *Wavelet* ϕ tal que para qualquer j a família de funções $\{\psi_{j,n}\}_{j,n\in Z} = \frac{1}{\sqrt{2^j}}\psi(\frac{t-2^j n}{2^j})$ é uma base ortonormal de W_j.

Com estes resultados, podem ser construídas as projeções ortogonais de uma função $f(x)$, nos espaços V_j e W_j denotadas respectivamente por $PV_j f$ e $PW_j f$:

$$PV_j f = \sum_{n=-\infty}^{+\infty} a_j[n]\phi_{j,n} \tag{2.23}$$

sendo

$$a_j[n] = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{2^j}}\phi(\frac{t-2^j n}{2^j}) \tag{2.24}$$

onde $a_j[n]$ corresponde à aproximação discreta de f na escala j, isto é, no espaço V_j, e no caso de

$$PW_j f = \sum_{n=-\infty}^{+\infty} d_j[n]\psi_{j,n} \tag{2.25}$$

sendo

$$d_j[n] = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{2^j}}\psi(\frac{t-2^j n}{2^j}) \tag{2.26}$$

$d_j[n]$ corresponde ao chamado sinal de detalhes no espaço W_j, que contem àquela informação do sinal contida em V_{j-1} e não presente em V_j.

Aplicando-se recursivamente este esquema, se define um Banco de Filtros $\{PV_j\}, \{PW_j\}$ que permitem obter uma decomposição do sinal em múltiplas escalas. Os detalhes em cada escala d_j representam a informação do sinal em um determinado intervalo de freqüências e localizado espacialmente.

Se aplicado até um nível K, pode ser demonstrado que a função f é completamente representada por:

$$(a_K, (d_j)_{0\leq j\leq K}) \tag{2.27}$$

e se considerado o processo inverso, f pode ser totalmente recuperada. Este esquema é a base da Transformada Rápida de *Wavelet* ou Piramidal [46, 45, 52] quando o processo é discreto. Neste caso, a mudança de escala deriva em uma sub-amostragem.

No caso de sinais bidimensionais (imagens), todo o raciocínio pode ser estendido de maneira similar à Transformada de Fourier. A propriedade de separabilidade permite que a versão bidimensional seja calculada em termos de combinações de transformadas em cada uma das duas dimensões. No entanto, a diferença fundamental está em que neste processo obtém-se três sinais de detalhes d_j^1, d_j^2, d_j^3 por escala, correspondentes à combinação de dimensões nas diferentes orientações junto com um sinal aproximado a_j.

Em geral, o esquema é representado na figura 2.2. O desenho representado em (a) mostra o conjunto de decomposições nas diferentes escalas e orientações $\{d_i^j\}$, as imagens de detalhe, assim como a imagem filtrada e aproximada em escala menor a_3. Em (b) observamos uma imagem original e em (c) o resultado da aplicação de duas iterações do processo. Pode-se observar em (c) como o processo vai produzindo as diferentes imagens de detalhes contendo a informação da imagem filtradas segundo direção e a escala.

Desta maneira a imagem é completamente representada por:

$$(a_K, (d_j^1), (d_j^2), (d_j^3)_{1 \leq j \leq K}) \tag{2.28}$$

A teoria da representação multiresolução por *Wavelets* é muito interessante para o processamento de textura. A partir da transformada, pode-se analisar, para cada pixel, ou para toda a imagem, o comportamento dos diferentes d_j^i através das escalas. Isto permite obter uma seqüência da relevância local dos diferentes intervalos de freqüências associados às escalas, que pode se interpretar como uma análise simultânea de contraste com escala, isto é, rugosidade. Por outro lado, em uma escala determinada, cada imagem de detalhe diferente concentra informação de filtros orientados (horizontal, vertical e diagonal).

O maior problema das *Wavelets* está em sua sensibilidade à rotação. Uma mesma textura alongada, sofrendo rotações em uma imagem, produz coeficientes muito diferentes nas mesmas escalas, dificultando a caracterização.

Outra abordagem [11] utiliza *Wavelets Packages* o que corresponde a considerar a aplicação da transformada também sobre os sinais de detalhes $\{(d_j)\}$, criando uma grande árvore de possibilidades em escala e espaço. A árvore mais adequada para representar determinados padrões de textura pode ser utilizada como caracterização de textura.

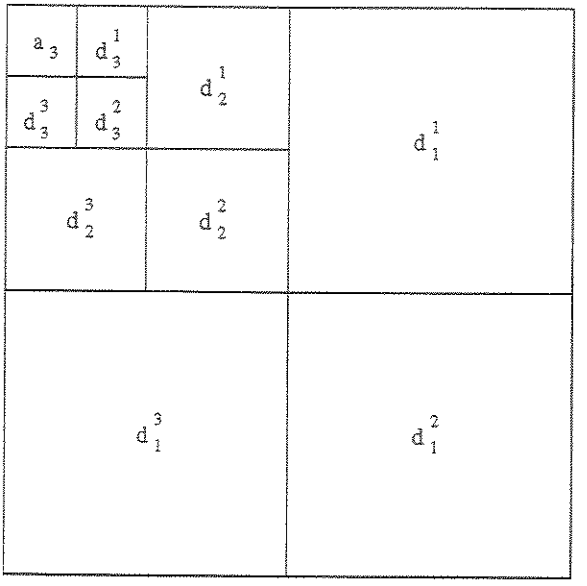
Wavelets são sem dúvida a abordagem mais usada para textura nos últimos tempos, tanto por sua capacidade de discriminação, em geral razoável, quanto por seu baixo custo computacional.

2.6.5 Filtros de Gabor

Filtros de Gabor [22] têm sido utilizados como descritores de textura. Uma função de Gabor tem a forma

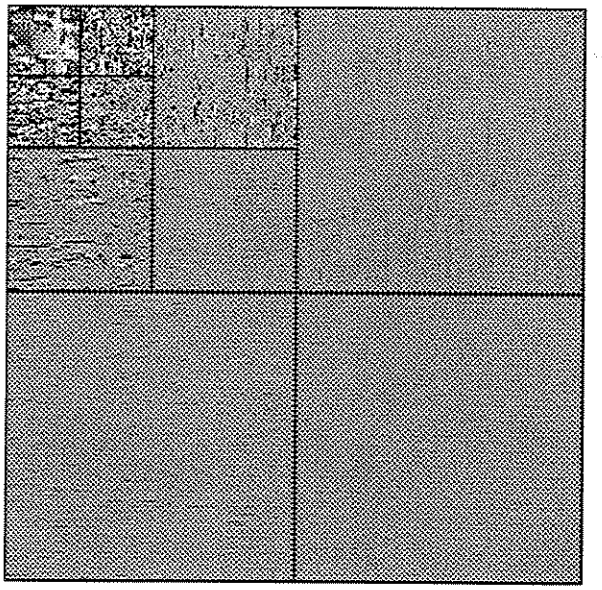
$$g(x, y) = e^{-j\omega_x x + \omega_y y} e^{-(\frac{x}{\lambda_x})^2 - (\frac{y}{\lambda_y})^2} \tag{2.29}$$

com centro de freqüência, na Transformada de Fourier, em $(\omega x, \omega y)$ e orientação no espaço de duas dimensões dada pelos parâmetros $\frac{\lambda_y}{\lambda_x}$, que pode ser definido a priori. Em seu papel



(a)

(b)



(c)

Figura 2.2: Processo de decomposição da imagem pela Transformada Wavelet. (a) Processo de decomposição. (b) Imagem original. (c) Processo de sub-amostragem da Transformada

de filtros, conseguem ser ótimos na representação simultânea em espaço e freqüência, segundo as restrições impostas pelo Princípio de Incerteza [22] . Em [47, 33] estas funções são utilizadas como *Mother Wavelets* mas introduzindo como parâmetros a variação de escala (s) e de orientação (ϕ) : $g_{s,\phi}(x,y)$. Desta forma, dada uma imagem $f(x,y)$ sua Transformada Wavelet de Gabor pode ser escrita como:

$$W_{s,\phi}(x,y) = \int f(x_1,y_1) g_{s,\phi}(x-x_1, y-y_1) dx_1 dy_1 \tag{2.30}$$

Esta Transformada pode ser entendida como filtros detectores de bordas na escala s e na orientação ϕ, onde a média e variância do sinal $W_{s,\phi}(x,y)$ é um descritor de textura conhecido. Aplicando a transformada para vários valores dos parâmetros s e ϕ, é construído um banco de filtros, cada um associado a diferentes centros de freqüência e orientações. Um vetor de características é construído considerando a média e variância associadas ao sinal obtido da aplicação de cada filtro.

2.6.6 Abordagem Morfológica

Uma importante abordagem morfológica para a caracterização da textura foi introduzida por Dougherty [16] a partir da noção de Granulometria. Granulometrias foram introduzidas por Matheron [49] e constituem ferramentas básicas da Teoria da Morfologia Matemática nas abordagens multi-escala.

Definição 2.1 (Granulometria[49]) *Seja $(\psi_\sigma)_{\sigma\geq 0}$ uma família de transformações dependentes de um único parâmetro σ. Esta família constitui uma* **Granulometria** *se e somente se satisfaz as seguintes propriedades:*

$$\begin{aligned} &\forall \sigma \geq 0, \psi_\sigma \text{ é crescente, i.e.,} (f \leq g \Rightarrow \psi_\sigma(f) < \psi_\sigma(g)) \\ &\forall \sigma \geq 0, \psi_\sigma \text{ é anti-estensiva, i.e.,} (\psi_\sigma(f) < Id) \\ &\forall\ \sigma \geq 0, \forall \mu \geq 0, \psi_\sigma \psi_\mu = \psi_\mu \psi_\sigma = \psi_{max(\sigma,\mu)} \end{aligned} \tag{2.31}$$

As últimas duas condições implicam na idempotência do operador ψ_σ e para $i \geq j \Rightarrow \psi_{\sigma_i} f \geq \psi_{\sigma_j} f$.

Definição 2.2 (Resíduo Granulométrico[49]) *Seja $(\psi_\sigma)_{\sigma\geq 0}$ uma granulometria. Os* **Resíduos Granulométricos** *são a diferença entre os resultados obtidos para dois níveis granulométricos sucessivos:*

$$\forall \sigma \geq 0, R_\sigma(X) = \psi_\sigma - \psi_{\sigma+1} \tag{2.32}$$

isto é, R_σ representa a informação preservada no nível σ e eliminada no nível $\sigma+1$ pela família de transformações $(\psi_\sigma)_{\sigma\geq 0}$

O operador morfológico de Abertura ($\circ$) [76] elimina as estruturas da imagem que não contêm a forma representada por um elemento estruturante, e é utilizado freqüentemente como operador das granulometrias.

Pode-se provar que a família de sucessivas aberturas:

$$(\gamma_g^\sigma(f)) = (f \circ g_\sigma)(x) \tag{2.33}$$

constitui uma granulometria, onde g_σ são funções estruturantes formadas pela dilatação de uma função estruturante primitiva $g(x)$:

$$g_\sigma = g_{\sigma-1} \oplus g \tag{2.34}$$

Os resíduos granulométricos, neste caso, correspondem a sucessivas aplicações da Transformação *Top Hat* [76].

$$R_1(X) = f(x) - (f \circ g)(x),$$
$$R_\sigma = (f \circ g_{\sigma-1})(x) - (f \circ g_\sigma)(x)$$

A partir destes conceitos, pode ser idealizada uma análise multi-escala e local da imagem, considerando a variação das suas estruturas e dos resíduos através da dimensão σ, representando a escala.

A abordagem natural para textura é a análise, na vizinhança de cada pixel, do comportamento de alguma magnitude, através das imagens nas diferentes escalas ou nos resíduos morfológicos.

Maragos [48] define o Espectro de Padrões (*Pattern Spectrum*) de uma imagem f (também denominado Espectro Granulométrico) como a função de distribuição de probabilidade δ_f formada pelos Resíduos Morfológicos entre as sucessivas imagens da granulometria de f.

$$\delta_f(i) = \frac{R_i(f)}{\sum_{j=1}^{n} R_j(f)} \tag{2.35}$$

Esta função descreve a variação das perdas de volume das imagens nas diferentes escalas e tem um comportamento diferente para a mesma imagem, dependendo da função estruturante utilizada.

Posteriormente, Dougherty em [16, 12] define a idéia de Espectro de Padrão Local (*Local Pattern Spectrum*) que corresponde à realização da mesma análise em uma janela na vizinhança de cada pixel. Neste caso, para cada pixel x da imagem é calculada uma

função de distribuição δ_x considerando-se apenas a vizinhança dada pela janela $W(x)$, isto é:

$$\delta_x(i) = \frac{R_i(W(x))}{\sum_{j=1}^{n} R_j(W(x))}, \forall \text{ pixel x} \in f \tag{2.36}$$

em que $R_i(W(x))$ corresponde ao Resíduo Morfológico Local para o pixel x, entre as imagens nos níveis i e $i-1$, e calculado sobre a janela $W(x)$.

Vários momentos podem ser calculados e associados a cada pixel considerando-se, por sua vez, várias funções estruturantes. Intuitivamente, a distribuição das perdas de volume deve fornecer informações relativas à forma do elemento estruturante e aos componentes da imagem na vizinhança do pixel.

A hipótese é que em texturas homogêneas a distribuição das perdas deve ser similar e, portanto, os momentos das distribuições podem ser bons descritores de textura.

Certamente outras granulometrias podem ser construídas. Li .et. al. [41] propõem uma granulometria com o operador de abertura por reconstrução que, segundo os autores, garante resíduos granulométricos menos sensíveis ao ruído na análise da textura. Seguindo esta abordagem são conseguidos bons resultados apresentados em [12]. No entanto, a sua principal desvantagem está no alto custo computacional. Simplificações deste esquema devem ser estudados visando garantir uma maior eficiência.

2.6.7 Comentários

Nenhuma abordagem para textura é totalmente bem sucedida, tanto para a segmentação como para a caracterização. Matrizes de co-ocorrência e campos randômicos não consideram o elemento escala e portanto não discriminam bem a rugosidade, além do custo computacional alto. A Transformada Wavelet aparece como a abordagem mais popular e melhor sucedida, dada sua capacidade de discriminar por escala e contraste e seu algoritmo de cálculo eficiente. No entanto, neste caso, texturas iguais com orientações diferentes são discriminadas como sendo diferentes. Filtros de Gabor aparecem em algumas aplicações como a melhor alternativa, pois possuem as vantagens das *Wavelets* com melhor comportamento com a mudança de orientação. A abordagem morfológica por granulometrias mostra resultados altamente precisos de discriminação mas seu custo computacional compromete seu uso em aplicações em tempo real.

O reconhecimento da textura é um problema em aberto. Para o caso de imagens coloridas ou multiespectrais, a abordagem tradicional tem sido a extração da sua componente intensidade ou luminosidade a partir de transformações dos modelos de cores. Sobre esta componente são aplicados então os métodos para imagens em níveis de cinza. Nos últimos tempos têm aparecido trabalhos que estendem alguns dos modelos anteriores para imagens coloridas, considerando a análise simultânea de todas as bandas [55, 34, 74].

2.7 Cor em Recuperação por Conteúdo

A cor corresponde a uma das características mais importantes em Recuperação por Conteúdo. Uma das primeiras decisões a serem tomadas é o espaço de cores a ser considerado. Neste sentido, espaços como HSI ou YIQ [25] em geral são escolhidos por duas razões: (1) resultados experimentais mostram a correspondência com a similaridade humana, quando aplicadas variantes das r-métricas de Minkowsky, (2) a quantidade importante de informação na componente de luminosidade (Y) ou intensidade (I) conduz a uma possível redução do problema ao processamento vetorial destas componentes.

Para a sua apresentação, as abordagens para Recuperação por Conteúdo para cor podem ser divididas em dois grupos: caracterização global ou caracterização local. Estas abordagens são descritas a seguir.

2.7.1 Caracterização global da cor

Histograma

O histograma corresponde à representação mais usualmente empregada na recuperação baseada na característica cor. Para isto, é realizada uma quantização de modo que cada *bin* (ponto do histograma) corresponde, na realidade, a vários valores de cor (valores em cada uma das três bandas). Este processo de quantização pode provocar perda de informação importante quando valores do histograma significativos na imagem, mas numericamente próximos, são colocados no mesmo *bin*.

Várias medidas de similaridade sobre histogramas têm sido propostas. A métrica de Minkowski L_1 é utilizada em [85]. No entanto, é conhecido que com esta métrica mudanças de iluminação podem causar alterações importantes nos valores de similaridade, gerando muitos falsos negativos [84].

Já em [54] é proposta uma interseção de histograma utilizando a métrica

$$sim(H, I) = \sum_{i=1}^{N} min(H(i), I(i)) \tag{2.37}$$

Em outra abordagem, os autores em [21, 10] utilizam uma métrica do tipo L_2 com

$$sim(H, I) = \sqrt{(H - I) \cdot A \cdot (H - I)^T} \tag{2.38}$$

onde A corresponde a uma matriz de similaridade entre diferentes valores de cor. A partir de uma transformação de bases, A pode ser diagonalizada, obtendo-se um novo espaço d_l sem correlações entre as cores, onde a similaridade pode ser dada por:

$$sim(H, I) = \sqrt{\sum_{l=1}^{n} w_l \cdot (hd_l - i_{d_l})^2} \tag{2.39}$$

sendo w_l os autovalores da matriz A.

Histograma Acumulativo

Dado que o histograma e sua quantização podem ser muito sensíveis ao ruído e mudanças de iluminação, e ainda podem ter valores muito dispersos, é introduzido o histograma cumulativo.

Na proposta em [84] para construir o histograma cumulativo é definida uma ordem no espaço de cores. Por exemplo, para as cores $c_j = (r_j, g_j, b_j)$ e $c_i = (r_i, g_i, b_i)$, $c_j < c_i$ se $r_j < r_i, g_j < g_i$ e $b_j < b_i$. O histograma cumulativo h_{cum} é construído a partir do histograma h da imagem como: $h_{cum}(c_j) = \sum_{c_l \leq c_j} h(c_l)$.

As medidas de similaridade são propostas são as diferentes métricas de Minkowski, fundamentalmente L_1, L_2 e L_∞. A maior vantagem do histograma cumulativo é a sua maior estabilidade em relação à quantização.

Outra abordagem [91] considera o histograma como um sinal e aplica sobre ele a Transformada Wavelet. Os coeficientes da transformada são usados então como descrição das características do histograma e utilizados como descritor de cor.

Momentos e conjuntos de Cor

Momentos de cor são introduzidos como forma de diminuir o efeito da quantização no histograma e também a dimensionalidade do vetor de característica para o histograma, em geral de tamanho elevado. Partindo de que os momentos de uma distribuição de probabilidade podem caracterizá-la completamente, alguns destes momentos são utilizados como descritor da cor. Basicamente três momentos são propostos: a média, o desvio padrão e a assimetria. Usando estes momentos, em [84] é proposta como medida de similaridade entre duas imagens H e I a função:

$$sim(H, I) = \sum_{i=1}^{r} w_{i1}|\mu_i(H) - \mu_i(I)| + w_{i2}|\sigma_i(H) - \sigma_i(I)| + w_{i3}|s_i(H) - s_i(I)| \qquad (2.40)$$

sendo r as diferentes bandas de cor e μ, σ e s média, desvio padrão e assimetria, respectivamente. Os valores w_{ik} correspondem a pesos associados a cada momento e cada banda.

Em [84] são combinadas duas abordagens, e a descrição da cor é dada pelos momentos do histograma cumulativo.

No caso dos conjuntos de cores, a idéia é selecionar valores mais significativos do histograma que possam descrever o conteúdo cor.

Por exemplo, em [83], depois de construir o histograma h, cada *bin* $h[m]$ é convertido a um valor $c[m]$ de um histograma binário, que toma valor 1 quando $h[m]$ é maior que um

determinado limiar e 0, em caso contrario. Posteriormente, é utilizada a métrica de tipo L_2 descrita na equação (2.38). Dada a natureza binária de $c[m]$, esta métrica é reduzida a:

$$d(H, I) = \mu_H + \mu_I - 2c_h^T r_t, \tag{2.41}$$

com $\mu_H = c_h^T \cdot A \cdot c_h$, $\mu_I = c_i^T \cdot A \cdot c_i$ e $r_t = A \cdot c_i$, sendo c_i e c_h os histogramas binários das imagens I e H respectivamente e A a matriz de similaridade entre cores. Como c_h é um vetor binário, então a expressão pode ser transformada em:

$$d(H, I) - \mu_H = \mu_I - 2 \sum_{\forall m, c_H[m]=1} r_I[m] \tag{2.42}$$

e desta forma para calcular a similaridade de uma imagem H com uma imagem I no banco de dados, basta armazenar μ_I e $r_I[m]$ os quais podem ser pré-calculados.

2.7.2 Caracterização local da Cor

As caracterizações globais da cor de uma imagem têm capacidade de discriminação limitada, pois não consideram aspectos espaciais ou topológicos da distribuição das cores. Como solução, outras abordagens são propostas para a caracterização da cor considerando descritores locais. Neste sentido, um processo de segmentação é geralmente incorporado.

Como primeiras abordagens para a descrição local, podem ser utilizadas as mesmas propostas para a descrição global, mas considerando apenas a informação de cada região segmentada. A seguir, apresentamos outras propostas que consideram a particularidade da descrição local.

Histograma Multinível

O chamado histograma multinível é apresentado pelos autores em[43]. O objetivo é desta proposta é considerar a similaridade simultânea das cores e sua distribuição espacial na imagem. Cada imagem é decomposta utilizando *quadtrees*, isto é, sucessivamente dividida em quadrantes regulares, até um nível decidido previamente. Para cada quadrante é construído um histograma dos pixels dentro dele. A similaridade entre duas imagens é calculada considerando os diferentes níveis. É proposta uma função de similaridade que considera a representação multinível do histograma. Inicialmente, define-se como interseção entre dois histogramas A e B uma função $h : A \times B \rightarrow H$ tal que o histograma resultante é dado por $H_i = min(A_i, B_i), 1 \leq i \leq m$, sendo m o número de cores diferentes. Introduzindo o histograma multinível, então no i-ésimo nível são comparados dois a dois os 4^{i-1} histogramas das duas imagens. Como resultado final da similaridade é proposto:

$$S_i(A, B) = \frac{1}{4^{i-1}} \sum_{j=1}^{4^{i-1}} \sum_{k=1}^{m} H_j^k \tag{2.43}$$

sendo H_j^k o k-ésimo *bin* do histograma do quadrante H_j obtido pela interseção dos histogramas A_j, B_j. Inicialmente, são comparados os histogramas do nível superior; se estes são similares em relação a um limiar, então é considerado o próximo nível, e assim sucessivamente. Uma imagem é recuperada quando a similaridade no nível das folhas é maior que um limiar.

Código de Cores

Para a caracterização local, os autores em [44] propõem realizar inicialmente uma segmentação supervisionada pelo usuário da Imagem, a partir de detecção de. Cada região obtida é então caracterizada utilizando-se uma abordagem de conjunto de cores.

Para definir o descritor de cor é introduzido um código de cores *CodeBook*. Um código de cores é uma quantização ótima do espaço de cores em 256 valores. Para isto é utilizado o algoritmo generalizado de Loyd [24] que quantiza o espaço de cores a partir de um conjunto inicial de imagens de treinamento, neste caso, uma população inicial do banco de dados.

Para cada região segmentada é escolhido um conjunto de k cores que melhor casa com o código de cores. Como descritor de cor é adotado então:

$$f_c = \{(I_j, P_j) | I_j \in \{1, 2, \cdots, 256\}, 0 \leq P_j \leq 1, \sum_{1 \leq j \leq N} P_j = 1, 1 \leq j \leq\} \tag{2.44}$$

onde I_j é o índice do código de cores e P_j é a porcentagem de cada cor. Note-se que esta abordagem é uma variante da abordagem dos conjuntos de cores apresentada na seção anterior.

Como medida de similaridade entre duas regiões é adotado o seguinte critério. Dadas duas regiões A e B com descritores de cor $\{(I_a, P_a) | 1 \leq a \leq N_a\}$ e $\{(I_b, P_b) | 1 \leq b \leq N_b\}$, respectivamente, define-se inicialmente uma medida entre duas cores quaisquer do código de cores como $W(I_a, I_b) = \|C_{I_a} - C_{I_b}\|$ cujo valor pode ser calculado a priori. A partir daqui é determinada a cor k da região B de menor distância em relação à cor I_a,

$$k = \min_{1 \leq b \leq N_b} W(I_a, I_b) \tag{2.45}$$

A seguir é calculado,

$$D[(I_a, P_a), B] = |P_a - P_k| \cdot W(I_a, I_k) \tag{2.46}$$

como sendo a distância entre uma cor de A, (I_a, P_a), e a região B. De forma similar, pode-se calcular a distância entre uma cor de B e a região A. Como distância final entre duas regiões é proposta:

$$d(A, B) = \sum_{1 \leq a \leq N_a} D[(I_a, P_a), B] + \sum_{1 \leq b \leq N_b} D[(I_b, P_b), A] \tag{2.47}$$

2.7.3 Comentários

Diferentes alternativas têm sido introduzidas para a recuperação de imagens por cor. A qualidade dos resultados neste caso depende do tipo de imagens e aplicações. Por exemplo, em imagens com muita variedade de cores, a abordagem por conjunto de cores é menos discriminante. Um dos problemas fundamentais está na sensibilidade das abordagens à mudança de iluminação ou ruído, produzindo falsos negativos.

Outro aspecto refere-se ao compromisso entre as representações mais compactas (conjuntos de cores, momentos) com fácil manipulação nas estruturas de indexação, mas com menor poder de discriminação, e as representações com vetores em múltiplas dimensões (por exemplo, histogramas), mais efetivas porém com maiores problemas na indexação.

Dependendo da aplicação, a cor pode ser uma característica relevante para a modelagem da similaridade do usuário. No entanto, expressa pouca informação semântica e é, em geral, utilizada em associação com outras características de conteúdo.

2.8 Propostas para Recuperação por Conteúdo

Existem várias propostas de sistemas para Recuperação de Imagens por Conteúdo. Estas propostas utilizam muitas das técnicas para análise de textura e cor apresentadas anteriormente. Estas técnicas são integradas a diferentes modelos de consulta, medidas de similaridade, estruturas de índice, etc. A seguir, serão apresentadas as principais idéias e contribuições de cada uma delas, principalmente no que se refere à textura e cor.

2.8.1 Sistema QBIC

O sistema QBIC (Query By Image Content) [21, 40] desenvolvido na IBM, oferece facilidades para a recuperação combinando características de forma, cor e textura. Para cor e textura são estabelecidos descritores globais. Para a característica de conteúdo forma é realizada uma segmentação da imagem gerando descritores locais.

No caso da textura, inicialmente a imagem colorida é transformada em imagem em tons de cinza. Sobre esta imagem é definido um descritor global baseado em extensões do modelo de Tamura (seção 2.6.1). Como critério de similaridade é utilizada uma distância euclidiana ponderada, no espaço tridimensional destas características de textura.

A cor é representada a partir de uma transformação da imagem em coordenadas do modelo de cores de Munsell [40] e construindo o histograma quantizando o espaço de cores. Como medida de similaridade, é adotada a métrica apresentada na equação (2.38).

Para a recuperação, uma consulta composta pode ser definida como uma composição de condições, utilizando recursos gráficos. Assim, é possível estabelecer porcentagens de cores nas imagens procuradas, desenhar graficamente uma forma próxima do objeto

procurado ou selecionar uma amostra de textura dentro de um dicionário que o sistema visualiza. Estes critérios podem ser combinados, sendo a similaridade total calculada a partir das similaridades de cada característica individual. Este resultado é obtido normalizando todas as medidas individuais de similaridade através da variância e combinando, posteriormente, a similaridade de cada característica em uma soma ponderada cujos pesos são definidos pelo usuário. A indexação é realizada através de *R-trees*.

2.8.2 Sistema PhotoBoook

O sistema PhotoBoook [56] é composto por diferentes módulos para reconhecimento de formas, rostos de pessoas e texturas. O sistema permite a anotação de características das imagens. Estas características podem ser combinadas em uma consulta a qual consiste em apresentar ao sistema um modelo da imagem procurada.

A textura é caracterizada a partir de um Campo Randômico segundo o modelo de Wold (seção 2.6.3). A distância entre imagens é definida a partir das distâncias entre componentes destes campos. Em um primeiro passo, a componente harmônica é calculada e sua periodicidade é analisada, a partir da função e correlação sobre a transformada de Fourier da sua função de espectro. Se a imagem é considerada como periódica, são comparados os picos harmônicos. Caso contrário, são avaliadas as componentes restantes, o que implica em um maior custo computacional. A principal contribuição deste sistema está no modelo original de textura adotado, assim como nos algoritmos para extração de formas e características do rosto humano.

2.8.3 Sistema MARS

No sistema MARS [54] uma imagem é representada por descritores globais de cor, textura e forma. A cor é representada pelo histograma no espaço HSI usando como similaridade a interseção entre histogramas definida na equação (2.37).

No caso da textura é utilizada a representação por *Wavelets*, onde as imagens são decompostas até três níveis, obtendo-se assim 9 bandas de detalhes mais a banda da imagem aproximada (seção 2.6.4). Para cada banda é calculado o desvio padrão dos coeficientes gerando-se um vetor 10-dimensional. A similaridade entre estes vetores utiliza a distância euclidiana. Sobre cada um destes descritores é realizado um processo de normalização para transformar as distâncias em valores no intervalo [0..1].

Para as consultas, o usuário determina as características de interesse sobre as quais são calculadas as similaridades globais, e o sistema automaticamente associa pesos a cada característica. Como linguagem de consulta são propostos dois modelos alternativos [54]: *fuzzy* e probabilístico, que combinam os resultados das similaridades segundo as diferentes características de conteúdo. O resultado da similaridade será o valor final do predicado

avaliado. Para a interação com o usuário é proposto um modelo de consulta com retroalimentação(*feedback*) [65]. Neste modelo, os usuários avaliam as imagens recuperadas e o sistema calcula novos pesos, a partir desta avaliação, em um modelo de refinamento, para realizar uma nova recuperação. O processo se repete até o usuário considerar o resultado aceitável.

2.8.4 Sistema *El niño*

Neste sistema [70] é definido um descritor de imagem global, a partir da Transformada *Wavelet* calculada até três níveis. O critério de similaridade adotado é o modelo de contraste com extensão *fuzzy* (seção 2.3.2). Os autores definem um mecanismo de consulta e navegação que chamam de "exploratório", no qual o usuário se aproxima, em várias iterações, da resposta mais adequada à sua consulta. Propõe-se uma interface de consulta na qual imagens recuperadas são representadas por ícones arranjados em um espaço bidimensional, segundo as distâncias relativas de similaridade que o sistema calcula. O usuário faz um rearranjo dos ícones representando espacialmente sua própria percepção da similaridade relativa entre as imagens. Este rearranjo produz uma modificação em parâmetros na função de similaridade que o sistema utiliza para uma nova recuperação. O processo é repetido até um resultado considerado adequado pelo usuário. Como modelo matemático de similaridade, nesta abordagem, os autores definem o Espaço de Representação como um espaço de Riemann cuja geometria e medidas vão mudando dinamicamente de acordo com as decisões tomadas pelo usuário [68].

2.8.5 Sistema Blobworld

No sistema **Blobworld** [10] o conteúdo das imagens é representado localmente. Uma imagem é tratada como um conjunto de alguns *blobs* que correspondem a regiões da imagem com uma homogeneidade não exata em termos de cor e textura.

A cor de cada região é representada por um histograma no espaço de cores L*a*b* [86]. São definidos 5 *bins* na dimensão L* e 10 *bins* nas dimensões a* e b* . A medida de similaridade adotada é a função descrita pela equação (2.38).

A textura é representada por duas características: contraste e anisotropia. Como medida de similaridade é utilizada a distância euclidiana entre os valores dos pares de valores destas características.

Uma segmentação das imagens é realizada considerando-se cada pixel como um vetor das componentes de cor, as duas características de textura e a posição do pixel. Estes valores são modelados como uma combinação de gaussianas. Um método é apresentado [9] para determinar um número ótimo de gaussianas na composição que produz um melhor

agrupamento de pixels e, finalmente, os pixels conectados são agrupados em *clusters* correspondentes aos *blobs*.

Uma consulta consiste em apresentar ao sistema uma imagem, selecionando seus *blobs* significativos e associando uma importância relativa a cada um. As estruturas de indexação são *R*-trees*.

2.8.6 Sistema VisualSEEK

O sistema **VisualSEEK** [83, 82] baseia-se na representação simultânea de informação de cor juntamente com informação espacial. A imagem é segmentada em regiões a partir da seleção global da melhor combinação de conjuntos de cores (seção 2.7.1) segundo uma técnica chamada de retropropagação de conjunto de cores [82]. Cada região é representada pelo conjunto de cores mais significativo assim como por informações de tipo espacial como área, centroíde, retângulo envolvente mínimo, dentre outras. Para cada uma destas características é definida uma própria medida de similaridade e calculados os conjuntos solução. O resultado da consulta é o conjunto interseção de todos estes conjuntos. Uma consulta é definida por um conjunto de regiões apresentados ao sistema. Cada região tem uma cor e uma distribuição espacial. São utilizadas *R-trees* para a indexação.

2.8.7 Sistema CANDID

A descrição do conteúdo de uma imagem em **CANDID** [39, 38] é de caráter global. Para tanto, é calculado um vetor por cada pixel a partir de várias características de conteúdo (cor, textura, forma). A partir destes vetores é construído um histograma multidimensional de toda a imagem. A seguir é realizada uma estimativa da função de probabilidade, associada ao histograma, como combinação de uma soma ponderada de gaussianas. Como resultado desta estimação são obtidos um vetor de médias, μ_i, e uma matriz de covariância, Σ_i, que servem como assinatura da distribuição e, portanto da imagem. Como função de similaridade entre duas assinaturas em [38] são propostas duas medidas baseadas nas métricas L_1 e L_2 considerando as funções de probabilidade geradas pelas assinaturas como vetores. Outra medida proposta considera como similaridade o produto interno das duas funções de probabilidade definidas pela assinatura.

2.8.8 Projeto ADL

No projeto da Biblioteca Digital Alexandria (*Alexandria Digital Library*) [44] a caracterização das imagens é local. Inicialmente uma segmentação é realizada nas imagens com intervenção do usuário. Esta segmentação é baseada em bordas utilizando a técnica introduzida em [44] chamada de *Edges Flow* e que utiliza características da textura e cor.

Como representação para textura são utilizados os coeficientes da decomposição da imagem segundo filtros de Gabor (seção 2.6.5). A cor de cada região é representada segundo a abordagens de Código de Cores (seção 2.7.2). Uma consulta é realizada a partir de uma imagem de exemplo da qual são extraídas, pelo usuário, regiões de interesse. As características destas regiões são comparadas com as informações locais associadas a cada imagem no banco de dados.

2.8.9 Abordagem de Shekoleshlami

Nos trabalhos de Shekoleshlami et.al.[77, 80, 78, 79, 81] a ênfase está centrada no problema da organização e indexação da informação.

A textura é utilizada como característica de conteúdo através de *Wavelets*. São aplicados dois níveis da transformada e para cada banda calculadas média e variância. Adicionalmente, são calculados o número de pixels de bordas nas três direções como forma de caracterização da direcionalidade, resultando em um vetor.

São criados descritores locais. A segmentação é orientada a região, sendo utilizada a decomposição regular sucessiva do espaço através de um *nonatree*, que é uma generalização do *quadtree* [66]. Cada imagem é dividida em quatro sub-regiões correspondentes aos quadrantes (como no *quadtree*) mais uma sub-região correspondente ao centro da imagem e quatro sub-regiões nas laterais, todas de mesmo tamanho. Como resultado do processo cada sub-região diferente é indexada

Este processo gera um número importante de segmentos para cada imagem porque segmentos adjacentes com a mesma textura não são fundidos. Para aliviar este problema é proposto um complexo mecanismo de indexação, utilizando técnicas de classificação supervisionada. Para a indexação é utilizado *R-tree*.

Para realizar uma consulta, o usuário fornece um padrão de textura extraído de uma imagem. O sistema determina sub-regiões onde podem existir padrões similares. Caso sejam recuperados vários segmentos associados a uma mesma imagem o maior valor de similaridade é considerado como a similaridade resultante. As imagens utilizadas nesta abordagem são ISR.

2.8.10 Abordagem de Vellaikal et. al.

Em vários trabalhos Vellaikal [91, 90, 92, 93] apresenta diferentes abordagens de consulta por conteúdo.

Para a caracterização das imagens uma análise espaço-espectral é utilizada. São definidos descritores locais como resultado de uma segmentação por decomposição regular utilizando *quadtree*. A imagem é dividida sucessivamente em quatro quadrantes de igual

tamanho até conseguir segmentos espectralmente homogêneos. Um segmento é considerado homogêneo quando a variância total dos seus pixels é menor que um limiar dado.

Cada segmento do *quadtree* é analisado e seus pixels são agrupados em um conjunto de *clusters* a partir de um algoritmo *k-mean* [25]. O número k de clusters por segmento é incrementado até que o erro quadrático médio em todas as bandas de espectro sejam menores que um limiar definido. O vetor de características associado a um segmento é formado pela média e o número de pixels associados a cada um dos *clusters* definidos dentro desse segmento. A proposta não faz nenhuma consideração quanto à utilização de estruturas de indexação.

A aplicação considera ISR. Uma consulta consiste em apresentar diferentes padrões ou classes ao sistema, (por exemplo classes de solo, nuvens, neve, etc) e considera-se que os valores multiespectrais dos pixels nestas classes possuem distribuição gaussiana. Um *cluster* em um segmento contém pixels de uma classe se a distância ponderada entre a média da classe e a média do *cluster* é menor que um determinada limiar. Este limiar e a definição de distância são ponderados com a variância, como reflexo da dispersão espacial dos pontos.

Como uma variação da abordagem anterior em [90], o critério de segmentação adotado é a subdivisão a priori da imagem numa rede de segmentos regulares. Os segmentos têm um tamanho predeterminado e sobre eles é realizado o mesmo processo de *clustering*.

2.9 Resumo

Este capítulo apresentou uma revisão bibliográfica das principais vertentes de caracterização de imagens visando sua descrição. Esta caracterização busca representar propriedades das imagens (tipicamente cor, forma e textura) através de descritores, que nada mais são do que conjuntos de valores que sintetizam estas propriedades. Estes valores são obtidos a partir da aplicação de vários tipos de algoritmos.

A recuperação baseada em conteúdo consiste em buscar, em um banco de dados de imagens, o conjunto que mais se aproxime de um conjunto de características desejadas. Tipicamente, as características desejadas são expressas por descritores.

A similaridade entre os descritores de entrada (características desejadas) e os de cada imagem do banco de dados é calculada de acordo com um conjunto de medidas preestabelecidas, associadas aos descritores. Desta forma, o processo de recuperação por conteúdo exige a caracterização de informação sobre os descritores e sobre as métricas associadas.

A extração de descritores é ainda tópico de pesquisa de ponta. A definição de métricas, também tópico de pesquisa, depende ainda dos objetivos dos usuários, do tipo de imagem considerada e das aplicações alvo.

O próximo capítulo descreve uma família de imagens para as quais descritores e

métricas são particularmente complexos - imagens de sensoriamento remoto, objeto desta tese.

Capítulo 3

Imagens de Sensoriamento Remoto

3.1 Introdução

O capítulo anterior apresentou uma revisão sobre os principais problemas para a recuperação de imagens por conteúdo, as técnicas de análise e processamento de imagens usadas como propostas de solução e, finalmente, alguns exemplos de sistemas orientados a este tipo de recuperação.

Este capítulo introduz as características das Imagens de Sensoriamento Remoto (ISR), e a maneira como estas são processadas pelos especialistas. Discute, igualmente, as dificuldades para a recuperação por conteúdo deste tipo de imagens, as limitações para a aplicação das abordagens apresentadas no capítulo anterior, e define um conjunto de aspectos que orientarão a proposta apresentada nesta tese.

3.2 Propriedades das ISR

Imagens de Sensoriamento Remoto (ISR) são obtidas a partir de equipamentos aéreos que, por meio de sensores, realizam uma coleta de dados da superfície terrestre, quantizados em um número discreto de níveis. Suas características mais relevantes são:

- As ISR são multi-espectrais, isto é, de diferentes bandas do espectro de sinais. As chamadas bandas *passivas* refletem as emissões de energia externas ao dispositivo sensor. Algumas refletem a disposição espacial da *radiação solar* nas bandas ultravioleta, visível e perto do infravermelho. Outras, chamadas de *termais*, estão associadas à distribuição espacial de *energia* emitida pela Terra e são obtidas nas bandas do espectro infravermelho. As bandas *ativas*, por sua vez, captam a resposta da Terra a emissões de energia por dispositivos, geralmente radares, no espectro de microondas.

- Os diferentes valores multi-espectrais captados dependem fortemente da natureza física dos objetos da Terra observados pelos sensores, com alta probabilidade de introdução de ruído, variações segundo as condições climáticas no momento da coleta dos dados (luz solar, nuvens). Uma ISR em sua forma original é difícil de visualizar e, portanto, precisa de filtragens, aumentos de contraste, dentre outras transformações, para a visualização pelos especialistas.

- O caráter geográfico da informação implica na associação de coordenadas aos pixels. Este aspecto introduz problemas de distorção. Enquanto a superfície da Terra é redonda, a informação é captada e representada de forma planar, precisando de correção geométrica, com introdução de possíveis erros. Esta informação geográfica representa, no entanto, uma propriedade imutável da imagem o que a define como um metadado importante associado a cada ISR.

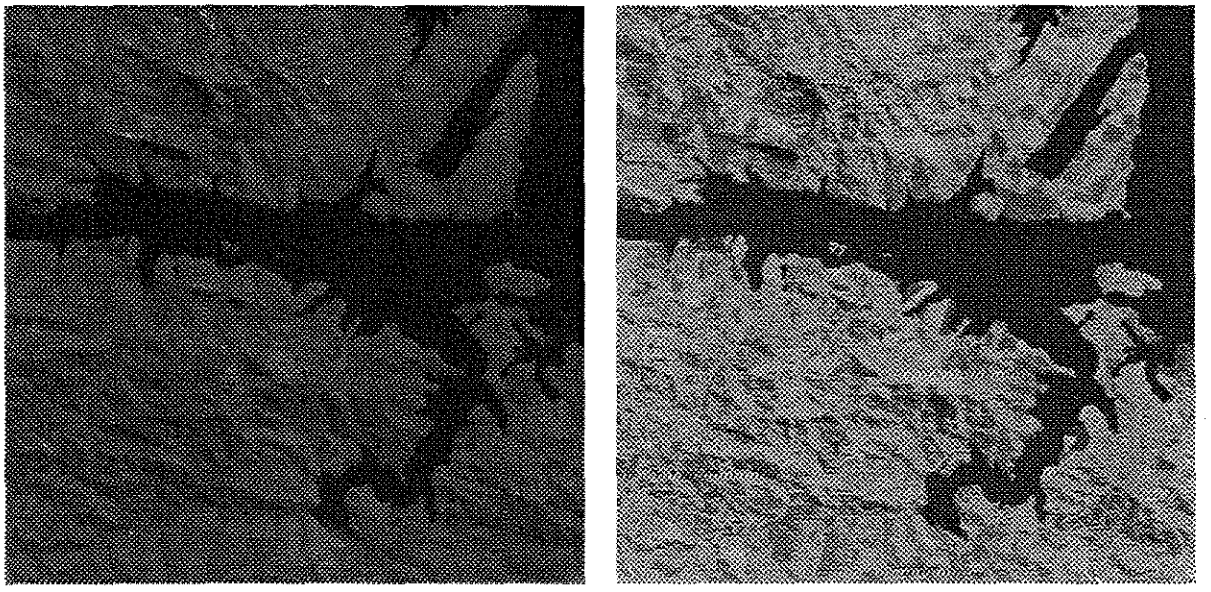
- O conteúdo das ISR é altamente complexo, diverso e não uniforme e as diferentes bandas das imagens não correspondem em geral ao espectro visível criando, portanto, percepções que não correspondem ao cenário visual real da superfície da Terra que representam

As imagens em que esta tese se concentra são obtidas do sistema de satélite **LANDSAT**, em particular, do sensor TM (*Thematic Maper*). O TM [62, 63] é um equipamento de sensoriamento mecânico que adquire os dados da Terra em faixas de largura de 185 km em 7 bandas do espectro eletromagnético. As bandas utilizadas são de número 3, 4 e 5, correspondentes a uma banda visível vermelha e duas de infravermelho.

A figura 3.1 representa à esquerda uma ISR em seu estado original, e depois de ter sido transformada com aumento de contraste. Pode-se constatar a diferença de possibilidades perceptuais entre uma e outra.

3.3 Processamento das ISR

Existem duas abordagens para a interpretação e manipulação de ISR [62]. A chamada **Foto-interpretação** é realizada diretamente pelo operador humano e a **Análise Quantitativa** de modo automático. A Análise Quantitativa permite explorar pixel por pixel da imagem, combinar imagens de diferentes espectros, assim como a análise de níveis de brilho da imagem que o especialista não consegue identificar. Por sua vez, a Foto-interpretação depende da experiência e habilidade do especialista. Em geral, este consegue distinguir



(a) (b)

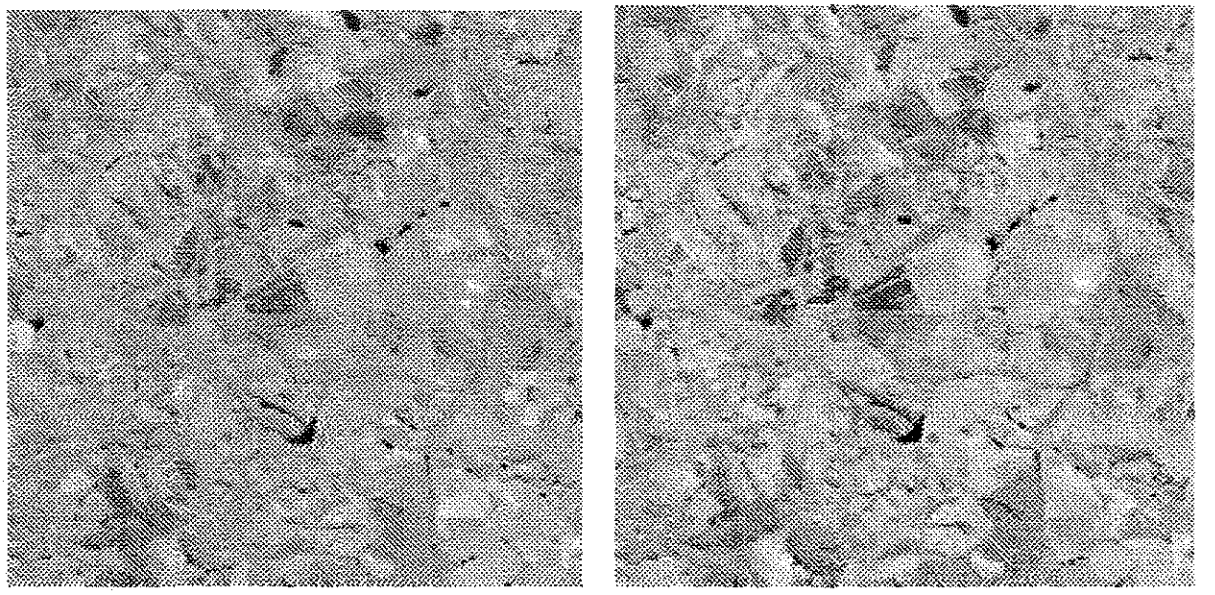
Figura 3.1: Imagem ISR original(a) e depois do aumento do contraste(b)

facilmente padrões globais na imagem, o que é difícil de ser realizado no processo automatizado. Ambas abordagens são complementares para a interpretação e análise de uma imagem.

Uma imagem de trabalho é obtida como resultado da aplicação simultânea e cíclica das duas abordagens para o processamento de ISR. Para isto, são aplicadas filtragens, aumentos de contrastes, dentre outras operações, para facilitar a "percepção das imagens", permitindo ao especialista reconhecer aqueles elementos de interesse de sua aplicação. Este processo de transformação facilita a percepção mas introduz distorções e mudanças na informação originalmente coletada.

O especialista identifica regularidades ou propriedades no espaço a partir da informação da imagem. Uma primeira decisão do especialista é a composição de cores utilizada, isto é, a associação das bandas espectrais às bandas visíveis. O fato de que as bandas de uma ISR não representam bandas do espectro visível faz com que a associação de cores a cada uma delas, para efeitos de visualização, seja um processo subjetivo. A decisão da associação de bandas da ISR a bandas visíveis cria "pseudocores" que determinam como os objetos ou características são percebidos pelo especialista. Esta decisão é muito dependente do especialista que, com o tempo, percebe como "natural", a maneira em que os objetos do mundo real são visualizados pela composição adotada. Algumas composições podem ser mais adequadas para determinados tipos de aplicações e objetos observados, segundo os usuários. No entanto, o fato real é que a decisão da composição tem uma componente subjetiva muito grande.

A figura 3.2 mostra duas composições de cores diferentes para a mesma ISR. Vemos que conseguimos identificar em ambas composições os mesmos objetos e perceber as duas como representando a mesma imagem, mas os detalhes e a percepção de cores são diferentes.



(a) (b)

Figura 3.2: Imagem ISR com associações de bandas a cores diferentes (a) composição de bandas 3-4-5 (RGB). (b) composição de bandas 4-5-3 (RGB)

O resultado final do processamento de ISR em geral é a segmentação e classificação destas em regiões dependendo do estudo sendo realizado. Estas regiões ou classes são associadas a diferentes categorias, tais como tipo de vegetação, culturas, tipos de solos, características geológicas, etc. O objetivo é definir mapas temáticos. Em todo este processo, a experiência do especialista e os interesses da aplicação influenciam ativamente o processo automatizado.

Múltiplos métodos de classificação são aplicados com diferentes graus de efetividade dependendo do tipo de imagem que se estuda. Embora utilizando a informação das múltiplas bandas, não existe um critério de classificação multi-espectral universal aceito. Portanto, os métodos de segmentação, classificação e filtragem são muito dependentes da aplicação, do tipo de imagem, das condições da sua coleta pelos sensores e da experiência do especialista.

3.4 Características das ISR e Recuperação baseada em conteúdo

A partir dos elementos discutidos anteriormente, apresentamos características das ISR que as diferenciam de outras classes de imagens.

3.4.1 Características de conteúdo relevantes para ISR

A distribuição de cores, as texturas, a forma e geometria dos objetos estão dentre as mais importantes características das ISR que os especialistas utilizam para seu processamento e, portanto, relevantes para sua recuperação baseada em conteúdo

A característica *cor*, embora muito importante, é sensível às condições da coleta, de tal forma que imagens que representam um mesmo tipo de objeto (por exemplo, o mesmo tipo de plantação) podem aparecer com variações nas colorações em imagens coletadas em momentos diferentes. Estas variações, em geral, aparecem como cores relativamente similares do ponto de vista visual, mas com graus de diferença. Neste sentido, antes de poder "comparar" as imagens visualmente, muitas vezes os especialistas aplicam técnicas de pré-processamento para tentar "normalizar" as imagens a serem analisadas [15, 59].

A característica *textura* é também altamente significativa nas aplicações e menos sensível às condições climáticas e ao ruído [15]. Porém, as técnicas existentes para análise de textura são imprecisas e, segundo os especialistas, as ferramentas que aparecem atualmente nos softwares comerciais (SIGs) não conseguem bons resultados de classificação. Em [73] são apresentados resultados alentadores para textura em ISR baseados na abordagem de Campos Markovianos apresentada no capítulo 2, mas com o alto custo de processamento que estas técnicas impõem.

O problema do reconhecimento de *formas e objetos* nas imagens está entre os mais complexos do ponto de vista de processamento de ISR, e as técnicas são dependentes do tipo de aplicação, assim como do conhecimento prévio da forma dos objetos procurados.

Existem vários problemas associados à natureza das ISR que dificultam sua recuperação por conteúdo. Estes problemas são discutidos a seguir.

3.4.2 O problema da composição de pseudocores

Como apresentado no capítulo anterior, muitas das técnicas de recuperação por conteúdo baseadas em cores exploram a conhecida diferença de sensibilidade da percepção humana às distintas bandas visíveis. Esta diferenciação humana tem levado à definição de vários sistemas de cores que associam pesos diferentes às diferentes bandas do espectro visual. Sistemas como HSI, L*a*b, YIQ dentre outros, partem destas hipóteses.

A grande vantagem destes sistemas está em que funções de distância ou métricas, definidas sobre eles, podem modelar de maneira mais efetiva a percepção humana de similaridade entre cores. Valores de distância pequenos representam realmente cores similares, e valores altos, cores perceptualmente diferentes.

No caso das ISR, a variação perceptual resultante da composição de bandas limita a utilização destes sistemas de cores. O fato de que cada especialista possa definir sua própria composição de cores, e inclusive mudar essa composição dinamicamente, torna difícil armazenar descritores de cores baseados nos sistemas mencionados.

O processamento eficiente de uma consulta por conteúdo deve basear-se necessariamente na utilização de estruturas de índice multidimensional que modelem e respondam eficientemente a consultas baseadas em uma determinada métrica. Uma mudança de composição implicaria numa mudança de métrica e, necessariamente, na necessidade de reconstrução dessa estrutura de índice.

3.4.3 O problema dos modelos de descrição de conteúdo

Um dos problemas fundamentais para a recuperação por conteúdo de ISR é a falta de modelos matemáticos de descrição do conteúdo que sejam universalmente efetivos.

Na seção anterior foi apresentado o problema da limitação no uso dos sistemas de cores. No caso da textura, a diversidade de padrões encontrados nas ISR faz com que seja difícil achar um modelo universal para sua segmentação, caracterização e modelagem da similaridade.

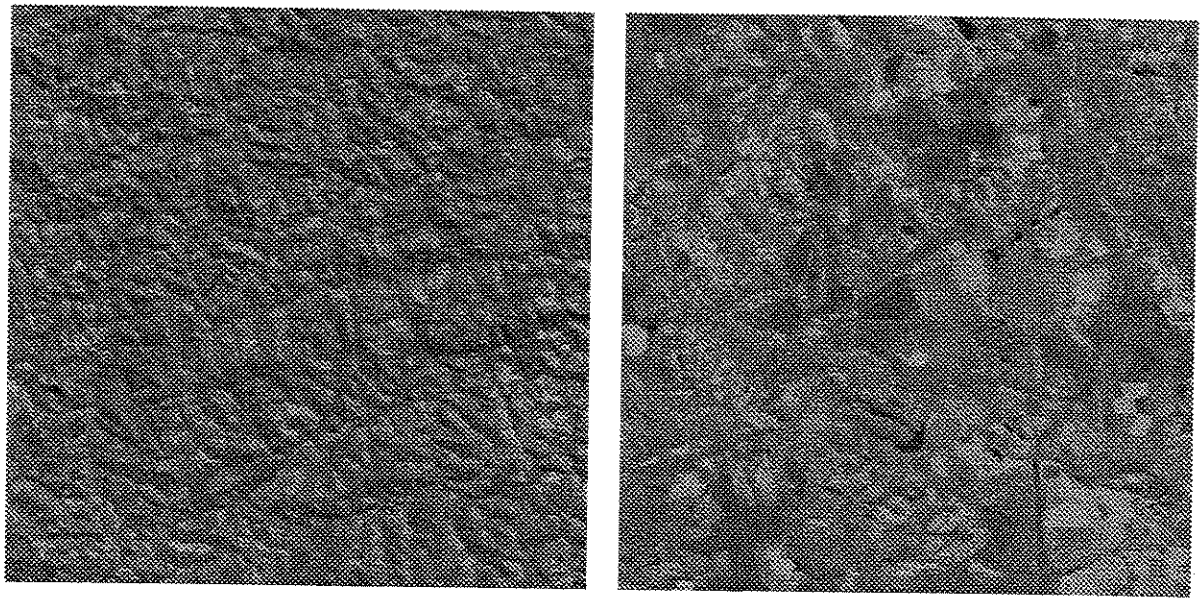
Na nossa experiência existem modelos que são efetivos para determinados padrões em ISR mas que não conseguem caracterizar corretamente outros.

3.4.4 O problema da homogeneidade/heterogeneidade

Um dos aspectos aceitos dentro da área de recuperação baseada em conteúdo é o fato de que uma especialização do tipo de imagens processadas aumenta a probabilidade da efetividade da sua recuperação. Desta maneira, imagens com estruturas bem definidas, como células humanas ou animais, texturas de tecidos, rostos, etc, facilitam a recuperação. Embora para algumas destas imagens o problema da definição de um critério de similaridade pode ser complexo (por exemplo, similaridade entre rostos) a homogeneidade da informação que se repete em todas elas permite a aplicação de técnicas específicas com maior probabilidade de sucesso. Um dos aspectos que esta homogeneidade oferece é decidir se a descrição da imagem deve ser global ou local, isto é, deve ser segmentada alguma parte da imagem ou deve ser processada como um todo.

Já o conteúdo representado em ISR pode ser, em determinadas ocasiões, homogêneo e em outras, diverso e complexo. A figura 3.3 mostra duas ISR. Na primeira um mesmo

padrão de textura e cor bastante homogêneo aparece em toda a imagem. Na imagem da direita pode-se perceber a variedade de texturas e cores.



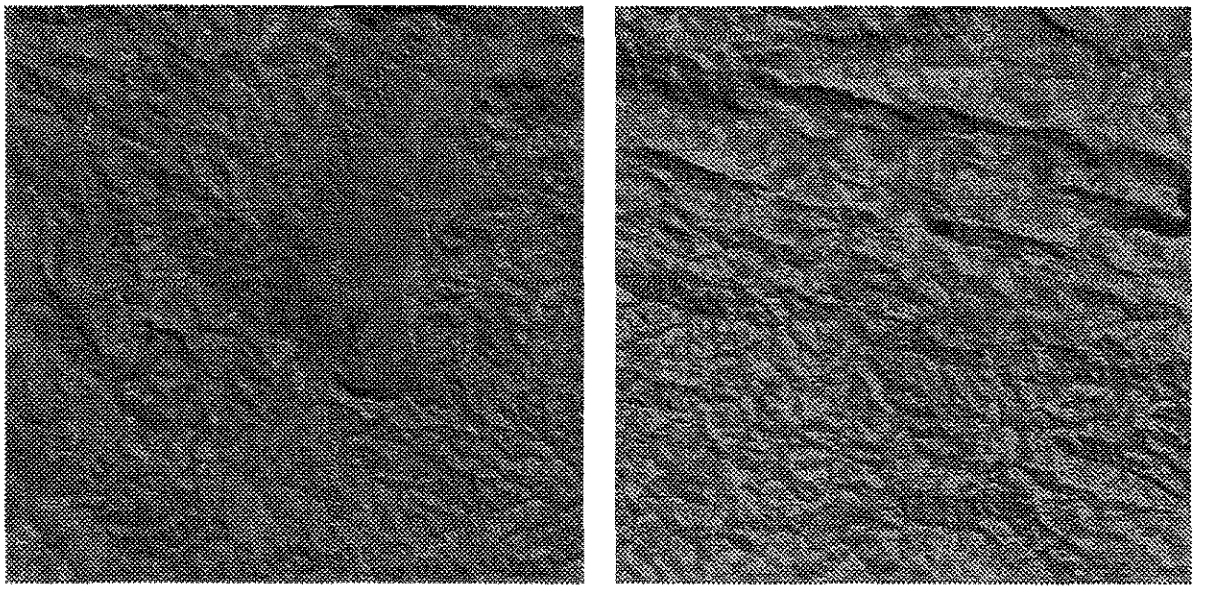
(a) (b)

Figura 3.3: Duas imagens com conteúdo diferente: (a) Imagem com homogeneidade no seu conteúdo. Textura de toda a imagem é relativamente homogênea. (b) Imagem com heterogeneidade no conteúdo, textura e cores na imagem com múltiplos padrões e regiões bem definidas

Um outro problema associado à diversidade do conteúdo das ISR está na delimitação visual das regiões homogêneas dentro das imagens, isto é, aquelas regiões para as quais conseguimos definir visualmente os limites. Esta delimitação é muito importante para um processo de segmentação que permita descrever o conteúdo localmente. Em algumas imagens esta delimitação é bem definida mas em outras, a definição desse limite é totalmente difuso.

A figura 3.4 ilustra este problema. A imagem da esquerda possui dois padrões de textura bem delimitados. Na imagem da direita, estes padrões não conseguem ser bem delimitados fundindo-se um no outro.

A partir desta discussão pode-se concluir que é complexa a decisão sobre que tipo de descritor (global/local) deve ser utilizado em ISR. Seguindo nesta mesma direção, no caso



(a) (b)

Figura 3.4: Delimitação de objetos da imagem: (a) Imagem onde a textura define limites de objeto. A imagem possui uma região central bem delimitada pela textura (b) Imagem onde existem mudanças de textura mas os limites entre elas não estão bem definidos

em que seja preciso algum tipo de segmentação para descrever localmente a imagem, este processo de segmentação é altamente complexo. De fato, não se conhecem ainda algoritmos universais de segmentação para ISR.

3.4.5 O problema da visualização e a similaridade

Como mencionado nas seções anteriores, uma ISR original é extremamente difícil de analisar, sendo necessários processos de composição de cores e de filtragens/transformações para permitir a visualização e interpretação. No entanto, estes processos levam a uma distorção ou mudança da informação original que a imagem carrega.

Como resultado, duas ISR com conteúdos similares podem sofrer transformações que distorçam essa similaridade e dificultem o processo de detectar esses conteúdos como similares.

Um principio que adotaremos nesta tese é considerar que a caracterização do conteúdo de uma imagem será sempre realizada a partir da informação básica, isto é. o processamento será realizado sempre na imagem original. Os efeitos de visualização das imagens apresentadas a um usuário podem ser resultantes de transformações, mas o processamento deve ser realizado sobre a informação obtida originalmente.

Uma hipótese básica do trabalho é considerar que imagens com características simi-

lares, ainda quando transformadas e modificadas para efeito de visualização, mantém algum grau de similaridade dessas mesmas características. O processo de recuperação por conteúdo parte da participação ativa do usuário. Se o usuário não consegue avaliar essa similaridade, o sistema de recuperação com intervenção humana perde o sentido.

3.5 Premissas para um sistema de recuperação de ISR baseada em conteúdo

Diferentes trabalhos reportam aplicações de recuperação por conteúdo em ISR [6, 38, 81, 80, 91].

O projeto CANDID [38] baseia-se em descrição global da imagem. A extração de características e a medida de similaridade consideram as diferentes bandas espectrais para a estimação de uma função de densidade de probabilidade multidimensional como soma de funções gaussianas.

Em [81] é utilizada a Transformada *Wavelet* para descrever textura e as consultas são realizadas a partir de apresentar um padrão de textura ao sistema

A abordagem apresentada em [73] define um padrão de textura a partir de uma ISR para a busca de imagens similares. A modelagem da textura é realizada a partir de Campos Randômicos de Gibbs.

Em geral, a principal abordagem nestes sistemas, em relação ao tipo de imagens, consiste em apresentar descritores que usam as múltiplas bandas espectrais. No entanto, não são oferecidos recursos para a interação e consulta deste tipo de imagens e usuários.

A partir dos aspectos analisados na seção anterior chegamos às seguintes conclusões:

- Não existem resultados conclusivos quanto a modelos e algoritmos para textura e cor em ISR. Quando uma ISR é processada e analisada, vários modelos são utilizados simultaneamente visando comparar os resultados.

- De maneira análoga à forma de trabalho seguida pelos especialistas, é previsível que os critérios de recuperação em um banco de dados de ISR sejam dependentes do tipo de usuário. Uma imagem aceita como similar por um usuário pode ser rejeitada por outro, empregando o mesmo modelo de consulta. Além disto, a relevância da textura e da cor pode variar de usuário para usuário ou de acordo com a aplicação considerada.

- Um importante grau de incerteza deve ser considerado no modelo de consulta. O usuário tem uma idéia de que elementos ou padrões está procurando, mas não sabe exatamente quais as imagens. A sua idéia da relevância de uma imagem depende

da comparação com outras, isto é, o critério de similaridade deve ser refinado no processo de recuperação já que as imagens procuradas podem ser visualizadas de diferentes maneiras.

- Regiões que representam um mesmo tipo de entidade ou objeto em imagens diferentes, podem apresentar diferenças em relação á cor e a textura. No entanto, existe uma semelhança perceptual entre estas regiões, que permite aos especialistas reconhece-las como representando a mesma entidade.

- A complexidade do conteúdo de uma ISR sugere formas de recuperação com descritores locais que implica numa segmentação da imagem no processamento. Em outros tipos de imagens, a semântica de alto nível tem maior importância na decisão da relevância de uma imagem. Nas ISR, as características de baixo nível são muito importantes como critério de relevância.

- A textura e cor em ISR são estabelecidas de modo diferente em função da composição de cores escolhida. Isto significa que não devem ser utilizadas técnicas dependentes de uma composição de cores predeterminada.

- A busca de imagens que contenham regiões ou padrões similares a alguns já conhecidos (caracterização local) é tão importante quanto um critério de similaridade da imagem como um todo (caracterização global).

- A relevância das diferentes bandas pode ser diferente, eventualmente, dado o conhecimento do especialista, sobre o tipo de informação contida em cada banda.

3.6 Resumo

Este capitulo apresentou uma caracterização das Imagens de Sensoriamento Remoto e alguns dos problemas intrínsecos a este tipo de imagens que, em geral, não se apresentam em outras classes de imagens. A natureza diferenciada das ISR impõe um conjunto de necessidades a um sistema de recuperação por conteúdo para tentar diminuir o impacto destes problemas. O capítulo seguinte apresenta um modelo de recuperação por conteúdo que tenta considerar estas necessidades.

Capítulo 4

O Modelo Proposto

4.1 Introdução

Este capítulo apresenta um arcabouço para a recuperação por conteúdo de imagens de sensoriamento remoto (ISR). Para sua realização foram consideradas as características próprias deste tipo de imagens e a manipulação destas pelos especialistas, como abordado no capítulo anterior.

Do ponto de vista do usuário, a proposta apresentada neste capítulo baseia-se em dois aspectos fundamentais: a utilização de mecanismos de *retroalimentação (relevance feedback)* e o conceito de *padrão* como elemento básico na definição de consultas. Adicionalmente, e de maneira transparente para a visão do usuário, é introduzido o uso de *múltiplos modelos de representação do conteúdo.* Um modelo de representação de conteúdo é um modelo matemático para descrever e diferenciar características de conteúdo de uma imagem como textura ou cor. O uso de múltiplos modelos é um dos aspectos importantes desta tese. Este aspecto é determinante para o gerenciamento do banco de imagens e das consultas. Estes elementos - padrões, retroalimentação e suporte a múltiplos modelos - correspondem ao núcleo da proposta e são apresentados e combinados ao longo do capítulo.

De acordo com o analisado no capítulo 3, um sistema de recuperação de imagens por conteúdo em geral não é baseado em consultas exatas e casamento exato e constrói a resposta a uma consulta na base de um critério de similaridade, com a incerteza e imprecisão associadas. Neste sentido, outro aspecto básico da proposta, apresentado neste capítulo, é a definição da medida de similaridade que será utilizada como base do processamento de consultas. Esta medida de similaridade incorpora os três elementos da proposta mencionados no parágrafo anterior.

4.2 Padrão como conceito de consulta

Para um especialista em ISR, a principal expectativa sobre um sistema de recuperação de imagens por conteúdo é obter imagens que contenham padrões visuais e espectrais similares a regiões de imagens já conhecidas e verificadas no mundo real. Estes padrões, baseados fundamentalmente em textura e cor, representam conceitos ou entidades que, se espera, apareçam com padrões mais ou menos similares em outras imagens, por exemplo: tipos de cultura, tipos de solos, dentre outros. No entanto, como apresentado no capítulo anterior certos fatores podem produzir diferenças nesses padrões, quando presentes em imagens distintas, que complicam a possibilidade de identificá-los como válidos (similares ao procurado).

Este trabalho propõe um conceito de padrão, baseado no tipo de requisitos de usuários do domínio de ISR.

4.2.1 Motivação do padrão: classificação

Uma das operações mais comuns no processamento de ISR é a classificação [62, 15, 59] cujo objetivo é particionar os pixels de conjuntos de imagens de acordo com determinadas características. Um algoritmo de classificação supervisionado define a priori um conjunto de classes $\{C_1, C_2, \cdots, C_n\}$ e junto com elas múltiplas amostras ou exemplos dos objetos que as compõem

O problema fundamental a ser resolvido em um algoritmo de classificação é o seguinte: dado um objeto não conhecido, determinar a qual das classes predefinidas ele pertence ou, eventualmente, decidir que não pertence a nenhuma. Quanto maior o número inicial de amostras em cada classe, mais precisa deve ser a resposta esperada. Há vários métodos de classificação supervisionada amplamente utilizados em várias áreas do conhecimento.

Sem perda de generalidade, os métodos de classificação podem ser descritos como processos onde intervêm três funções:

- Uma função que processa todas as amostras de uma classe e gera uma descrição da classe.
- Uma função de processamento do objeto a ser classificado que extrai uma descrição de suas propriedades.
- Um "classificador"que determina um valor da "semelhança"da descrição do objeto a ser classificado com as descrições das classes e determina a qual destas classes o objeto pertence.

Os métodos de classificação são amplamente usados em processamento de ISR de acordo com o seguinte procedimento:

1. Determinam conjuntos de pixels em uma ou várias imagens que representam áreas conhecidas de um mesmo conceito ou entidade (definidas por trabalho de campo, dados comprovados, etc) agrupados em classes, por exemplo: zona urbana, cultura de cana de açúcar, fonte de água, etc.

2. Classificam por meio de uma função de classificação os pixels não identificados das imagens em alguma dessas classes.

3. Obtêm mapas temáticos onde a classe associada a cada pixel determina a entidade do mundo real que, espera-se, deve corresponder à localização geográfica associada a esse pixel.

4.2.2 Padrões e consultas

Fazendo uma analogia com o problema da classificação, o problema da recuperação de imagens por conteúdo pode ser entendido como a definição de várias entidades que representam classes significativas para uma aplicação. O objetivo da recuperação é "classificar" as diferentes regiões das imagens armazenadas no Banco de Dados como pertencentes ou não a uma das classes. O grau de "semelhança" de uma imagem com objetos de uma classe determina quais imagens devem compor o conjunto resposta.

Como cada classe deve possuir um número significativo de amostras predefinidas, de maneira análoga podemos considerar que as entidades procuradas no banco de dados devem estar bem definidas. Cada entidade deve ser representada inicialmente por um número de amostras bem conhecidas que descrevam as diferentes formas com que o mesmo conceito pode aparecer. Esta idéia corresponde ao conceito de *padrão* considerado neste trabalho.

Definição 4.1 (Padrão) *Um* **Padrão** p_j *é definido como um conjunto de regiões* $p_j = \{r_1[p_j], r_2[p_j], \cdots, r_m[p_j]\}$ *de uma ou várias imagens que um usuário associa a uma mesma entidade.*

Definição 4.2 (Consulta por padrões) *Uma* **consulta por padrões** $Q = (P, N)$ *é definida como uma consulta cujo argumento é um conjunto de padrões* $P = \{p_1, p_2, \cdots, p_{np}\}$ *e um número máximo de imagens* N *a serem recuperadas e que tenham similaridade com o conjunto* P.

Em outras palavras, a consulta por conteúdo passa a ser consulta por padrões. Um padrão corresponde a um conjunto de regiões extraídas pelo usuário de diferentes imagens por ele conhecidas, que chamamos de **Imagens de Consulta**. São estas imagens que são utilizadas pelo usuário na formulação inicial da sua consulta.

A figura 4.1 mostra um diagrama de especificação de uma consulta pelo usuário. Inicialmente, a partir das imagens de consulta no lado esquerdo, o usuário seleciona diferentes regiões obtendo dois padrões: o padrão "mais escuro"e o padrão "mais claro".

Esta noção de padrão generaliza a noção usual de padrão: ao invés de utilizar uma imagem de consulta, é possível utilizar várias imagens como fornecedoras de parâmetros de consulta.

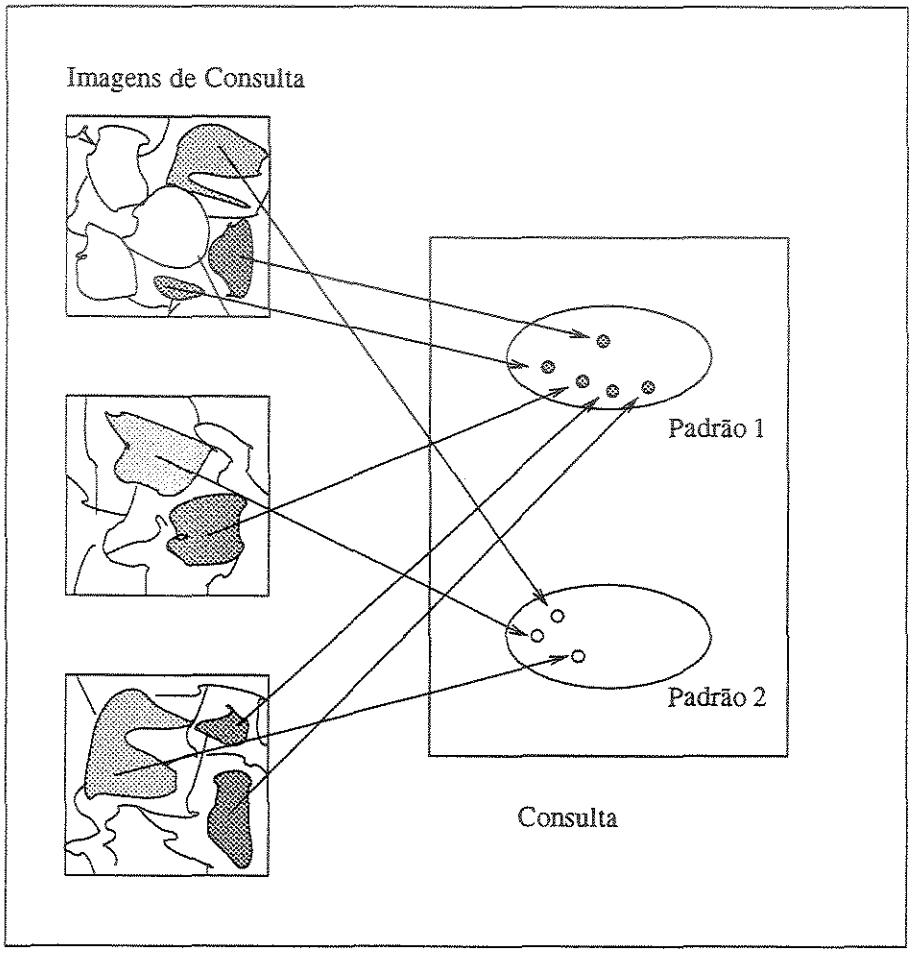


Figura 4.1: Definição de uma consulta. Cada elipse corresponde a um padrão, criado a partir de amostras de diferentes imagens de consulta.

Juntamente com a definição da consulta, é necessário definir o resultado esperado. O sistema deve retornar o conjunto das imagens contendo regiões, que segundo sua avaliação de similaridade, sejam conjuntamente as mais similares a cada um dos padrões da consulta, isto é, as imagens com as regiões melhor classificadas conjuntamente como pertencentes às classes definidas pelos padrões.

Definição 4.3 (Resultado de Consulta) *O resultado para uma consulta por padrão $Q = (P, N)$ é uma tripla $A_r = (J_r, S_r, R_r)$ tal que:*

$J_r = \{J_1, J_2, \cdots, J_N\}$ representa o conjunto das N imagens resultantes, ordenado por grau de similaridade decrescente.

$S_r = \{s_1, s_2, \cdots, s_N\}$ representa o conjunto dos valores de similaridade das imagens J_r tais que $s_i > s_j, \forall i \geq j$.

$R_r = \{R_1, R_2, \cdots, R_N\}$ representa o conjunto das regiões mais similares a cada padrão em cada imagem. Cada R_i é definido como $R_i = \{r_1^i, r_2^i, \cdots, r_{np}^i\}$ onde r_j^i é a região dentro da imagem J_i considerada pelo sistema como mais similar ao padrão $p_j \in P$. O resultado fundamental de uma consulta é o conjunto das imagens mais similares J_r. No entanto, este conjunto permite identificar em cada uma destas imagens quais as regiões que o sistema selecionou como mais similares a cada padrão.

4.3 Retroalimentação como mecanismo de refinamento

Como analisado no capítulo 3, o processo de recuperação de imagens por conteúdo tem uma natureza imprecisa, refletida nas medidas de similaridade. Esta incerteza está associada em grande parte à ausência de modelos matemáticos universais para modelar a noção humana de similaridade para características de conteúdo (como textura, cor ou forma). Este aspecto é ainda mais complexo quando tenta-se modelar elementos da semântica representada nas imagens. Como conseqüência, os resultados gerados pelos sistemas de recuperação por conteúdo possuem, geralmente, imagens fora do escopo procurado. Uma solução a este problema são os mecanismos de iteração do processo de consulta que permitam ao sistema refinar sucessivamente a busca. Propostas neste sentido incluem abordagens do tipo "exploratória"[69] ou com retroalimentação (*Feedback*)[65] como mencionado no capítulo 3.

Em geral, podemos descrever estes modelos de recuperação a partir de um processo iterativo de consulta e aproximação. A idéia básica é que a resposta a uma consulta seja utilizada para parametrizar uma próxima consulta, com a intervenção do usuário e a supervisão do sistema. O resultado de uma consulta por refinamento, no banco de dados, é obtido depois de uma sucessão de consultas $\{Q_0, \cdots, Q_n\}$, onde a consulta Q_i é parametrizada com informações do resultado da consulta anterior Q_{i-1}. Formalmente:

Definição 4.4 (Consulta por refinamento) *Seja $Q = \{Q_1, \cdots, Q_m\}$ um conjunto de consultas por conteúdo, $A = \{A_1, \cdots, A_m\}$ o conjunto de resultados a tais consultas e $P = \{P_1, \cdots, P_m\}$ o conjunto de parâmetros utilizados. Uma* **Consulta por refinamento** *é uma sucessão $\{(Q_1, A_1), \cdots, (Q_m, A_m)\}$ tal que:*

$$\begin{aligned} Q_1 &= \Psi() \\ Q_i &= \Psi(Q_{i-1}, A_{i-1}) \\ P_1 &= \Gamma(Q_1, \emptyset) \\ P_i &= \Gamma(Q_i, P_{i-1}) \\ A_i &= \chi(P_i) \end{aligned} \tag{4.1}$$

onde $Q_i \in \mathcal{Q}$ é uma consulta por conteúdo cujo resultado é $A_i \in \mathcal{A}$. $\mathcal{Q}$ é chamado de Espaço de Consultas e $\mathcal{A}$ de Espaço de Resultados. Neste processo $P_i \in \mathcal{P}$ é a parametrização da consulta Q_i sendo $\mathcal{P}$ chamado de Espaço de Parâmetros.

A consulta Q_1 é denominada **Consulta Inicial** *e A_m é chamado de* **Resultado Final** *da Consulta por Refinamento e corresponde a uma resposta aceitável para o usuário.*

$\Gamma : \{\mathcal{Q} \cup \emptyset\} \times \mathcal{P} \rightarrow \mathcal{P}$ é a função de parametrização, e $\Psi : \{\mathcal{Q} \cup \emptyset\} \times \mathcal{A} \rightarrow \mathcal{R}$ é a função de Avaliação de Consulta. $\chi : \mathcal{P} \rightarrow \mathcal{A}$ é a função de Execução da Consulta.

Esta definição, ilustrada na figura 4.2, descreve a aproximação do resultado final de uma consulta por refinamento.

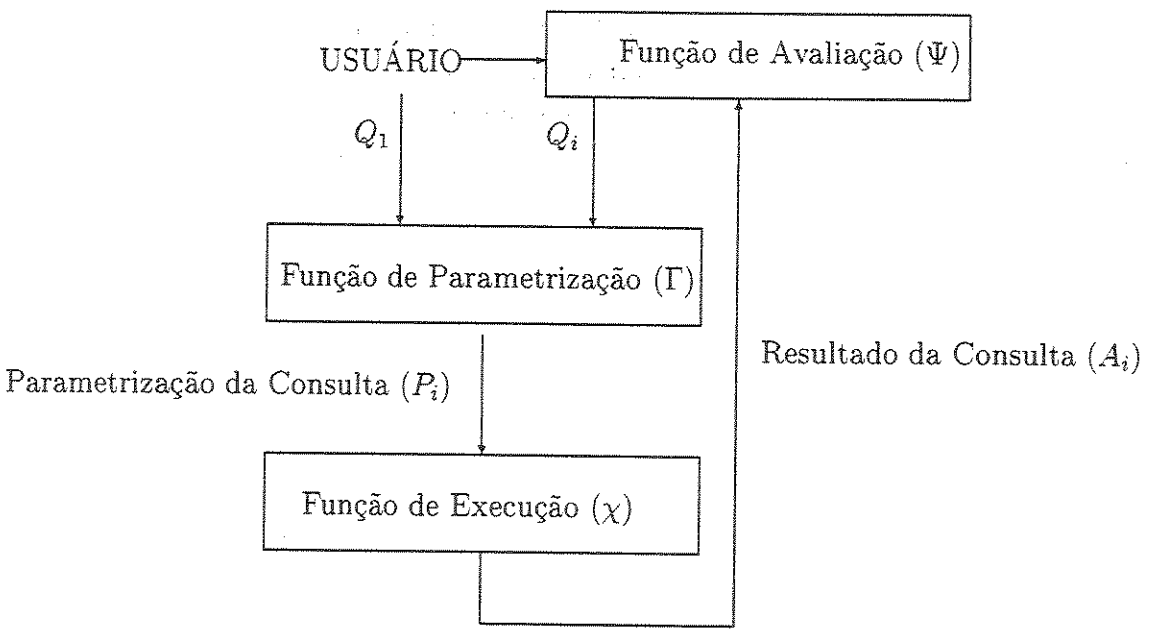


Figura 4.2: Consulta por Refinamento

O processo começa pela definição por parte do usuário de uma especificação inicial Q_1 do resultado que deseja obter. Sobre esta consulta, é aplicada a função Γ para gerar a

parametrização P_1, a qual é processada pela função de execução de consulta para obter o primeiro resultado de consulta A_1. A partir deste momento, repete-se um ciclo em que o usuário avalia subjetivamente o resultado A_{i-1} que junto com a consulta anterior (Q_{i-1}) permite gerar uma nova consulta Q_i.

A função Γ cria uma nova parametrização utilizando os parâmetros da consulta anterior P_{i-1} junto com a nova consulta Q_i. Esta parametrização é processada gerando um novo resultado A_i.

A consulta por refinamento se produz com a intervenção do usuário e do sistema. O refinamento por parte do usuário acontece na avaliação dos resultados das sucessivas recuperações, através da função Ψ para gerar uma nova consulta. No caso do sistema, o refinamento se produz no processo de parametrização.

Em vários destes sistemas a noção de similaridade é bastante precisa do ponto de vista humano, i.e. diferentes pessoas vão reconhecer com alto grau de concordância as mesmas imagens como similares. Nestes casos a dificuldade maior para o sistema está na necessidade de modelar essa similaridade e o recurso iterativo tenta uma aproximação com a resposta ideal.

No caso de imagens de sensoriamento remoto é maior a discordância entre especialistas em relação à similaridade, e podem acontecer mais facilmente diferenças entre as respostas relevantes de especialistas diferentes. Este aspecto pode ser entendido a partir da noção de padrão introduzida na seção anterior. Diferentes especialistas podem escolher um conjunto de amostras diferentes para representar um padrão, mesmo que em todos os casos estas representem um mesmo conceito do mundo real.

A retroalimentação é baseada na função Ψ de avaliação da recuperação. Na nossa proposta, esta função corresponde à confirmação, por parte do usuário, das regiões de interesse dentro das imagens resultantes selecionadas pelo sistema. Estas regiões são associadas a padrões pelo mecanismo de recuperação e o usuário deve avaliar se cada uma delas é aceitável ou não como exemplo desse padrão.

Como resultado do processo, aquelas regiões confirmadas positivamente pelo usuário são automaticamente incorporadas como novas amostras na definição de cada padrão, isto é, os padrões são redefinidos. Esta redefinição implica em que na nova iteração a consulta possa ser realizada com parâmetros reajustados para cada um dos diferentes padrões. A expectativa é que cada iteração produza resultados mais precisos e recupere regiões que aproximem mais da percepção do usuário sobre os padrões.

Na realidade, este processo não apenas exige refinamento por parte do sistema, mas também refinamento da noção de similaridade dos padrões por parte dos usuários. Cada especialista possui sua própria noção de similaridade para cada padrão e o registro permanente destas definições deve ser realizado para futuras consultas.

Este processo é ilustrado na figura 4.3. Cada resultado é um conjunto de imagens com

regiões contendo padrões (figura 4.3(a)). O usuário confirma/rejeita estas regiões (figura 4.3(b)) e o mecanismo de retroalimentação refina o conjunto de regiões do padrão (figura 4.3(c)), que é utilizado como entrada na consulta seguinte.

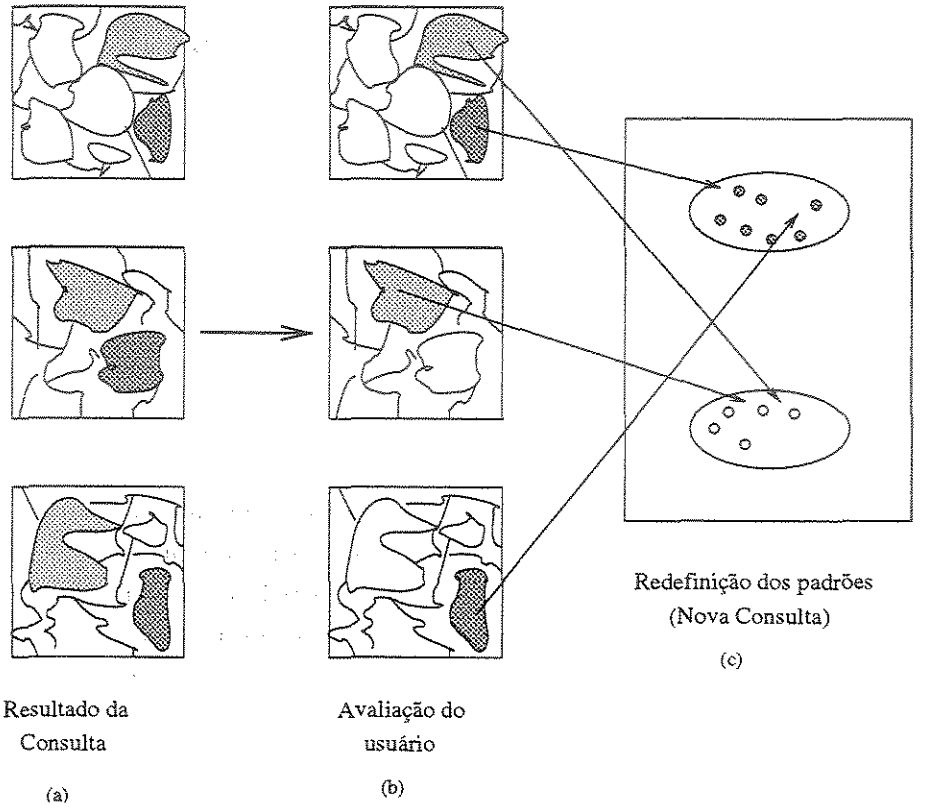


Figura 4.3: Processo de Retroalimentação. (a) Resultado da consulta: cada imagem com a região mais similar associada a cada padrão. (b) O usuário confirma as associações região-padrão que ele considera relevantes. (c) As regiões confirmadas pelo usuário são incorporadas à redefinição dos padrões.

4.4 Múltiplos modelos de representação do conteúdo

O capítulo 3 apresentou alguns modelos matemáticos usados em processamento de imagens e reconhecimento de padrões para textura e cor. Também constatou que não existem modelos matemáticos universais para a caracterização do conteúdo em ISR, especialmente textura e cor. Isto implica que determinados padrões de textura ou cor podem ser bem caracterizados por certos modelos mas pessimamente caracterizados por outros. Adici-

onalmente, modelos matemáticos considerados mais exatos [60] são os mais caros computacionalmente, acarretando restrições para seu uso em Sistemas de Recuperação por conteúdo que exigem tempo de resposta curto.

Por estas razões, propomos o uso simultâneo de vários modelos matemáticos. A intenção com o uso de vários modelos é permitir ao sistema avaliar o casamento entre a similaridade associada ao modelo e a noção de similaridade do usuário. O sistema pode então considerar mais confiável a resposta de um modelo para um determinado padrão e um determinado usuário. Como metáfora, poderíamos considerar que os modelos "cooperam"e competem para se adaptar à noção de similaridade de um usuário para um determinado padrão. Os modelos escolhidos devem ser preferencialmente simples e de baixo custo computacional.

Exemplos de modelos simples para textura são os modelos de Tamura [87] e a Transformada Wavelet[46, 11]. Para Cor, um modelo de representação simples é o uso de histograma [84, 85].

4.4.1 Definição de um modelo de representação de conteúdo

Os modelos matemáticos são usados para extrair informação que caracteriza o conteúdo associado à noção de distância conduzindo implicitamente uma noção de similaridade.

Nossa proposta introduz o conceito de *Modelo de Representação de Conteúdo* (MR). Este conceito descreve a maneira como os modelos matemáticos são utilizados. Como nosso modelo baseia-se na idéia de padrão, motivada pelo processo de classificação, exigimos que um modelo matemático inclua funções equivalentes àquelas descritas na seção 4.2.1 como parte do processo de classificação.

Definição 4.5 (Modelo de Representação de Conteúdo (MR)) *Um Modelo de Representação* $R =< E, T, s >$ *corresponde a um modelo matemático sobre o qual são definidas as funções:*

- **Uma função E de extração de características:** Esta função gera um *vetor característico* V_c, onde $V_c = E(r[J])$ ou $V_c = E(J)$. Esta função existe de maneira natural em todo modelo matemático que tenta modelar alguma característica de conteúdo de imagens.

  Por exemplo, no caso do Modelo de Histograma a função E corresponde ao algoritmo de construção do histograma e o resultado é o vetor representando os diferentes "bins" de distribuição de cores da imagem.

- **Uma função T de parametrização ou definição de classe:** Esta função processa um conjunto de regiões associadas a um mesmo padrão p_j e gera um descritor

desse padrão. este descritor geralmente aparece na forma de um *vetor de parametrização* V_p que caracteriza ou sintetiza um conteúdo comum a todas essas regiões: $V_p = T(r_1[p_j], r_2[p_j], \cdots, r_m[p_j])$

No exemplo de Histograma a função de parametrização pode ser o cálculo do histograma cuja distância média a todos histogramas dos diferentes segmentos processados é menor. Este vetor sintetiza a informação de cor das diferentes regiões.

- **Uma função s de distância :** Esta função estima a similaridade entre um vetor característico V_c e um vetor de parametrização V_p. Esta função gera um valor no intervalo $[0, 1]$ que estima a similaridade entre o conteúdo representado por ambos vetores, isto é, $s(V_c, V_p) \in [0, 1]$. O valor da função de distância deve estar no intervalo $[0, 1]$ para garantir uma equalização dos valores de similaridade. Esta equalização garante que os valores de similaridade dos diferentes MR sejam comparáveis, isto é. tenham aproximadamente o mesmo intervalo de valores. Em geral, esta função está associada a uma métrica.

  No exemplo de histograma, a métrica Euclidiana pode ser utilizada como função de distância.

4.4.2 Normalização das funções de distância

De modo geral, a função de distância s está baseada em uma métrica d aplicada sobre os vetores característicos. Uma métrica em geral pode ter valores em um domínio $[0, k]$ onde k eventualmente pode ser infinito. No entanto, como foi mencionado, a função de distância s se define no domínio $[0, 1]$, Esta propriedade garante o mesmo intervalo de valores de similaridade para qualquer MR e portanto a possibilidade de combinar seus valores. Por esta razão, é necessária a introdução de uma *função de normalização* $\tau : [0, k] \rightarrow [0, 1]$ tal que

$$s(V_c, V_p) = \tau(d(V_c, V_p)) \tag{4.2}$$

Note-se que quanto maior o valor de distância menor deve ser o valor de s representando menor similaridade. Um valor da distância (0) deve ser transformado em um valor de similaridade (1) e um valor de distância próximo de k deve gerar um valor se similaridade próximo de (0).

Existe um outro aspecto onde a normalização pode ser necessária. Vetores característicos são definidos sobre espaços multi-dimensionais e a distribuição de valores em cada uma de estas dimensões pode ser totalmente distinta. Se aplicadas métricas sobre estes vetores, a influência de cada dimensão no valor de distancia final pode ser diferente. Uma maneira de aliviar esta situação é aplicar funções de normalização em cada dimensão

para distribuir seus possíveis valores no intervalo $[0, 1]$. Esta transformação garante que cada dimensão influencie da mesma maneira o resultado de distância final.

Várias propostas de funções de normalização podem ser encontradas na literatura [2].

4.4.3 Modelos de Representação combinados com os padrões e retroalimentação

A introdução dos modelos de representação permite completar a definição de retroalimentação.

O sistema necessita processar os padrões considerando os diferentes MR. De fato, para cada modelo R_i a função de parametrização T_i é aplicada sobre cada padrão, obtendo-se diferentes *vetores de parametrização* que caracterizam este padrão.

A figura 4.4 mostra a especificação da consulta pelo usuário selecionando as diferentes regiões, envolvendo dois padrões. A seguir, as funções de parametrização dos diferentes MR são aplicadas, gerando-se os vetores de parametrização. Na figura, cada elipse corresponde a um padrão. Em (b) cada "versão"da elipse corresponde a um vetor de parametrização para um MR diferente do mesmo padrão. Assim. por exemplo, a elipse mais escura corresponde à aplicação do MR 1 obtendo-se da aplicação de T_1 para o padrão 1 (superior) e 2 (inferior).

Desta forma, uma consulta pode ser descrita a partir de um conjunto de P_C de *vetores de parametrização* associados a np diferentes padrões e nr modelos de representação, isto é,

$$P_C = [P_C[p_1], P_C[p_2], \cdots P_C[p_{np}]] \tag{4.3}$$

onde $P_C[p_j]$ representa o conjunto dos diferentes vetores de parametrização para o padrão p_j

$$P_C[p_j] = \{VR_1[j], VR_2[j], \cdots, VR_{nr}[j]\} \tag{4.4}$$

Cada $VR_i[j]$ corresponde ao vetor de parametrização para o padrão p_j no modelo R_i, ou seja, $VR_i[j] = T_i(p_j)$.

A retroalimentação refina os padrões, isto implica que as funções de parametrização T_i dos diferentes MR devem ser novamente aplicadas em cada iteração sobre as definições dos padrões.

Este processo é ilustrado na figura 4.5. O resultado de uma consulta é um conjunto de imagens com regiões contendo padrões (figura 4.5(a)). O usuário confirma/rejeita estas regiões (figura 4.5(b)) e o mecanismo de retroalimentação refina o conjunto de regiões do padrão (figura 4.5(c)), que é utilizado como entrada na seguinte consulta a partir da aplicação das funções de parametrização.

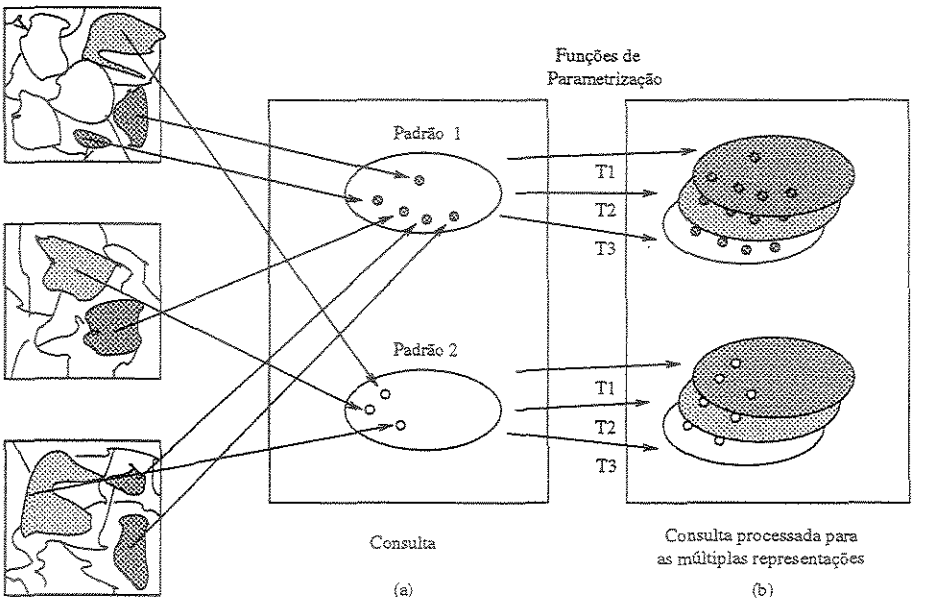


Figura 4.4: Definição de uma consulta. (a) Cada elipse corresponde a um padrão, criado a partir de amostras de diferentes imagens de consulta. (b) Cada padrão é processado para se obter seus diferentes vetores de parametrização.

O processo de retroalimentação deve permitir ao sistema avaliar o casamento de cada MR com o conceito de similaridade definida pelo usuário, ou seja, a relevância de cada MR para aquela consulta e usuário. A cada iteração, o sistema deve atribuir maior relevância aos valores de similaridade associados aos MR considerados mais significativos.

4.5 Inserção das imagens no Banco de Dados

A seção 4.2.2 definiu o padrão, constituído por um conjunto de regiões, como elemento básico de uma consulta. Por outro lado, o resultado esperado em uma consulta inclui as regiões similares a esses padrões. Estes elementos apontam à necessidade de inserir no banco de dados as regiões que compõem cada uma das imagens.

A proposta baseia-se no seguinte procedimento de inserção. Quando uma imagem J é incorporada ao Banco de Imagens, inicialmente é segmentada em $N(J)$ regiões homogêneas $\{r_1[J], r_2[J], \cdots, r_{N(J)}[J]\}$ segundo a cor e textura. A seguir, cada região segmentada $r_k[J]$ é processada utilizando cada uma das funções de extração de características E_i definidas para os diferentes MR. Como resultado, para cada região $r_k[J]$ é obtido um conjunto de vetores $\{r_k^1[J], r_k^2[J], \cdots, r_k^{nr}[J]\}$ onde $r_k^i[J] = E_i(r_k[J])\}$ descreve o conteúdo da região $r_k[J]$ segundo o MR R_i. Para cada MR é criada uma estrutura onde são armazenados os vetores característicos gerados por essa representação para cada uma das

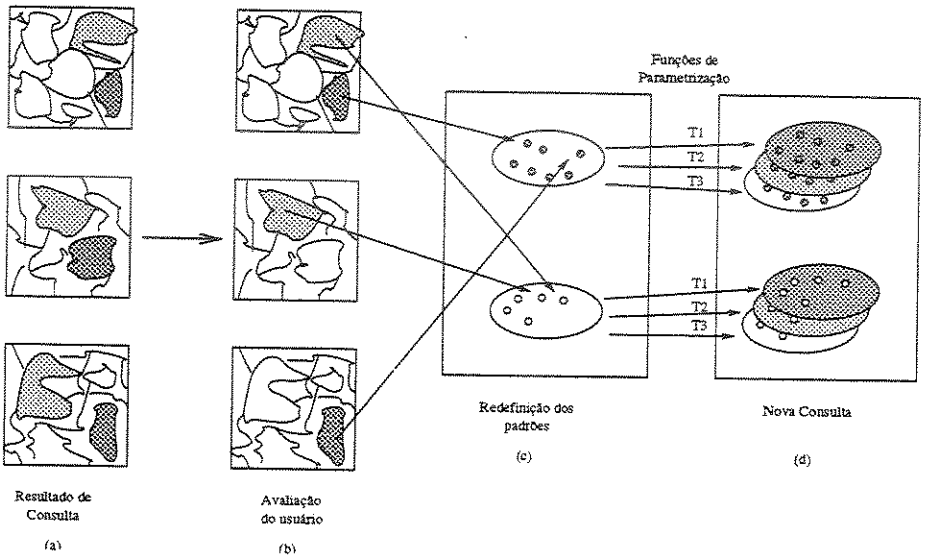


Figura 4.5: Processo de Retroalimentação. (a) Resultado da consulta: cada imagem tem uma região associada a cada padrão. (b) O usuário confirma as regiões associadas aos padrões que ele considera similares. (c) As regiões selecionadas pelo usuário são incorporadas à definição dos padrões. (d) São processados os "novos "padrões segundo as funções de parametrização dos múltiplos MR.

regiões de todas as imagens do Banco de Dados.

A figura 4.6 mostra este processo de segmentação e processamento. Os MR diferentes são ilustrados como retângulos, que correspondem às várias representações de uma mesma imagem e suas regiões.

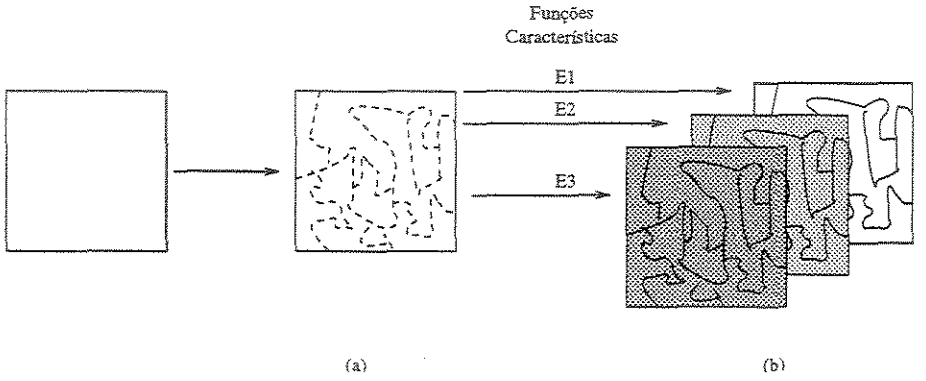


Figura 4.6: Inserção de uma imagem. (a) A imagem é inicialmente segmentada em regiões homogêneas segundo textura e cor. (b) Para cada região de cada imagem é gerado um vetor característico segundo cada MR.

4.6 Função de Similaridade

Um aspecto fundamental no processamento de uma consulta é a definição do critério de similaridade que será utilizado. Este critério deve considerar tanto as múltiplas representações como as definições de múltiplos padrões.

A similaridade de uma imagem J do banco de dados com a descrição de conteúdo especificada pela consulta P_C pode ser entendida como o resultado da aplicação de uma função $Sg(P_C, J)$ que devolve um valor de similaridade no intervalo $[0, 1]$.

Processar uma consulta corresponde a determinar quais as N imagens do banco de dados com maior valor da função Sg.

Para definir este valor de similaridade, é necessário estabelecer uma relação entre os padrões e a imagem. Esta relação deve considerar que cada padrão é descrito por várias representações e que cada imagem é definida por regiões segmentadas, representadas igualmente por múltiplas representações.

A figura 4.7 mostra os diferentes aspectos que a função Sg deve considerar para o cálculo da similaridade de uma imagem dada com uma consulta definida pelo usuário: os múltiplos padrões, as diferentes formas de representação e o conjunto de regiões associadas

a cada imagem. Cada tonalidade de cinza, na figura, corresponde a um MR e cada grupo de elipses corresponde a uma descrição de padrão. Neste caso a consulta consiste de dois padrões.

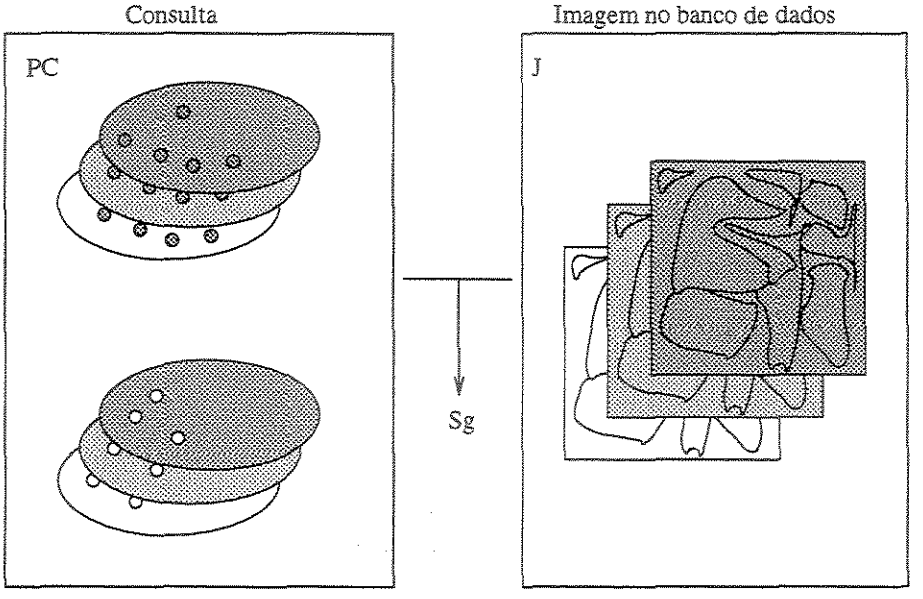


Figura 4.7: Modelo de similaridade: O cálculo da similaridade depende da relação entre os diferentes padrões em múltiplas representações (grupos de elipses) e os segmentos das imagens (regiões delimitadas dentro das imagens) também nas suas múltiplas representações.

O cálculo do valor de similaridade Sg de uma imagem do banco de dados em relação aos padrões de uma consulta pode ser entendido como um processo hierárquico em vários níveis, como ilustrado na figura 4.8. Este processo define uma *árvore de similaridade* cujos nós devem ser avaliados para o obter o resultado final de similaridade. No próximo capítulo será apresentada uma proposta de algoritmo de cálculo que reduz a complexidade deste cálculo.

O processo de cálculo da similaridade é dividido em duas etapas:

1. *Similaridade Intra-Padrão (Sp):* Esta função determina a similaridade entre uma imagem J do Banco de Dados e um padrão p_j descrito na consulta P_C como $P_c[p_j]$. Esta função considera em sua definição as diferentes regiões da imagem e os diferentes MR.

2. *Similaridade Global (Sg):* A função Sg combina todos os valores de similaridade Intra-Padrão em um único valor de similaridade global associado a toda a imagem.

As seções seguintes apresentam estas duas etapas detalhadamente.

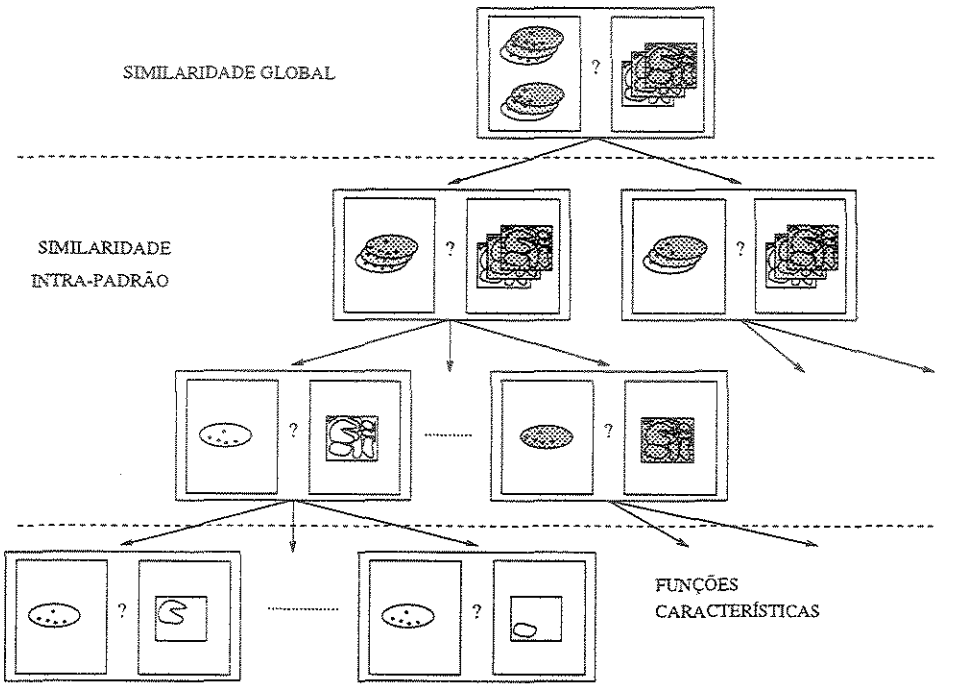


Figura 4.8: Arvore de cálculo de similaridade: Diferentes níveis no cálculo da similaridade de uma imagem com padrões de consulta. O cálculo é feito de maneira ascendente (*bottom-up*)

4.6.1 Função de Similaridade Global

A função de Similaridade Global determina o valor final de similaridade entre uma descrição de consulta P_C e uma imagem J.

Formalmente, a função Sg é definida como $Sg : \mathcal{P} \times \mathcal{J} \rightarrow [0,1]$ sendo $\mathcal{J}$ o conjunto das imagens no Banco de Dados e $\mathcal{P}$ o Espaço de Parâmetros definido na seção 4.3.

Este valor final de similaridade (global) combina as similaridades intra-padrão Sp de cada um dos padrões da consulta com a imagem, como apresentado na figura 4.8. Motivados pela abordagem apresentada por Fagin em [17] este cálculo é entendido como uma função Mg que combina em um único valor final os valores de similaridade intra-padrão:

$$Sg(P_C, J) = Mg\Big(Sp(P_C[p_1], J), \cdots, Sp(P_C[p_j], J), \cdots, Sp(P_C[p_{np}, J])\Big) \tag{4.5}$$

Na figura 4.8, a avaliação de cada similaridade intra-padrão $Sp(P_C[p_j], J)$ representa o relacionamento entre o padrão p_j, descrito por um conjunto de elipses, com a imagem descrita pelas diferentes representações e regiões.

A função Mg adotada neste trabalho será apresentada na seção 4.7.1.

4.6.2 Função de Similaridade Intra-Padrão

A função Sp determina a similaridade de um padrão com a imagem. Esta similaridade pode ser definida igualmente por uma função Mp que combina em um único valor as similaridades entre as diferentes representações de um padrão e as diferentes representações e regiões associadas a uma imagem.

Formalmente, a função Sp é definida como $Sp : \mathcal{P} \times \mathcal{J} \rightarrow [0, 1]$, sendo sendo $\mathcal{J}$ o conjunto das imagens no Banco de Dados e $\mathcal{P}$ o Espaço de Parâmetros de um MR específico.

$$\begin{aligned} Sp(P_C[p_j], J) = Mp\Big(& s_1(r_1^1[J], VR_1[j]), \cdots, s_1(r_{N(J)}^1[J], VR_1[j]), \\ & s_2(r_1^2[J], VR_2[j]), \cdots, s_2(r_{N(J)}^2[J], VR_2[j]), \\ & \cdots \\ & s_{nr}(r_{nr}^1[J], VR_{nr}[j]), \cdots, s_{nr}(r_{N(J)}^{nr}[J], VR_{nr}[j])\Big) \end{aligned} \tag{4.6}$$

onde $N(J)$ é o número de regiões da imagem J.

$r_m^i[J]$ corresponde ao i-ésimo vetor característico da região r_m da imagem J, isto é, a descrição do conteúdo da região segundo o MR R_i. Estes vetores são gerados e armazenados no momento da inserção da imagem J no banco de dados, como descrito na seção 4.5.

A função Mp adotada neste trabalho será apresentada na seção 4.7.2.

Note-se que $VR_i[j]$ corresponde ao vetor de parametrização do padrão p_j segundo o MR R_i. Este outro nível de cálculo pode ser considerado como uma sub-árvore que representa a integração de todas as possíveis combinações de regiões da imagem e de representações, como apresentado na figura 4.8.

A determinação da similaridade entre um padrão e as regiões da imagem, descrita na figura 4.7, foi incorporada à árvore, agora considerando vários padrões.

O nível de cálculo inicial da função de similaridade está na avaliação de $s_i(r_m^i[J], VR_i[j])$ que determina o valor de similaridade da região r_m da imagem J com o padrão p_j segundo

o MR R_i. Note-se que s_i corresponde à função de distância de R_i como definido na seção 4.4.1

A avaliação de $s_i(r_m^i[I], VR_i[j])$ pode ser entendida como uma função de classificação que estima o grau de pertinência do vetor $r_m^i[J]$ à classe descrita pelo vetor de parametrização $VR_i[j]$, para o modelo R_i.

Da análise anterior, infere-se que o cálculo da similaridade de uma imagem implica em uma procura em toda uma árvore de diferentes padrões, representações e segmentos como apresentado na figura 4.8. Como veremos no capítulo 5, o cálculo pode ser reduzido a partir da poda de sub-árvores desta árvore.

4.7 Modelo de similaridade e retroalimentação adotados

As funções Mg e Mp estabelecem, na realidade, condições sobre as similaridades intra-padrão e funções de distância, respectivamente. Um abordagem natural seria descrever estas condições como predicados No entanto, a lógica booleana não pode ser utilizada para expressar estes predicados, dado que o valor de similaridade não é booleano, mas um número real. Portanto, outros modelos são necessários.

Como metáfora para a compreensão de proposta para estas funções, considera-se a existência de um conjunto de n "juízes"ou votantes $\{V_1, \cdots, V_n\}$ que realizam uma apreciação sobre algum aspecto e uma tupla X em um banco de dados. Não necessariamente todos os juizes julgam o mesmo aspecto. Posteriormente, todos estas avaliações são integradas em um valor final de acordo com um determinado critério ou função. Exemplos típicos desta abordagem são os sistemas de avaliação em esportes como ginástica ou salto ornamental. Fagin em [20] define o conceito de **função de pontuação** (*Scoring function*) para denominar este tipo de função. Uma **função de pontuação** $f(x_1, x_2, \cdots, x_n)$ calcula um resultado final associado a uma tupla X como integração dos valores $\{x_i\}$ que são as avaliações dos respectivos $\{V_i\}$.

Este modelo pode ser adaptado ao nosso problema considerando Mg e Mp como funções de pontuação. No caso, uma tupla corresponderia a uma imagem.

Para a função de similaridade intra-padrão , cada x_i pode ser interpretado como a avaliação da similaridade da imagem com um padrão segundo a "visão"de um determinado MR. Cada MR pode ser interpretado como um "juiz"diferente. A função Mp combina todos estes resultados em uma pontuação final para esse padrão.

No caso da função de similaridade global, cada x_i pode ser interpretado como a avaliação da similaridade desta imagem com um padrão. A função de pontuação Mg combina estes resultados em uma pontuação final para a imagem. Neste caso estaríamos interpre-

tando a avaliação de cada padrão como sendo realizada por um "juiz" diferente.

O conceito de função de pontuação é suficientemente geral para considerar sua utilização em vários modelos de avaliação numéricos, em particular lógica *fuzzy*. No modelo de lógica *fuzzy*, um predicado corresponde a uma expressão que combina vários operadores *fuzzy* e cujo resultado é um valor entre 0 e 1 que define o grau de pertinência de um elemento a um conjunto, isto é, um grau de verdade. No caso de uma tupla X em um banco de dados, uma formula em lógica *fuzzy* pode ser transformada em uma função de pontuação que devolve um valor *fuzzy* para essa tupla.

4.7.1 Função de Similaridade Global adotada

A função de pontuação Mg estabelece a integração dos valores das similaridades intra-padrão. Se analisado do ponto de vista da lógica booleana, a forma mais natural desta integração seria uma conjunção, isto é, exigir que a imagem satisfaça simultaneamente a condição de similaridade para todos os padrões. O operador de conjunção *fuzzy* é definido como o mínimo entre todos os termos da conjunção. Em nosso caso, corresponderia ao valor mínimo entre todas as similaridades intra-padrão, isto é,

$$\begin{aligned} Mg\Big(& Sp(P_C[p_1], J) \cdots, Sp(P_C[p_j], J), \cdots, Sp(P_C[p_{np}], J)\Big) \\ & = Sp(P_C[p_1], J) \cdots, \wedge Sp(P_C[p_j], J), \cdots, \wedge Sp(P_C[p_{np}], J) \\ & = \min_{1 \leq j \leq np}\{Sp(P_C[p_j], J)\} \end{aligned} \tag{4.7}$$

Embora este seja o modelo que adotaremos para a Similaridade Global, outros modelos poderiam ser aplicados. Por exemplo, em um modelo probabilístico, a Similaridade Global poderia ser considerada como a probabilidade de que uma imagem seja similar à consulta. Considerando que essa probabilidade depende da probabilidade de a imagem ter regiões similares em todos os padrões, então a Similaridade Global seria o produto de todas as probabilidades, isto é,

$$Sg(P_C, J) = \prod_{1 \leq j \leq np} (Sp(P_C[p_j], J)) \tag{4.8}$$

4.7.2 Função de similaridade Intra-Padrão adotada

No caso da Similaridade Intra-Padrão várias abordagens podem ser utilizadas para definir Mp, combinando as diferentes similaridades envolvidas. No entanto, continuando com a analogia com a lógica booleana, a forma natural a adotar seria uma disjunção, isto é, considerar que é suficiente que uma região da imagem em qualquer uma das representações seja similar para considerar a imagem similar. Trasladado à lógica *fuzzy* adotada, seria

aplicado o operador de disjunção *fuzzy* que é definido como o máximo entre todos os valores $s_i(r_k^i[J], VR_j[i])$, isto é,

$$\begin{aligned} Mp\Big(& s_1(r_1^1[J], VR_1[j]), \cdots, s_1(r_{N(J)}^1[J], VR_1[j]), \\ & s_2(r_1^2[J], VR_2[j]), \cdots, s_2(r_{N(J)}^2[J], VR_2[j]), \\ & \cdots \\ & s_{nr}(r_1^{nr}[J], VR_{nr}[j]), \cdots, s_{nr}(r_{N(J)}^{nr}[J], VR_{nr}[j])\Big) \\ = & \\ & s_1(r_1^1[J], VR_1[j]) \vee \cdots \vee s_1(r_{N(J)}^1[J], VR_1[j]) \\ & s_2(r_1^2[J], VR_2[j]) \vee \cdots \vee s_2(r_{N(J)}^2[J], VR_2[j]) \\ & \cdots \\ & s_{nr}(r_1^{nr}[J], VR_{nr}[j]) \vee \cdots \vee s_{nr}(r_{N(J)}^{nr}[J], VR_{nr}[j]) \\ = & \max_{1 \leq i \leq nr, 1 \leq k \leq N(J)} \{s_i(r_k^i[J], VR_i[j])\} \end{aligned} \tag{4.9}$$

Considerando a análise de cada imagem como uma função independente, podemos redefinir esta equação da seguinte forma:

$$\begin{aligned} Mp & \Big(SimIm_1(VR_1[j], J), \cdots, SimIm_{nr}(VR_{nr}[j], J)\Big) \\ = & \; SimIm_1(VR_1[j], J) \vee \cdots \vee SimIm_{nr}(VR_{nr}[j], J) \\ = & \max_{1 \leq i \leq nr} \{SimIm_i(VR_i[j], J)\} \end{aligned} \tag{4.10}$$

onde o elemento $SimIm_i(VR_i[j], J)$ é definido como:

$$SimIm_i(VR_i[j], J) = \max_{1 \leq k \leq N(J)} \{s_i(r_k^i[J], VR_i[j])\} \tag{4.11}$$

4.7.3 Inclusão de pesos na similaridade

Como apresentado anteriormente, o processo de retroalimentação deve provocar modificações no cálculo das similaridades em cada iteração.

Como parte do cálculo da similaridade, o sistema estima quantitativamente a capacidade ou relevância de cada MR para modelar o senso de percepção do usuário para cada padrão p_j. Esta relevância é representada por pesos (θ_j^i) associados a cada padrão p_j para o MR R_i

Nesta seção é apresentada a introdução destes pesos na função de Similaridade Intra-Padrão.

Abordagem de Fagin A extensão de uma função de pontuação para incluir pesos pode ser vista como a associação de pesos a cada termo, na expressão original da função.

Em outras palavras, dada uma função de pontuação $f(x_1, x_2, \cdots, x_n)$, a inclusão de pesos $\theta_1, \theta_2, \cdots, \theta_n$ é realizada como $f_{(\theta_1,\theta_2,\cdots,\theta_n)}(x_1, x_2, \cdots, x_n = f(\theta_1 x_1, \theta_2 x_2, \cdots, \theta_n x_n)$. Como condição adicional é razoável exigir que $\sum_i \theta_i = 1$.

No entanto, esta extensão não garante certas propriedades desejáveis na função resultante. Por exemplo, considerando a função de pontuação *min*, a função estendida segundo o raciocínio anterior ficaria:

$$\min_{(\theta_1,\theta_2,\cdots,\theta_n)}(x_1, x_2, \cdots, x_n = min(\theta_1 x_1, \theta_2 x_2, \cdots, \theta_n x_n) \tag{4.12}$$

Seja, por exemplo, a existência de dois termos na função $\min_{\theta_1,\theta_2}(x_1, x_2) = min(\theta_1 x_1, \theta_2 x_2)$. Consideremos então o comportamento da função em relação ao peso θ_1. Quando seu valor é 0, o resultado final da função $\min_{0,1}(x_1, x_2) = x_1$. Por outro lado, quando o valor de θ_1 é 1, o resultado seria $\min_{1,0}(x_1, x_2) = x_2$. Finalmente no caso em que o valor dos dois pesos é o mesmo (1/2) é razoável considerar que os pesos não tenham nenhum impacto, e portanto, a função com peso deve comportar-se como a função original, isto é: $\min_{\frac{1}{2},\frac{1}{2}}(x_1, x_2) = min(x_1, x_2)$ mas neste caso não se garante esta propriedade pois o resultado obtido é $\frac{1}{2} min(x_1, x_2)$

Outro problema surge em outras valorações dos pesos, onde nem sempre é possível um comportamento contínuo da função, isto é, com uma pequena variação do peso acontecem saltos no seu resultado.

Considerando esta situação, Fagin propõe em [20] um método para transformar uma função de pontuação sem pesos em uma função com pesos (*Weighted Scoring Function*). Este método baseia-se nas seguintes definições e teorema.

- Dizemos que uma função de pontuação com pesos $f_{(\theta_1,\theta_2,\cdots,\theta_n)}$ *é baseada* em uma função de pontuação sem pesos f se $f_{(\frac{1}{2},\cdots,\frac{1}{2})}(x_1, x_2, \cdots, x_n) = f(x_1, x_2, \cdots, x_n)$.
- Dizemos que uma função de pontuação com pesos $f_{(\theta_1,\theta_2,\cdots,\theta_n)}$ *é compatível* se $f_{(\theta_1,\theta_2,\cdots,\theta_{n-1},0)}(x_1, x_2, \cdots, x_n) = f_{(\theta_1,\theta_2,\cdots,\theta_{n-1})}(x_1, x_2, \cdots, x_{n-1})$

Teorema 4.1 (Função de Pontuação com pesos(Fagin) [20]) *Para toda função de pontuação sem pesos f existe uma função de pontuação com pesos $f_{(\theta_1,\theta_2,\cdots,\theta_n)}$, compatível, baseada em f e contínua em $(\theta_1, \theta_2, \cdots, \theta_n)$. Se $I = \{1, 2, \cdots, n\}$ e X e Θ são índices sobre I tais que $\theta_1 \geq \theta_2, \geq \cdots, \geq \theta_n$ então a função de pontuação com pesos tem a forma:*

$$\begin{aligned} f_\Theta(X) = \; & (\theta_1 - \theta_2) \cdot f(x_1) + \\ & 2 \cdot (\theta_2 - \theta_3) \cdot f(x_1, x_2) + \\ & 3 \cdot (\theta_3 - \theta_4) \cdot f(x_1, x_2, x_3) + \\ & \cdots + \\ & n \cdot \theta_m \cdot f(x_1, x_2, \cdots, x_n) \end{aligned} \tag{4.13}$$

Este teorema permite transformar facilmente uma função de pontuação sem pesos em uma equivalente com pesos.

Aplicação da abordagem ao contexto da tese

Consideremos que $\Theta_j = \{\theta_j^1, \theta_j^2, \cdots, \theta_j^{nr})\}$ corresponda ao conjunto de pesos para os diferentes MR em relação ao padrão p_j.

Aplicando o teorema 4.1, a Similaridade Intra-Padrão Sp apresentada em 4.7.2 pode ser transformada em:

$$\begin{aligned} Sp_{\Theta_j}(P_C[p_j], J) = \quad & Mp_{\Theta_j}\Big(\quad SimIm_1(VR_1[j], J), \\ & \qquad SimIm_2(VR_2[j], J), \\ & \qquad \cdots, \\ & \qquad SimIm_{nr}(VR_{nr}[j], J)\Big) \\ = \quad & \\ & (\theta_j^1 - \theta_j^2) \cdot SimIm_1(VR_1[j], J) + \\ & 2 \cdot (\theta_j^2 - \theta_j^3) \cdot max_{1 \leq i \leq 2}\{SimIm_i(VR_i[j], J)\} + \\ & \cdots + \\ & nr \cdot \theta_j^{nr} \cdot max_{1 \leq i \leq nr}\{SimIm_i(VR_i[j], J)\} \end{aligned} \tag{4.14}$$

onde θ_j^i representa o peso que tem a representação R_i para o padrão p_j.

Note-se que cada peso está associado a um termo que corresponde ao maior valor de similaridade dentre todas as regiões da imagem segundo a representação, isto é, não estamos associando pesos à similaridade de cada região da imagem mas aos diferentes MR.

A equação 4.14 permite associar um valor de similaridade a uma imagem J do banco de dados.

4.7.4 Mecanismo de Retroalimentação

O mecanismo de retroalimentação origina-se da avaliação pelo usuário das regiões associadas a cada padrão nas imagens resultantes da consulta. Esta avaliação modifica parâmetros associados à maneira em que os MR e os padrões intervêm na função de similaridade.

Estes parâmetros são de dois tipos:

- *Os pesos associados a cada padrão e cada forma de representação* (θ_j^i). O sistema deve determinar automaticamente a relevância que cada forma de representação tem

para o cálculo da Similaridade Intra-Padrão de cada um dos padrões. A forma com que os pesos intervêm nesta similaridade já foi abordada anteriormente.

- *Os vetores de parametrização $VR_i[j]$. Para cada padrão e cada forma de representação é gerado um vetor de parametrização $VR_i[j]$. Cada iteração do processo de consulta implica na redefinição do conjunto de regiões que descreve um padrão e portanto, na redefinição destes vetores. Estes vetores participam na definição função de similaridade através da avaliação das funções de distância de cada MR* $s_i(r_m^i[J], VR_i[j])$. Os valores da avaliação destas funções em cada iteração são diferentes e modificam o resultado final da função de similaridade.

Com a mudança destes parâmetros, a função de similaridade é redefinida em cada iteração permitindo o processo de refinamento da consulta. O cálculo dos vetores de parametrização em cada iteração é resultado da definição de das funções de parametrização T_i de cada modelo R_i. No entanto, no caso dos pesos é necessário um critério para seu cálculo. A seção seguinte apresenta detalhadamente com este cálculo é realizado.

4.7.5 Cálculo dos pesos

O processo de cálculo dos pesos θ_j^i tem como objetivo fundamental dar maior relevância àqueles MR que melhor representem a noção de similaridade do usuário responsável pela consulta. Por esta razão, deve ser estabelecida uma relação entre as melhores regiões para um MR, segundo um padrão, e as melhores regiões para um usuário, segundo este mesmo padrão. A partir desta relação, deve ser definido o peso θ_j^i do MR R_i para o padrão p_j.

Como se pode deduzir da expressão de similaridade intra-padrão (Equação 4.14), as N melhores regiões para cada MR R_i, em cada padrão p_j, são aquelas com maior valor de similaridade $s_i(r_m^i[J], VR_i[j])$. No entanto, estas regiões não necessariamente são as que têm o maior valor de similaridade intra-padrão. Isto quer dizer que as melhores regiões para cada MR não têm de ser necessariamente aquelas escolhidas pelo sistema na resposta a uma consulta.

Definamos os seguintes conjuntos:

- Seja RM_j^i o conjunto das N regiões de maior similaridade para o MR R_i para o padrão p_j, isto é, as de maior valor de $s_i(r_m^i[J], VR_i[j])$.

- Seja RS_j o conjunto das N regiões de maior valor de $Sp_{\Theta_j}(P_C[p_j], J)$ e portanto na resposta do sistema considerando o padrão p_j isoladamente.

- Seja RU_j o conjunto das regiões de maior similaridade com o padrão p_j selecionadas pelo usuário.

- Seja $M_j^i = RM_j^i \cap RS_j$ o conjunto das regiões de maior similaridade para o MR R_i e o padrão p_j que também pertencem ao conjunto de maior similaridade intra-padrão para p_j, isto é, aquelas imagens incluídas na resposta do sistema para o padrão p_j que estão no conjunto das mais similares para o MR R_i.

A partir destes conjuntos, podemos estabelecer uma medida da relevância de um determinado MR no conjunto das regiões de maior similaridade intra-padrão para um p_j determinado. Para isto é definido o conceito de compatibilidade do sistema.

Definição 4.6 (Compatibilidade do sistema) *Denominamos* **compatibilidade do sistema com o MR R_i para o padrão p_j** *a seguinte relação:*

$$CS_j^i = \frac{|M_j^i|}{N} \tag{4.15}$$

isto é, a proporção entre o número de regiões de maior similaridade para o MR R_i que formam parte do conjunto do resultado para p_j (M_j^i) em relação ao número total de regiões desse resultado (RS_j) que sempre é N.

CS_j^i pode assumir valores discretos no conjunto $\{0, \frac{1}{N}, \frac{2}{N}, \cdots, 1\}$. Um valor 1 indica que todas as regiões no conjunto das mais similares com o padrão p_j, segundo o MR R_i, estão incluídas na resposta do sistema para o padrão p_j. um valor 0 indica que nenhuma das regiões neste conjunto estão incluídas na resposta. Assim, este valor CS_j^i descreve quantitativamente a relevância que tem o MR R_i no resultado da similaridade intra-padrão.

O fato que CS_j^i pode ter valores menores que 1 impõe uma restrição ao confronto deste MR com o usuário. O usuário confirma apenas aquelas regiões que o sistema lhe apresentou e a partir dessa confirmação devem ser redefinidos os pesos para cada MR. No pior dos casos, quando $CS_j^i = 0$, o usuário não tem como avaliar a relevância do MR R_i pois a seleção da similaridade intra-padrão não lhe oferece regiões que permitam realizar esta avaliação.

Esta situação define uma primeira premissa na proposta de pesos: um MR deve ser avaliado em relação à confirmação pelo usuário das imagens por ele escolhidas , sem prejuízo de ter sido menos favorecido pela função de similaridade intra-padrão.

A partir desta premissa, definimos uma medida de quão relevante é um determinado MR para o usuário.

Definição 4.7 (Compatibilidade do usuário) *Denominamos de* **compatibilidade do usuário com o MR R_i para o padrão p_j** *a seguinte relação:*

$$CU_j^i = \frac{|RU_j \cap RM_j^i|}{|M_j^i|} \tag{4.16}$$

isto é, a proporção entre o número de regiões confirmadas pelo usuário para o padrão p_j, dentre o conjunto das mais similares segundo o MR R_i para o padrão p_j, em relação ao número total de regiões selecionadas pelo MR R_i que **o usuário está em condições de avaliar**, *isto é que fazem parte da resposta do sistema.*

CU_j^i pode assumir valores discretos no conjunto $\{0, \frac{1}{|M_j^i|}, \frac{2}{|M_j^i|}, \cdots, 1\}$

Fazendo a mesma análise que no caso anterior, um valor 1 indica que o usuário confirmou todas as regiões selecionadas pelo MR R_i fazendo parte da resposta. Por outro lado, um valor 0 indica que o usuário não confirmou nenhuma das regiões de RM_j^i na resposta. O valor de CU_j^i é um indicador quantitativo do grau de relevância do MR R_i, nas confirmações realizadas pelo usuário.

Quando o conjunto M_j^i é vazio, então a resposta do sistema para o padrão p_j não contem nenhuma das imagem incluídas no conjunto das mais similares segundo o MR R_i. Neste caso a compatibilidade do usuário é indefinida.

Partindo da definição destas compatibilidades a relação:

$$T_j^i = \frac{CU_j^i}{CS_j^i} \tag{4.17}$$

define um critério de quando e quanto aumentar ou diminuir cada peso.

Note-se que teoricamente T_j^i pode assumir valores no conjunto discreto dado por $\{0, \frac{1}{N}, \frac{2}{N}, \cdots, 1, 2, \cdots, N\}$, sendo N o número de imagens requisitadas na consulta.

A condição $T_j^i > 1$ indica que a compatibilidade do usuário com o MR é maior que a compatibilidade do sistema com o mesmo MR. Esta condição representa um critério para que na próxima iteração a função de similaridade intra-padrão considere um aumento na relevância desse MR.

No caso contrario, a condição $T_j^i < 1$ indica que a compatibilidade com o sistema é maior que a compatibilidade com o usuário e, portanto, é uma medida de que na próxima iteração a função de similaridade intra-padrão deve diminuir a relevância desse MR. O caso em que $T_j^i = 1$ indica que a relevância para o sistema e para o usuário é a mesma e, portanto, o peso está correto.

O próprio valor de T_j^i é uma boa medida da magnitude em que o peso deve ser alterado. O valor de T_j^i , além de indicar que o peso deve aumentar ou diminuir, indica em que magnitude o usuário considera esse MR melhor ou pior do que o sistema considera e, portanto, em que proporção o valor do peso deve ser alterado.O maior aumento do peso T_j^i acontece quando $T_j^i = N$ e o menor quando $T_j^i = 0$

Finalmente, como mencionado, a compatibilidade do usuário (equação 4.16) pode ser indefinida implicando também a indefinição da equação de T_j^i. Como foi analisado anteriormente, esta situação corresponde ao caso em que a resposta da função de similaridade

intra-padrão não inclui nenhuma das regiões da resposta do MR R_i. Este aspecto obviamente deve ser considerado. Nossa proposta, neste caso, é considerar que $T_j^i = 1$, isto é, como nesta iteração não existe critério para avaliar o MR R_i, então ele não deve sofrer nenhuma alteração na sua relevância.

A partir da definição de T_j^i, uma primeira forma de definir os pesos pode ser a normalização de cada T_j^i em relação à soma de todos eles, isto é,

$$\theta_j^i = \frac{T_j^i}{\sum T_j^i} \tag{4.18}$$

Desta maneira o peso θ_j^i de cada MR é estabelecido de acordo com a proporção do valor do seu T_j^i associado em relação ao resto dos MR. No entanto, esta forma de definição dos pesos não considera o valor destes na iteração anterior, isto é, o histórico dos pesos. Desta maneira, este critério estaria considerando unicamente o resultado da última iteração.

Um dos pressupostos do nosso trabalho é considerar uma consulta como um processo de aproximação sucessiva do resultado esperado pelo usuário. A partir deste pressuposto, definimos uma segunda premissa: os novos pesos devem ser calculados a partir da correção dos pesos anteriores, isto é, a história acumulada nos pesos até a iteração anterior deve ser considerada no novo cálculo.

Para garantir esta premissa, definimos:

$$\delta_j^i = \theta_j^i + T_j^i \cdot \theta_j^i \tag{4.19}$$

em que δ_j^i representa o valor do peso θ_j^i corrigido com o valor de T_j^i, mas proporcionalmente ao seu próprio valor.

Como passo final, realizamos a normalização dos valores de δ_j^i para obter os pesos θ_j^i para uma nova iteração:

$$\theta_j^i = \frac{\delta_j^i}{\sum \delta_j^i} \tag{4.20}$$

Esta abordagem garante uma correção gradual dos pesos, permitindo, porém, uma correção significativa no caso de um valor de T_j^i muito relevante, em relação ao resto dos MR.

4.8 Parâmetros de precisão do modelo

De maneira ideal, o modelo proposto nas seções anteriores, deve permitir o refinamento sucessivo da consulta e da resposta do sistema, e concluir com o melhor resultado que o sistema possa oferecer ao usuário. No entanto, determinados fatores contribuem para que esta melhor resposta não seja garantida. Alguns destes fatores são:

- *Falta de um critério de finalização da consulta:*

  O processo de consulta por refinamento proposto deve considerar dois cenários possíveis de conclusão: (1) o resultado obtido pelo usuário está de acordo com suas expectativas e ele finaliza a consulta; (2) o resultado obtido pelo usuário é insatisfatório e o usuário desiste de refiná-lo. Em ambos os casos, existe a possibilidade de que próximas iterações consigam devolver um resultado mais preciso. No entanto, o usuário não tem um critério para decidir se deve continuar ou não o processo de consulta.

- *Usuários inconsistentes:*

  A definição de um padrão é totalmente dependente do usuário, tanto na sua definição inicial como no seu refinamento ao longo do processo de consulta. Um usuário que não mantém coerência na seleção de regiões pertencentes a um padrão induz o sistema a respostas imprecisas.

- *Modelos de Representação inadequados:*

  Alguns MR utilizados pelo sistema ou a combinação deles podem ser inadequados para representar a noção de similaridade do usuário. Esses MR não contribuem para a determinação do resultado e eventualmente podem ser removidos do cálculo da similaridade.

Estes fatores, que afetam o resultado de uma consulta, não são possíveis de ser evitados, sobretudo se consideramos que os dois primeiros casos dependem totalmente de decisões subjetivas. No entanto, é possível para o sistema estabelecer medidas que permitam identificar estas situações e indicá-las ao usuário.

Consideremos o processo de consulta como uma aproximação do modelo de similaridade do sistema ao critério de similaridade do usuário. Em cada iteração da consulta, o usuário deve confirmar os resultados a ele apresentados Como conseqüência, os vetores de parametrização $VR_i[j]$ e os pesos θ_j^i são redefinidos. Se em duas iterações sucessivas as mudanças produzidas no processo de retroalimentação são pequenas, podemos interpretar que o sistema e o usuário estão convergindo em suas valorações. Se consideramos a evolução destes valores em iterações sucessivas, é possível identificar alguns comportamentos específicos à consulta.

Se todos os valores de $\{T_j^i\}$ são muito próximos de 1, isto indica que o usuário e o sistema chegaram a uma convergência em relação à importância de cada MR para cada padrão, isto é, em relação aos pesos. Esta situação indica que a retroalimentação do usuário não está contribuindo mais para o refinamento dos pesos e que este atingiu uma situação de estabilidade.

De maneira análoga, suponhamos que a similaridade entre os vetores de parametrização, $VR_i[j]'$ e $VR_i[j]''$, de duas iterações sucessivas associados a um MR, atinge um valor muito alto, isto é, muito próximo a 1. Esta situação indica que a retroalimentação do usuário não está contribuindo muito para a definição do padrão p_j no MR R_i e, igualmente, é sinal de uma estabilização na definição do padrão para o usuário.

Se ambos os processos de estabilidade ocorrem para todos os MR então isto indica a convergência entre o modelo de similaridade do sistema e a noção de similaridade do usuário em todos os MR. Como conseqüência, o sistema não vai conseguir melhorar os resultados da consulta. Diante de uma tal situação, um usuário pode reiniciar a consulta redefinindo sua percepção sobre os padrões, ou aceitar a última resposta do sistema.

Por outro lado, é possível que durante o processo de consulta nenhum MR consiga atingir a estabilidade dos pesos e dos padrões. Esta situação pode ser um indicador de que o usuário não está sendo coerente na definição dos padrões ou que nenhum MR é adequado para a noção de similaridade desse usuário.

Para o caso da estabilidade em relação à definição dos padrões, nossa proposta é oferecer ao usuário os valores mínimos e máximos das diferenças entre os vetores de parametrização das duas últimas iterações. Chamaremos estes valores de indicador de estabilidade dos padrões.

O **Indicador de estabilidade dos padrões** IP é definido como um par de termos $IP = (IP_m, IP_M)$ onde $IP_m = \min\{s_i(VR_i[j]', VR_i[j]'')\}$ e
$IP_M = \max\{s_i(VR_i[j]', VR_i[j]'')\}$ sendo $VR_i[j]'$ e $VR_i[j]''$ os vetores de parametrização de duas iterações sucessivas para o MR R_i e o padrão p_j. s_i corresponde à função de distância do MR R_i e calcula a similaridade entre estes vetores. O indicador tem valores no intervalo $[0, 1]$ satisfazendo $IP_m \leq IP_M$.

Quando ambos valores IP_m e IP_M são próximos de 1, indicam poucas mudanças nas definições dos padrões. No caso em que IP_M seja próximo a 1 mas IP_m distante de 1, indica que ao menos um MR está estabilizado. Finalmente, um valor de IP_M distante de 1 indica que todas a definição do padrão está em mudança em todos os MR.

De maneira análoga, para descrever a estabilidade no ajuste dos pesos podemos analisar os valores intervalos de valores do conjunto $\{T_j^i\}$.

Como apresentado na seção 4.7.5, valores de uma variável T_j^i iguais a 0 ou N indicam os casos de maior mudança de pesos. No caso do valor N indica o maior incremento possível do peso, enquanto o valor N indica a maior diminuição deste. Estes casos correspondem à menor estabilidade possível dos pesos.

Para definir um indicador de estabilidade de pesos compatível com o indicador de estabilidade dos padrões os valores das variáveis $\{T_j^i\}$ devem ser normalizados no intervalo $[0, 1]$.

No caso que $T_j^i < 1$ seja menor que 1 o próprio valor representa a magnitude da mudança de peso e, por tanto, sua instabilidade. No caso $T_j^i > 1$ é necessária uma normalização do intervalo $[1, N]$ no intervalo $[0, 1]$ de modo tal que o caso de maior instabilidade, isto é $T_j^i = N$ seja associado com o valor 0.

A seguinte função define o processo de normalização das variáveis T_j^i.

$$N(T_j^i) = \begin{cases} T_j^i & se \quad T_j^i \leq 1; \\ \frac{1}{(1-N)}(T_j^i - N) & se \quad T_j^i > 1. \end{cases} \tag{4.21}$$

A partir desta função é definido o indicador de estabilidade dos pesos.

O **Indicador de estabilidade dos pesos** $I\theta$ é definido como um par de termos $I\theta = (I\theta_m, I\theta_M$ onde $I\theta_m = \min\{N(T_j^i)\}$ e $I\theta_M = \max\{N(T_j^i)\}$. Os dois elementos do indicador têm valores no intervalo $[0, 1]$ e satisfazem $I\theta_m \leq I\theta_M$.

Note-se que quando os valores de $I\theta_m, I\theta_M$ estão ambos próximos de 1, indica que os pesos de todos os MR sofrem poucas mudanças. No caso contrario, temos que todos os pesos estão sofrendo mudanças importantes.

Em geral, um processo de consulta deve começar com valores dos indicadores distantes de 1. Ao longo do processo de consulta, deve ocorrer uma aproximação de ao menos algum dos termos de $I\theta$ e de IP ao valor 1. Se ambos os termos de cada indicador se aproximam a 1, então indica uma convergência com o usuário. Se nenhum dos dois termos se aproximam então indicaria problema de coerência do usuário ou dos MR.

Assim, um critério de finalização da consulta pode ser definido como a impossibilidade, depois de algumas iterações, de aproximar algum dos indicadores aos valores esperados.

4.8.1 O processo de consulta. Arquivos de *profile*

O modelo de consulta apresentado parte de um processo cíclico de aproximação, dependente do usuário. Este processo de aproximação pode implicar em um esforço importante do usuário para atingir um resultado adequado. Em posteriores consultas, o mesmo pode desejar realizar outras consultas utilizando alguns dos padrões já definidos.

Para utilizar os resultados de consultas anteriores, devem ser criados arquivos *profiles* com a descrição dos parâmetros de definição de um padrão para um usuário . O conteúdo destes arquivos corresponde aos vetores de parametrização $VR_i[j]$ e os pesos θ_j^i para o padrão p_j. Estes valores descrevem exatamente a função de similaridade desse padrão para futuras consultas e para esse usuário em específico.

A partir destes dados, quando a consulta é realizada para mais de um padrão, a maneira inicial mais efetiva de fazê-la é refinar cada padrão individualmente. Depois deste refinamento o seu arquivo *profile* é gerado e, posteriormente, a consulta com mais de um padrão é realizada.

4.9 Resumo

Este capítulo apresentou o modelo de sistema de recuperação por conteúdo proposto. Neste modelo, o principal elemento de consulta é o padrão, que corresponde a um conjunto de regiões de imagens associadas pelo usuário a um mesmo conceito. Como parte integrante do modelo, foi introduzido um esquema de refinamento que considera uma consulta como um processo iterativo que permite ao sistema uma retroalimentação em cada passo, a partir da avaliação pelo usuário dos resultados apresentados. Um outro elemento da proposta refere-se aos múltiplos modelos de representação do conteúdo, como mecanismo para aliviar a inexistência de modelos matemáticos universais de descrição do conteúdo das ISR. O mecanismo de retroalimentação adotado permite o refinamento tanto das definições dos padrões como da associação de pesos aos diferentes modelos de representação do conteúdo, indicando a relevância atribuída pelo sistema a cada um dos MR em relação à retroalimentação do usuário.

O próximo capítulo descreve alguns aspectos relativos à implementação dos conceitos aqui apresentados .

Capítulo 5

Aspectos de implementação do modelo

5.1 Introdução

Este capítulo descreve alguns aspectos de implementação do sistema proposto . Esta implementação está restrita a dois níveis: interface e algoritmos de extração de descritores. O processamento de consultas completo não foi implementado mas está totalmente especificado. Inicialmente, é apresentada a arquitetura para o sistema de recuperação. Adicionalmente, o capítulo apresenta os algoritmos associados a um dos módulos mais importantes do sistema - o processamento das consultas. Este processamento implementa a medida de similaridade definida no capítulo anterior e realiza um processo de otimização visando a redução da complexidade do cálculo das imagens similares. Para tanto, foi projetado um algoritmo de processamento de consultas que explora características da medida de similaridade, assim como a existência de estruturas de índice espacial associadas a cada Modelo de Representação (MR).

5.2 Arquitetura do Sistema

A figura 5.1 apresenta a arquitetura do sistema para o modelo proposto. Esta arquitetura é composta das seguintes camadas:

- Interface: responsável por intermediar a interação do usuário com o uso do sistema. Possui módulos para a inserção de imagens e processamento de consultas.
- Processamento: responsável pelo processamento das imagens a serem inseridas e pelo processamento das consultas.

- Armazenamento: responsável pelo armazenamento das imagens e dos índices para recuperação.

Mas detalhadamente, os módulos nas diferentes camadas são os seguintes:

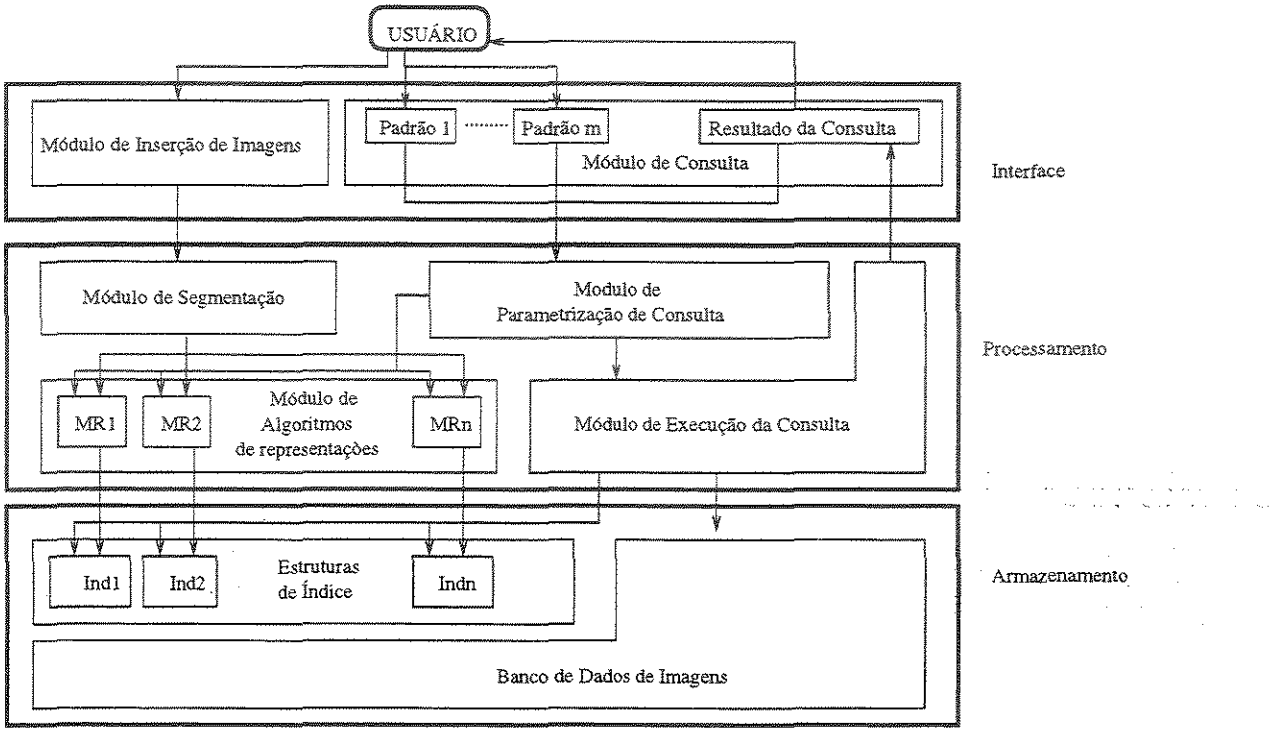


Figura 5.1: Arquitetura do Sistema.

- **Módulo de Consulta:** Este módulo implementa a comunicação com o usuário para a definição das consultas e a visualização dos resultados. Deve oferecer recursos visuais para a definição dos padrões.

- **Módulo de Inserção de Imagens:** Este módulo implementa a comunicação com o usuário para a inserção de novas imagens no Banco de Dados. Este módulo deve ser de uso apenas do administrador do sistema.

- **Módulo de Parametrização de Consulta:** Este módulo processa os padrões e gera parâmetros para cada um deles, a serem utilizados na a execução da consulta. Estes parâmetros são: pesos associados pela função de similaridade a cada padrão e representação (Equação 4.20); e vetores de parametrização para cada padrão (Equação 4.3). Para estes últimos o módulo utiliza o Módulo de Algoritmos de Representação.

- **Módulo de Execução de Consulta.** Este módulo implementa o algoritmo de execução da consulta que será apresentado em detalhes na seção 5.3. Recebe os parâmetros do Módulo de parametrização e utiliza os recursos das estruturas de índices.

- **Módulo de Segmentação.** Este módulo implementa o algoritmo de segmentação das imagens e é utilizado pelo Módulo de Interface de Inserção.

- **Módulo de Algoritmos de Representação:** Este módulo implementa os diferentes algoritmos dos modelos de textura e cor adotados. Na etapa de inserção, cada um dos algoritmos é aplicado em cada região da imagem obtida pelo módulo de segmentação, gerando, assim, para cada segmento, um vetor de características para cada MR. Na etapa de consulta, o Módulo de Parametrização aplica os algoritmos sobre as amostras associadas a cada padrão, utilizando os resultados para gerar os parâmetros a serem utilizados na execução da consulta.

- **Estruturas de Índice.** Para cada MR é definida uma estrutura de índice à qual é associada uma distância. Cada estrutura de índice armazena os vetores de característica das regiões de todas as imagens do Banco de Dados para a forma de representação associada. Uma estrutura de índice deve oferecer como recurso básico a possibilidade de determinar uma lista ordenada dos k vetores de regiões de imagem mais próximos a um vetor característico, isto é, um vetor de referência que representa uma consulta sobre esta estrutura. A seção 5.5 apresentara detalhadamente o uso das estruturas de índice.

As seções seguintes abordam mais detalhadamente as etapas de uma consulta e em particular o Módulo de Execução da Consulta.

5.3 Algoritmo de Processamento de Consultas.

Como apresentado no capítulo anterior, o resultado de uma consulta no banco de dados pode ser entendido como a busca das N imagens com maior valor de similaridade global$Sg(P_C, J)$ (Equação 4.7). A expressão da função de similaridade Sg é aplicada a cada imagem armazenada no banco de dados. Os termos desta expressão são representados na árvore de cálculo da similaridade (figura 4.8) que reflete o aumento da complexidade do cálculo com o aumento do número de padrões, de MR e de regiões segmentadas das imagens. Devido a esta complexidade, e tentando evitar a busca exaustiva em todo o banco de dados, é necessário definir um algoritmo que reduza o custo computacional permitindo o processamento da consulta de maneira eficiente. Esta seção apresenta o processo de

execução de consulta, com a introdução de estruturas de índice espacial para facilitar esta execução e, finalmente, uma proposta de algoritmo para este processamento.

Resumidamente, uma consulta do usuário tem a forma: $Q = (P, N)$ sendo $P = \{p_1, p_2, ..., p_{np}\}$ a definição dos diferentes padrões e N o número máximo de imagens a serem recuperadas. Cada padrão $p_i = \{r_1(p_i), r_2(p_i), \cdots, r_m(p_i)\}$, é definido inicialmente pelas regiões de diferentes imagens de consulta escolhidas como amostras do padrão. Este processo de definição dos padrões é realizado pelo usuário através do Módulo de Consulta apresentado na figura 5.1.

A partir desta definição o processamento de uma consulta pode ser visto como a execução de várias etapas que são representadas na figura 5.2. Estas etapas são as seguintes:

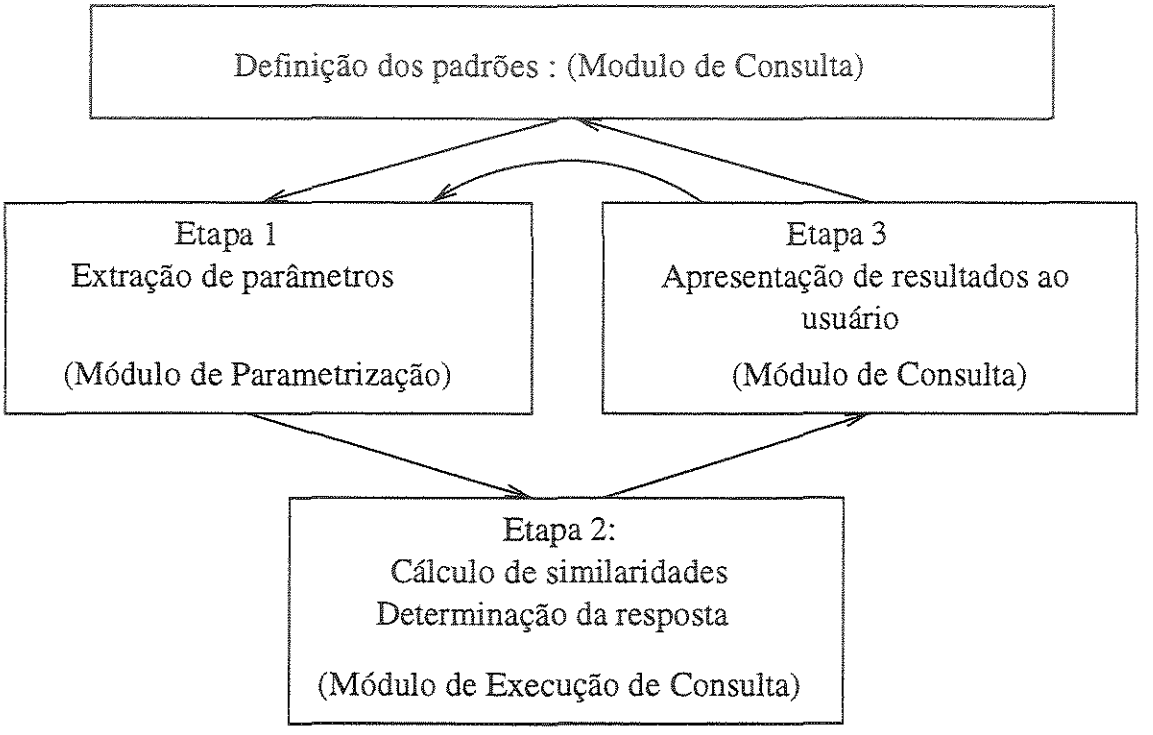


Figura 5.2: Etapas do processamento de uma Consulta

1. Extração de parâmetros dos diferentes padrões. Esta etapa é realizada pelo Módulo de Parametrização.

2. Cálculo das similaridades e determinação das imagens no conjunto resposta. Esta etapa é realizada pelo Módulo de Execução da consulta.

3. Apresentação dos resultados ao usuário. Este confirma a relevância dos resultados apresentados e o processamento volta à etapa 1 para o cálculo de novos parâmetros para uma nova consulta. Esta etapa é realizada no Módulo de Consulta.

Estas diferentes etapas serão apresentadas a seguir, considerando aspectos de implementação.

5.3.1 Etapa 1: Extração de parâmetros.

Como vimos no capítulo anterior, a extração de parâmetros dos padrões é realizada a partir da aplicação das funções de parametrização T_i dos distintos MR. Esta extração gera a estrutura de dados $P_C = (P_C[p_1], \cdots P_C[p_{np}])$ (Equação 4.3) que contém o conjunto de vetores de parametrização. Cada vetor de parametrização $P_C[p_j] = \{VR_1[j], \cdots, VR_k[j], \cdots, VR_{nr}[j]\}$ está composto pelas descrições do padrão para cada MR, isto é, $VR_i[j]$ corresponde à aplicação da função de parametrização T_i do MR R_i sobre as regiões que definem o padrão p_j (Equação 4.4).

O algoritmo 5.1 descreve resumidamente o processo de extração e construção dos vetores de parametrização.

5.3.2 Etapa 2: Cálculo das similaridades e determinação das imagens no conjunto resposta.

O modo "natural"de determinação das imagens mais similares em relação a uma consulta é analisar uma a uma as imagens do banco de dados e calcular sua similaridade global. Aquelas com maior valor de similaridade são as escolhidas. Este tipo de processamento não é eficiente mas será adotado, inicialmente, para apresentar mais claramente os algoritmos. Posteriormente, na seção 5.5 será apresentada uma versão modificada e eficiente.

Etapa 2.1: Cálculo da similaridade global das imagens no banco de dados

O procedimento Sg calcula a similaridade global dos padrões com cada imagem do banco de dados. Inicialmente, o procedimento Sp é usado para calcular o valor de similaridade intra-padrão da imagem com cada padrão p_j. Como resultado, obtem-se uma dupla $(s[j], r[j])$ que corresponde ao valor de similaridade intra-padrão $s[j]$ da imagem com esse padrão e a região $r[j]$ da imagem onde essa maior similaridade é atingida.

A similaridade global associada à imagem é definida como o mínimo das similaridades $\{s[1], s[2], \cdots, s[np]\}$ (Equação 4.7).

Como resultado final deste procedimento, são devolvidas as N imagens com maior valor de similaridade, seus valores de similaridade, e as regiões dentro da imagem mais similares a cada padrão. O algoritmo 5.2 descreve este procedimento.

Algoritmo 5.1 (Extração de parâmetros)

Forma Geral: $Extract(p_1, p_2, \cdots, p_n) \rightarrow P_C$
Entrada:

$p_1, p_2, \cdots, p_n$: Padrões da consulta onde:
$p_i = \{r_1(p_i), r_2(p_i), \cdots, r_m(p_i)\}$ são as
regiões de diferentes imagens que formam o padrão p_i

Saída:

P_C: Resultado da extração de parâmetros de cada
padrão p_j em cada MR R_i

Passos:
Para cada padrão p_j
Para cada MR R_i
$VR_i[p_j] \leftarrow T_i(r_1(p_j), r_2(p_j),), ..., r_m(p_j))$
$P_C[p_j] \leftarrow \{VR_1(p_j), VR_2(p_j), ..., VR_{nr}(p_j)\}$
$P_C = \{P_C[p_1], P_C[p_2], \cdots, P_C[p_{np}]\}$
return P_C

Algoritmo 5.2 (Cálculo das N imagens mais similares)

Forma Geral: $S_g(P_C, N) \rightarrow (Jr, Sr, Rr)$
Entrada:

P_C: parâmetros de consulta resultantes da etapa 1
N: número de imagens na resposta

Saída:

$Jr = \{J_1, J_2, \cdots, J_N\}$: K imagens de maior similaridade com a parametrização da consulta P_C
$Sr = \{S_g(J_1), \cdots, S_g(J_N)\}$ similaridade destas imagens com cada um dos padrões
$Rr = \{reg(J_1), \cdots, reg(J_N)\}$ regiões mais similares a cada padrão dentro de cada imagem selecionada em Jr

Passos:
Para cada imagem $J \in BD$ (banco de dados)
  Para cada padrão $p_j \in P_C$
    $(s[j], r[j]) \leftarrow Sp(P_C[p_j], J)$ /*similaridade intra-padrão */
  $s_g[J] \leftarrow min(s[1], s[2], \cdots, s[np])$ /*similaridade global de J */
  $reg[J] \leftarrow \{r[1].r[2], \cdots, r[np]\}$
SortDB $\leftarrow$ Ordenar as imagens $J \in BD$ pelo valor de $s_g[J]$
$Jr = \{J_1, J_2, ..., J_N\}$, /* N primeiras imagens devolvidas por SortDB */
$Sr = \{s_g[J_1], s_g[J_2], ..., S_g[J_N]\}$
$Rr = \{reg[J_1], reg[J_2], ..., reg[J_N]\}$
return (Jr, Sr, Rr)

Algoritmo 5.3 (Cálculo da similaridade intra-padrão)

Forma Geral: $Sp(P_C[p_j], J) \rightarrow (sim, reg)$
Entrada:
$P_C[p_j]$: Parâmetros de consulta para o padrão p_j
J: Imagem do BD sendo processada.
Saída:
sim : Similaridade entre o padrão p_j e a imagem J.
reg : Região de J onde é reconhecida a maior similaridade com o padrão p_j
Passos:
Para cada MR R_i
$(sim[i], reg[i]) \leftarrow SimIm(VR_i[j], J)$ /* similaridade entre o padrão e cada imagem e a região dentro dessa imagem onde a similaridade é atingida para o MR R_i */
Seja mx tal que $sim[mx] = Max\{sim[i]\}$
return $(sim[mx], reg[mx])$

Etapa 2.2: Cálculo da similaridade intra-padrão de uma imagem

Dentro do procedimento Sg definido anteriormente, é utilizado o procedimento Sp que calcula o valor de similaridade intra-padrão entre uma imagem e um padrão determinado, considerando cada modelo de representação. O cálculo segundo cada MR é realizado pela função $SimIm$ que analisa a similaridade do padrão com cada região da imagem, escolhendo como valor de similaridade o maior valor entre todas as regiões (Equação 4.10). Para calcular cada um destes valores de similaridade entre uma região e um padrão é utilizada a função de distância s_i de cada MR(Equação 4.11).

Utilizando a metáfora dos juízes, consideramos que cada um deles seleciona sua melhor região para a imagem com um valor de similaridade. A seguir é aplicada a função de pontuação sobre estas seleções (neste caso o máximo) para calcular o valor final de similaridade intra padrão (Equação 4.10). A região devolvida como mais semelhante ao padrão é a região escolhida pelo juiz que associou maior valor de similaridade, sendo, portanto, sua similaridade escolhida como similaridade final.

O algoritmo 5.3 apresenta o cálculo da similaridade intra-padrão. Lembremos que neste algoritmo $SimIm(VR_i[j], J)$ corresponde ao valor máximo da função de distância s_i da representação R_i aplicada sobre todas as regiões de J (Equação 4.11).

A próxima seção descreve a última etapa do processamento de uma consulta corres-

Algoritmo 5.4 (Apresentação de resultados)

Forma Geral: $Display(Jr, Sr, Rr)$
Entrada:(Jr, Rr, Sr) : `Resultado da anterior (Etapa 2)`

Passos:
Para cada $J \in Jr$
`Mostrar Imagem J`
`Mostrar valor de Similaridade` $s_g[J]$
Para cada `região` $r[k] \in Rr[J]$
`Apresentar dentro de J o retângulo envolvente mínimo da região` $r[k]$.

pondente à apresentação dos resultados

5.3.3 Etapa 3: Apresentação dos resultados ao usuário.

Os resultados obtidos na etapa anterior são apresentados ao usuário para sua avaliação. Este processo deve mostrar ao usuário a cada passo: o valor de similaridade atribuído pelo sistema a cada imagem da resposta e a visualização da região da imagem considerada a mais similar em relação a cada um dos padrões da consulta. O algoritmo 5.4 descreve resumidamente este processo sem entrar nos detalhes de interface gráfica da apresentação dos resultados.

As três etapas descritas anteriormente definem uma iteração do processamento de consulta. Este processo deve se repetir a partir da redefinição dos padrões pelo usuário e, conseqüentemente, a extração de novos parâmetros.

A Etapa 2 descrita anteriormente foi apresentada de maneira simplificada sem incluir os pesos associados a cada MR e cada padrão. A seção a seguir introduz os pesos no processamento da consulta, e redefine os algoritmos da Etapa 2.

5.4 Introdução da retroalimentação.

O modelo de refinamento proposto neste trabalho implica numa sucessão de iterações para melhorar o resultado final, a partir da avaliação pelo usuário da relevância dos resultados apresentados.

Seja RU_j, como anteriormente, o conjunto das regiões do resultado da consulta que o usuário seleciona como relevantes para o padrão p_j. Como foi analisado no capítulo 4 o impacto desta seleção é introduzido em dois momentos do processamento:

1. Extração de novos parâmetros. Como o conjunto RU_j corresponde a exemplos relevantes do padrão p_j, a definição deste padrão deve mudar com a inclusão das regiões de RU_j, isto é, $p_j = p_j \cup RU_j$. Esta extração corresponde à etapa 1 do processamento da consulta apresentada no algoritmo 5.1 *Extract*)

2. Introdução de pesos no cálculo das similaridades. O segundo aspecto em que a retroalimentação do usuário influencia o processamento corresponde à introdução de pesos dentro da medida de similaridade intra-padrão Sp_{θ_j} que utiliza pesos θ_j^i para cada MR R_i em cada padrão p_j (Equação 4.14). Como apresentado na seção 4.7.5, para o cálculo destes pesos, além dos conjuntos RU_j, são necessários os conjuntos RS_j correspondentes às regiões selecionadas na resposta do sistema para cada padrão e, RM_j^i correspondentes às regiões selecionadas para cada MR, para cada padrão. A introdução de pesos redefine os algoritmos da etapa 2 apresentados anteriormente. A próxima seção descreve estas modificações.

5.4.1 Redefinição do cálculo da similaridade global Sg com pesos

Com a introdução de pesos, a medida de similaridade apresentada na equação 4.7 deve ser avaliada para cada imagem.

O procedimento 5.2 (Sg) deve ser modificado para incluir o cálculo dos pesos. A modificação consiste basicamente na introdução dos procedimentos *InicializarPesos*(Θ), *AtualizarPesos* e *InicializarSelec*().

O procedimento *InicializarPesos*(Θ) é executado na primeira iteração de uma consulta. Este inicializa todos os pesos com um mesmo valor $\frac{1}{nr}$, sendo nr o número de representações. Esta inicialização corresponde a considerar inicialmente a mesma relevância para todos os MR em todos os padrões.

Nas sucessivas iterações, o procedimento *AtualizarPesos* recalcula os pesos considerando os critérios apresentados na Equação 4.7.5 do capítulo anterior.

O procedimento *InicializarSelec*(), inicializa como vazios todos os conjuntos RS_j e RM_j^i que são redefinidos em cada nova iteração.

O algoritmo 5.5 descreve o procedimento Sg com as modificações resultantes da introdução dos pesos discutidas anteriormente.

A próxima seção descreve o procedimento de cálculo da similaridade intra-padrão com a introdução dos pesos.

Algoritmo 5.5 (Cálculo das N imagens mais similares com retroalimentação)

Forma Geral $S_g(P_C, N, RU_j, \{RS_j\}, \{RM_j^i\}, flag) \rightarrow (Jr, Sr, Rr, \{RS_j\}, \{RM_j^i\})$
Entrada:

P_C: parâmetros de consulta resultantes da etapa 1
N: número de imagens na resposta
RU_j: Conjunto das regiões relevantes para o usuário em cada padrão na iteração anterior
$\{RS_j\}$: Conjuntos das respostas do sistema para cada padrão na iteração anterior
$\{RM_j^i\}$: Conjuntos das regiões selecionadas por cada MR R_i como mais similares a cada padrão na iteração anterior
flag : Indica se esta chamada é a primeira iteração da consulta

Saída:

(Jr, Sr, Rr): Similar ao apresentado no algoritmo 5.2
$\{RS_j\}, \{RM_j^i\}$: Atualizações dos conjuntos de regiões selecionadas

Passos
Se flag
$InicializarPesos(\Theta)$
$InicializarSelec(\{RS_j\}, \{RM_j^i\})$
senão
$AtualizarPesos(Selec, Relev, \Theta)$
Para cada imagem $J \in DB$
Para cada padrão p_j
$(s[j], r[j]) \leftarrow Sp(P_C[p_j], J, \Theta)$
$sim[J] \leftarrow Min\{s[1], s[2], \cdots, s[np]\}$
$reg[J] = \{(r[1]., r[2], \cdots, r[np]\}$

SortDB $\leftarrow$ Ordenar as imagens $J \in DB$ pelo valor de $sim[J]$
Determinar os conjuntos $\{RS_j\}$ como o conjunto das regiões das imagens em Jr mais similares a cada padrão p_j.
Determinar os conjuntos $\{RM_j^i\}$ de cada MR R_i mais similares a cada padrão p_j.
$Jr = \{J_1, J_2, ..., J_N\}$, N primeiras imagens em SortDB
$Sr = \{s_g[J_1], s_g[J_2], ..., S_g[J_N]\}$
$Rr = \{reg[J_1], reg[J_2], ..., reg[J_N]\}$
return (Jr, Sr, Rr)

5.4.2 Redefinição do cálculo da similaridade intra-padrão *Sp* com pesos

Como visto anteriormente, o procedimento *Sg* utiliza a função *Sp* que calcula a similaridade intra-padrão. O algoritmo de *Sp* deve avaliar a expressão de similaridade intra-padrão com pesos que foi apresentada na Equação 4.14.

Notemos, inicialmente, que o algoritmo de cálculo de *Sp* sem pesos (algoritmo 5.2), seleciona como região com maior similaridade àquela associada ao MR R_i com maior valor de similaridade da função $SimIm(VR_i[j], J)$. Isto acontece porque a função *Mp* que calcula a similaridade intra-padrão final é a função *max*.

O algoritmo de *Sp* com pesos deve escolher também uma região da imagem como a mais similar ao padrão processado. No entanto, analisando a expressão da função Mp_{Θ_j} (Equação 4.14) se constata que desaparece a noção de qual MR determina o valor final de similaridade, como no caso da função *Sp* sem pesos. Isto é, não é possível selecionar a melhor região só analisando os valores resultantes das chamadas a $SimIm(VR_i[j], J)$ que se combinam para produzir o valor de similaridade intra-padrão.

Seguindo a metáfora, cada votante seleciona um valor de similaridade para a imagem, a partir de uma região sua mais similar com o padrão p_j. Como o valor final de similaridade da imagem combina todas as similaridades dos votantes, não é possível determinar qual das diferentes regiões propostas pelos votantes deve ser a indicada como a que de maior similaridade com esse padrão.

Como solução, é necessário aplicar a função Mp_{Θ_j} sobre cada região selecionada por cada votante associada ao maior valor de $SimIm(VR_i[j], J)$. Dentre estes valores resultantes, aquele com maior valor será escolhido como valor de similaridade intra-padrão *Sp* e, conseqüentemente, a região associada a esse MR, será a escolhida como mais similar.

Novamente, seguindo a metáfora, isto corresponderia ao processo em que cada votante pedisse aos demais votantes para avaliar a similaridade da imagem utilizando a região por ele escolhida. Os valores de todos os votantes se combinam para obter a similaridade *Mp* para essa região. Isto será realizado para a região selecionada por cada votante. Ao final, teremos tantos valores de similaridade quanto de votantes. Aquele maior valor corresponderá ao valor de similaridade intra-padrão devolvido e sua região associada, a selecionada.

Definindo formalmente esta idéia, seja $\{reg_1, reg_2, \cdots, reg_k, \cdots, reg_{nr}\}$ o conjunto de regiões selecionadas da imagem J tal que a região reg_k é aquela com maior valor de similaridade sim_k com o padrão p_j, segundo o MR R_k, isto é, $(sim_k, reg_k) = SimIm(VR_k[j], J)$. Assim, o valor de similaridade intra-padrão para J é calculado como:

$$Sp_{(\Theta_j)}(P_C[p_j], J) = \max_{1 \geq k \geq nr} \{Mp_{\Theta_j}(reg_k)\} \tag{5.1}$$

em que:

$$\begin{aligned} Mp_{\Theta_j}(reg_k) = \; & (\theta_j^1 - \theta_j^2) \cdot s_1(reg_k^i[J], VR_i[j]) \\ & +2 \cdot (\theta_j^2 - \theta_j^3) \cdot max_{1 \leq i \leq 2}\{s_i(reg_k^i[J], VR_i[j])\} \\ & + \cdots \\ & +m \cdot (\theta_j^m - \theta_j^{m+1}) \cdot max_{1 \leq i \leq m}\{s_i(reg_k^i[J], VR_i[j])\} \\ & + \cdots \\ & +nr \cdot \theta_j^{nr} \cdot max_{1 \leq k \leq nr}\{s_i(reg_k^i[J], VR_i[j])\} \end{aligned} \tag{5.2}$$

$reg_k^i[J]$ corresponde ao descritor de conteúdo da região $reg_k[J]$ no banco de dados, segundo o MR R_i e nr o número de MR.

Note-se que os valores dos pesos determinam a ordem dos MR na avaliação da função. Esta modificação traz um impacto importante no cálculo de Sp. Com a modificação introduzida é preciso determinar inicialmente as regiões mais similares segundo cada MR e, posteriormente, calcular para cada uma delas a função Mp.

As idéias discutidas implicam na modificação do algoritmo 5.3 de cálculo de Sp. Estas modificações são apresentadas resumidamente no algoritmo 5.6. Este algoritmo permite recuperar as N imagens mais similares aos padrões fornecidos originalmente, considerando a introdução de pesos. No entanto, esta solução não é viável computacionalmente pelo seu elevado custo. Este custo está determinado pela necessidade de analisar exaustivamente todas as imagens do banco de dados. Para cada uma destas imagens á avaliada uma função de similaridade complexa que envolve a análise exaustiva de todas as regiões da imagem considerando os diferentes padrões e MR.

A próxima seção apresenta uma proposta de algoritmo que reduz o custo deste processamento.

5.5 Otimizando o processamento de consultas

Alguns aspectos podem ser considerados na otimização do algoritmo de determinação das imagens mais similares (algoritmo AlgGlobal2) como por exemplo:

- Evitar processar extensivamente o banco de dados diminuindo o número de imagens e de regiões a analisar.
- Evitar avaliar toda a expressão da função de similaridade para uma imagem (Equação 5.2), realizando "podas"da árvore de similaridade que representa esta expressão.

As seções seguintes exploram estes aspectos mencionados.

Algoritmo 5.6 (Cálculo da similaridade intra-padrão com retroalimentação)

Forma Geral $Sp(P_C[p_j], J, \Theta_j) \rightarrow (sim, reg)$
Entrada:

$P_C[p_j]$: Parâmetros de consulta para o padrão p_j
J: Imagem do BD sendo processada
Θ_j : Pesos para o padrão p_j

Saída:

sim : Similaridade entre padrão p_j e a imagem J.
reg : Região de J onde é reconhecida a maior similaridade com o padrão p_j

Passos:
Para cada MR R_i
$(sim[i], reg[i]) \leftarrow SimIm(VR_i[j], J)$ /* similaridade do padrão com cada imagem e região dentro da imagem onde a maior similaridade é atingida para o MR R_i*/
$Atualizar(RM_j^i, (sim[i], reg[i])$
Para cada região $reg[i]$ selecionada
$Result[j] \leftarrow Mp_{\Theta_j}(reg[j])$ /* Avaliação de Mp para cada região selecionada por cada MR */
Seja mx tal que $Result[mx] = max\{Result[j]\}$
return $(sim[mx], reg[mx])$

5.5.1 Otimização introduzindo listas de regiões

Com o objetivo de evitar o processamento de todas as regiões das imagens armazenadas no banco de dados, propomos a introdução de listas em diferentes níveis do processamento. Estas listas vão combinando seus valores de acordo com as funções de similaridade global e intra-padrão apresentadas até determinar as N imagens mais similares.

A abordagem de processamento exaustivo do banco de dados, apresentada na seção anterior, descreve a determinação das N imagens mais similares de uma maneira descendente (*top-down*). Na abordagem atual, este processo é realizado de maneira ascendente (*bottom-up*), o que permite reduzir o número de imagens e regiões processadas.

Suponhamos que dentro do banco de dados seja possível construir um conjunto de listas ordenadas $\{X_j^i\}$ cada uma das quais associada com o MR R_i e o padrão p_j. Cada lista ordenada X_j^i contém todas as regiões de imagens no banco de dados dinamicamente ordenadas segundo o valor da função de distância s_i, desse MR, aplicada sobre essas regiões e o vetor de parametrização do padrão p_j, isto é, $VR_i[j]$. Chamaremos estas listas de **Listas de modelos de representação**.

Todas as listas associadas a um padrão p_j, isto é, o conjunto de listas $\{X_j^1 X_j^2, \cdots, X_j^{nr}\}$, são utilizadas para o processamento da similaridade intra-padrão associada a p_j. Processar a similaridade intra-padrão para p_j significa, nesta abordagem, criar uma nova lista Sp_j que contém todas as regiões do banco de dados ordenadas segundo seu valor de similaridade intra-padrão. Estas são chamadas de **Listas de similaridade intra-padrão**. Como resultado deste segundo nível, são criadas tantas listas quanto o número de padrões, isto é, o conjunto $\{Sp_j\}$. Finalmente, todas estas listas de segundo nível são processadas para criar uma última lista Sg com todas as regiões do banco de dados ordenadas segundo sua similaridade global. Esta lista é chamada de **Lista de similaridade global**. Os N primeiros elementos desta última lista representam o resultado final da consulta.

A figura 5.3 ilustra este processo. Para cada combinação de padrão p_j e MR R_i é criada uma lista de representação X_j^i. A seguir, o conjunto de listas associadas a um mesmo padrão é processado resultando na nova lista de similaridade intra-padrão Sp_j. Finalmente, a figura mostra no último passo, como todas as listas de similaridade intra-padrão são processadas gerando-se a lista de similaridade global Sg.

Do ponto de vista de criação, estas listas não são calculadas completamente e são implementadas por demanda, isto é, os elementos da lista vão sendo inseridos à medida que são requisitados.

Nas próximas seções, veremos como estas listas são implementadas e como podem ser combinadas para realizar o cálculo da similaridade, utilizando as equações de similaridade intra-padrão e global. A apresentação será feita de maneira ascendente, isto é, começaremos apresentado as listas de representação X_j^i até se atingir o nível superior

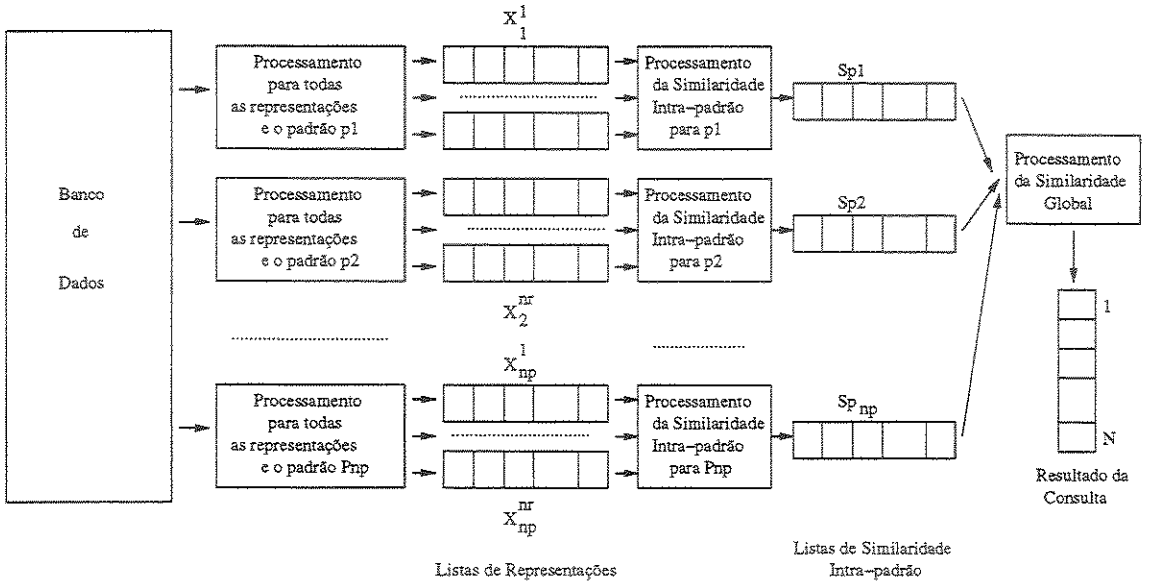


Figura 5.3: Esquema geral do processamento de uma consulta utilizando listas.

correspondente à lista de similaridade globalSg. cujos N primeiros elementos constituem o resultado da consulta.

5.5.2 Listas de modelos de representação

Cada lista de modelo de representação $\{X_j^i\}$ deve garantir uma ordenação das regiões do banco de dados, segundo a similaridade com o padrão p_j, utilizando como critério de ordenação o valor da função s_i da representação R_i, isto é, $X_j^i = \{reg_1, reg_2, \cdots, reg_{NB}\}$, tal que, $s_i(reg_k^i, VR_i[j]) \geq s_i(reg_m^i, VR_i[j]), \forall m \geq k$ sendo NB o número total de regiões armazenadas no banco de dados e reg_k^i e reg_m^i são os vetores característicos das regiões reg_k e reg_m segundo o MR R_i..

Lembremos que cada região reg_k de uma imagem no banco de dados foi processada no momento da sua inserção. Este processamento resultou na geração de um vetor característico reg_k^i para cada modelo de representação R_i e seu armazenamento na estrutura de índice correspondente.

Para a implementação destas listas são utilizadas as estruturas de índice espacial mencionadas na seção 5.2. Cada MR R_i terá associada uma estrutura de índice espacial $\mathcal{X}_i$.

Como foi discutido no capítulo 3, uma estrutura de índice espacial armazena e organiza vetores multidimensionais, de maneira tal que consultas de tipo espacial possam ser realizadas utilizando uma função de distância definida sobre essa estrutura. Dado que as funções de distância ou de similaridade s_i adotadas por cada MR estão baseadas em

métricas, é possível considerar que cada $\mathcal{X}_i$ utiliza s_i como critério de organização dos vetores armazenados. Tal estrutura de índice espacial pode ser sempre construída com uma árvore M [13] que exige unicamente que a função de distancia associada à estrutura seja uma métrica. Pela definição de MR, cada s_i satisfaz esta condição. Outras estruturas de índice espacial, como a família das árvores R [29, 75, 8, 37], consideram eficientemente funções de distância baseadas em métricas de Minkowski.

Dentre as consultas clássicas que estruturas de índice espacial temos as **consultas de proximidade**, isto é:

"Dado um vetor multidimensional de consulta V e uma estrutura de índice espacial $\mathcal{X}$ recuperar os N pontos mais próximos de V segundo a função de distância definida em $\mathcal{X}$

Considerando que cada MR R_i tem associada uma estrutura de índice $\mathcal{X}_i$, esta consulta espacial pode ser resolvida em qualquer MR. O vetor de parametrização $VR_i[j]$ descreve o padrão p_j no modelo de representação R_i. Uma consulta de proximidade em $\mathcal{X}_i$, usando como vetor de consulta $VR_i[j]$, devolve os vetores característicos das regiões do banco de dados mais próximos a $VR_i[j]$, isto é, as regiões segundo esse MR mais similares ao padrão p_j.

Consideremos que para uma lista X_j^i são definidas as seguintes funções:

- $Head(X_j^i) \rightarrow (reg, I, sim)$: devolve a região reg da imagem I que estiver no início da lista juntamente com o seu valor de similaridade com $VR_i[j]$. Durante a construção por demanda desta lista, esta região é em cada momento a de maior similaridade com $VR_i[j]$ em todo o banco de dados.

- $Next(X_j^i)$: Atualiza o inicio da lista com a próxima região mais similar com $VR_i[j]$.

Dado que as listas são construídas por demanda, o processamento fundamental está concentrado na função $Next$. Quando a região no início da lista é requisitada e a lista fica vazia, então desencadeia-se o processo de procurar a próxima região, segundo o critério de ordenação dado pela função s_i.

O algoritmo 5.7 apresenta a definição desta função $Next$. Quando a função $Head$ é utilizada são devolvidos os valores associados à região no inicio da lista de representação. As listas de representação associadas a um mesmo padrão são combinadas em uma lista de similaridade intra-padrão. A próxima seção discute como esta lista é implementada.

5.5.3 Listas de similaridade intra-padrão

O objetivo de cada lista de similaridade intra-padrão Sp_j é manter ordenadas as regiões do banco de dados segundo a similaridade intra-padrão em relação a p_j, isto é, $Sp_j = \{reg_1, reg_2, \cdots, reg_{NB}\}$, tal que: $Mp_{\Theta_j}(reg_k) \geq Mp_{\Theta_j}(reg_m), \forall m \geq k$, considerando a função Mp_{Θ_j} da equação 5.2.

Algoritmo 5.7 (Próxima região na lista de representação X_j^i)

Forma Geral: $Next(X_j^i)$
Entrada: X_j^i: Lista de Representação para o padrão p_j e o MR R_i
Passos:
Se a lista tem mais de um elemento, avançar na prõxima posição
senão
    Realizar uma consulta de proximidade em $\mathcal{X}_i$ para recuperar as próximas N regiões de menor distância $s_i(r_k^i, VR_i[j])$.
    Armazenar em X_j^i as regiões recuperadas

Como vimos na seção 5.4.2, para calcular a similaridade intra-padrão de uma região reg_k é necessário o cálculo da similaridade desta em todos os MR R_i para combinar estes resultados em um valor final $Mp_{\Theta_j}(reg_k)$.

Uma lista de representação X_j^i corresponde às regiões do banco de dados ordenadas pelo valores de similaridade com o MR R_i. Utilizando a metáfora dos juizes, cada votante vai gerando sua lista com as regiões do banco de dados ordenadas segundo sua percepção de similaridade com um padrão considerado. Somente quando uma região aparece nas propostas de todos os votantes, é que se pode avaliar como possível região mais similar em relação ao padrão correspondente. Isto implica que para cada região que aparece em uma lista de representação , é necessário avaliar a expressão de similaridade Mp_{Θ_j} com os valores de similaridade dessas regiões em todas as outras listas de representação. As regiões já avaliadas são então ordenadas segundo estes valores resultantes para formar a lista de similaridade intra-padrão Sp_j.

Para o caso das listas de similaridade intra-padrão podemos supor também que existem duas operações fundamentais:

- $Head(Sp_j, \Theta_j) \rightarrow (reg, J, sim)$: devolve a região reg da imagem J que estiver no início da lista juntamente com o seu valor de similaridade intra-padrão Mp_{Θ_j}. Durante a construção por demanda desta lista, esta região é em cada momento a de maior similaridade intra-padrão para p_j em todo o banco de dados.
- $Next(Sp_j)$: Atualiza o início da lista com a próxima região mais similar a p_j.

Da maneira similar às listas de representações, para construir a lista Sp_j não é necessário ordenar todo o banco de dados segundo sua similaridade intra-padrão. Para

tanto, basta garantir que toda vez que a função *Next* seja chamada, esta devolva exatamente a próxima região, na ordem definida pela similaridade intra-padrão, ou seja, a lista é construída por demanda. No entanto, como as regiões aparecem aleatoriamente nas diferentes listas de representação é necessário armazenar temporariamente os resultados de uma delas até completar todo o processamento da consulta.

Para armazenar a informação temporária das listas de representação, é definida uma estrutura $ImProc[J]$. Esta estrutura contém todas as imagens sendo processadas em um determinado momento, isto é, aquelas imagens para as quais ao menos uma região já apareceu na lista de representação de algum MR. Por sua vez, para cada imagem já incluída nesta estrutura, é mantida uma lista $ImProc[J].RegProc[k]$ que contém as regiões dessa imagem em fase processamento, isto é, que apareceram em alguma das listas de representação.

Finalmente, para cada região $reg_k[J]$ é mantida uma estrutura $Termos[m]$ que contém o valor de $s_i(reg_k^i[J], VR_i[j])$, isto é, o valor de similaridade para a região $reg_k[J]$ obtido da lista X_j^i. Analisando a equação 5.2 este valor corresponde ao elemento variável do m-ésimo termo da equação 5.2 de similaridade intra-padrão para a região $reg_k[J]$.

Note-se a importância da ordem entre os diferentes MR na expressão de similaridade: o m-ésimo MR estará determinado pelo valor do seu peso.

A idéia básica do processamento é a seguinte: selecionar a região no início da lista que tem o maior valor de similaridade dentre todas as listas, isto é, processar a região $X_v^j.reg$ tal que $Head(X_v^j).sim = \max_{1 \geq i \geq nr}\{Head(X_i^j).sim\}$. Esta região selecionada reg_k é inserida na estrutura $ImProc$ atualizando o v-ésimo termo, isto é,

$$ImProc[Head(X_v^j).I].RegProc[Head(X_v^j).reg].Termos[v] \leftarrow Head(X_v^j).sim \qquad (5.3)$$

A figura 5.4 ilustra este processo . À esquerda na figura, a região reg_k com maior valor de similaridade está na lista X_v^j. Esta região é inserida em $ImProc$ na entrada correspondente à sua imagem, isto é, $ImProc[J]$. Para a imagem J existe uma estrutura $RegProc$ associada. A entrada $ImProc[J].RegProc[k]$ dentro desta estrutura que corresponde à região reg_k que está sendo processada é atualizada. $ImProc[J].RegProc[m]$. Finalmente de acordo com a lista de onde a região foi obtida, será atualizada sua entrada correspondente em de $ImProc[J].RegProc[m].Termos[v]$.

Este processo é repetido até se conseguir determinar qual a próxima região de maior similaridade intra-padrão.

Analisando a equação 5.2, vemos que para o cálculo do valor de similaridade intra-padrão de uma região $reg_k[J]$ precisamos determinar nr termos, A expressão do m-ésimo termo

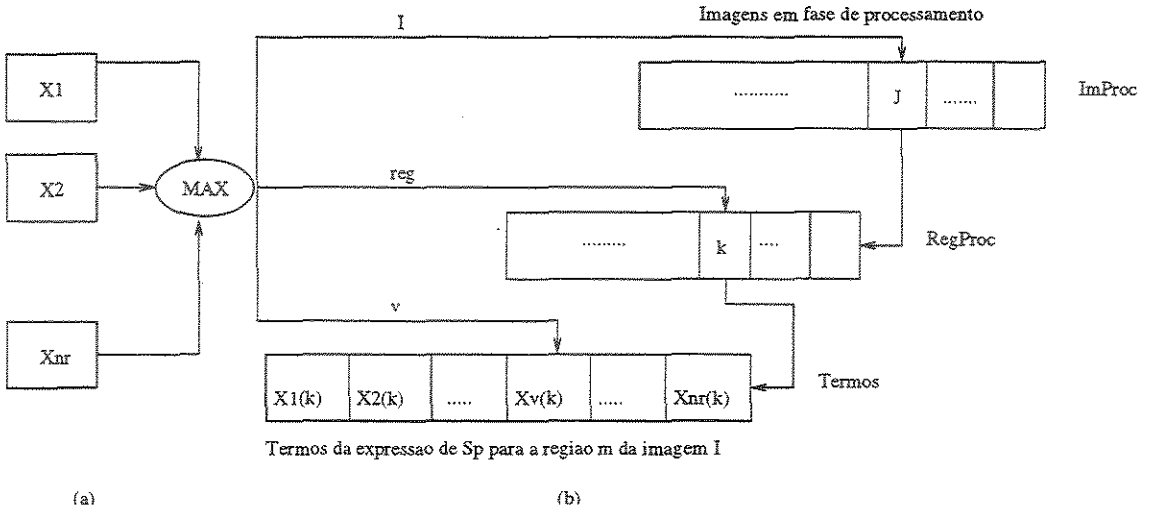


Figura 5.4: Processamento da similaridade intra-padrão: (a) inicialmente é selecionada a região com maior valor de similaridade dentre as Listas de Representação. (b) para esta região são atualizadas as estruturas $ImProc$, $RegProc$ e $Termos$

$$m \cdot ((\theta_m^j - \theta_{m+1}^j) \cdot \max_{1 \leq i \leq m} \{s_i(reg_k^i[J], VR_i[j])\} \tag{5.4}$$

indica que para seu cálculo é preciso obter os valores de similaridade da região $reg_k[J]$, isto é, $s_i(reg_k^i[J], VR_i[j])$, para todos os MR $\{R_1, \cdots, R_m\}$. Portanto, o cálculo da similaridade intra-padrão de uma região $reg_k[J]$ já armazenada na estrutura $ImProc$ requer os valores de similaridade com todos os MR. Considerando a estrutura de dados aqui proposta, isto implica que todos os valores associados a $Termos$ estejam previamente definidos.

O objetivo de armazenar na m-ésima entrada de $Termos$ o valor de $s_m(reg_k^m[J], VR_m[j])$ é permitir calcular todos os termos da expressão de similaridade.

Dada uma região $r_k[J]$, dizemos que a entrada $ImProc[J].RegProc[k].Termos[m]$ está **definida** quando o valor de $s_m(reg_k^m[J], VR_m[j])$ estiver determinado.

Dizemos que a região $r_k[J]$ está **completa** quando todas as entradas da sua estrutura $Termos$ estão definidas. Uma região completa caracteriza aquela para a qual é possível calcular a similaridade intra-padrão. Este valor será armazenado em $ImProc[J].RegProc[k].Sp$.

Vejamos como a definição das entradas para uma região pode ser facilitada pelo fato de ser *máximo* a função utilizada na equação de similaridade intra-padrão (Equação 5.2).

Como explicado anteriormente, em cada passo do algoritmo, a região $X_v^j.reg$ sendo processada é aquela de maior similaridade dentre todas as listas, isto é, $Head(X_v^j).sim = \max_{1 \geq i \geq nr}\{Head(X_i^j).sim\}$. Chamemos de RM_v^k a esta região O valor de similaridade da região RM_v^k segundo o MR R_v é maior que o valor de similaridade em todas as listas

seguintes e neste caso temos que $Head(X_v^j).sim \geq Head(X_m^j).sim$ para todo $m > v$. Este fato garante ainda que a similaridade segundo o MR R_v dessa região RM_v^k é maior que a similaridade dessa mesma região segundo os MR associados às listas seguintes cujos termos não estejam definidos, isto é, $s_v(RM_v^k, VR_m[j]) \geq s_m(RM_v^k, VR_m[j])$ para todo $m > v$ onde $Termos[m]$ não esteja definido. Então, necessariamente temos que o m-ésimo termo da equação de similaridade 5.2, para a região RM_v^k, dada por:

$$m \cdot (\theta_m^j - \theta_{m+1}^j) \cdot \max_{1 \leq i \leq m} \{s_i(RM_v^k, VR_i[j])\} \tag{5.5}$$

pode ser escrito como:

$$m \cdot (\theta_m^j - \theta_{m+1}^j) \cdot \max_{1 \leq i \leq v} \{s_i(RM_v^k, VR_i[j])\} \tag{5.6}$$

isto é, o valor da similaridade para o MR R_m não tem nenhuma relevância no cálculo da similaridade intra-padrão dessa região, já que o fato de existir um valor maior que ele, faz con que o mesmo nunca seja escolhido pela função *máximo*.

Com esta propriedade, garantimos que para a região RM_v^k sendo processada, é possível estabelecer como **definidas** todas as entradas não definidas das listas seguintes a v. Em outras palavras, quando a região RM_v^k é inserida na estrutura $ImProc[J].RegProc[k].Termos[v]$, além de definir a entrada correspondente a $Termos[v]$ também pode definir todas suas entradas seguintes.

Desta maneira, os valores das entradas de $Termos$ ainda não definidas da região RM_v^k, ou seja,

$$ImProc[J].RegProc[RM_v^k].Termos[m], \tag{5.7}$$

tais que $m > v$, podem ser substituídos pelo valor

$$ImProc[J].RegProc[RM_v^k].Termos[v] \tag{5.8}$$

Aa partir deste resultado, elimina-se a necessidade de obter os valores de similaridade da região RM_v^k para as listas posteriores. Este resultado corresponde a uma poda na árvore de similaridade (Figura 4.8) associada ao cálculo da similaridade global de uma imagem.

A figura 5.5 ilustra este processo: (a) A região selecionada à esquerda como a de maior valor entre as listas de representação (RM_v^k) é utilizada para atualizar as estruturas de dados Em (b) com o valor de similaridade dessa região é possível atualizar todas as entradas seguintes da estrutura $Termos$ associada a essa região.

A partir desta redução importante para o cálculo, o próximo passo é determinar qual região, dentre aquelas sendo processadas, é s de maior valor de similaridade intra-padrão e qual é este valor.

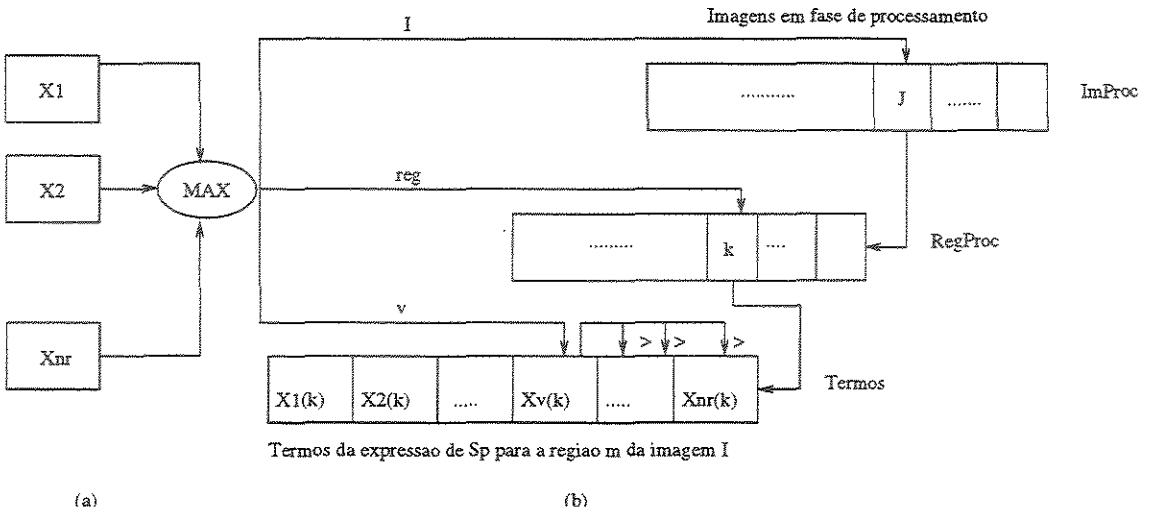


Figura 5.5: Substituição de termos do cálculo da expressão de similaridade intra-padrão: (a) é selecionada a região com maior valor de similaridade dentre as listas de Representação. (b) com a região selecionada são atualizadas todas as entradas da estrutura *Termos* associada a essa região

Quando uma região é **completa**, seu valor de similaridade pode ser totalmente calculado, mas ela não é necessariamente a de maior valor de similaridade intra-padrão. Pode acontecer que alguma outra região ainda não **completa** tenha maior valor de similaridade. No entanto, é possível determinar se uma região completa é a região de maior similaridade. Na estrutura $ImProc[J]$, podemos guardar para cada região **não completa** o maior valor de **similaridade possível** que poderia atingir. Para este cálculo é suficiente substituir os valores das entradas incompletas pelo valor de similaridade das regiões que estão no inicio das listas associadas a essas entradas. O cálculo realizado com esses valores estabelece um limite superior do valor real de similaridade intra-padrão dessa região.

Desta maneira, dizemos que uma região **completa** é **maximal** se seu valor de similaridade é maior que todos os valores de **similaridade possível** das regiões **não completas** ainda sendo processadas. Quando uma região é **maximal** então ela corresponde àquela com maior valor de similaridade e, portanto, a correspondente ao resultado.

Como foi apresentado no início desta seção, a lista de similaridade intra-padrão Sp_j é construída por demanda. Por tanto, cada chamada à função *Next* deve determinar a próxima região da lista. Esta região vai corresponder em cada momento, àquela região com maior valor de similaridade intra-padrão dentre as que estão sendo processadas.

A partir destas considerações, uma descrição geral da atualização das listas Sp_i pelo procedimento *Next* é apresentada no algoritmo 5.8. Note-se que o procedimento *Next* unicamente determina qual a região com maior similaridade, o valor desta similaridade, e a imagem à que esta região pertence. Estas informações permanecem nas variáveis

---

Algoritmo 5.8 (Próxima região na lista de similaridade intra-padrão Sp_j)

Forma Geral: $Next(Sp_j, \Theta_j)$
Entrada:
Sp_j: Lista de similaridade intra-padrão Sp_j
Θ_j): Pesos para o padrão p_j
Passos:
Enquanto não existe uma imagem I com alguma região Rc *completa e maximal*
$mx = max(Head(X_1^j), \cdots, Head(X_m^j), \cdots, Head(X_{nr}^j))$
$(reg, J, sim) = Head(X_{mx}^j)$
$Next(X_{mx}^j)$
$Atualizar(RM_j^i, s, reg, I)$
Inserir- Atualizar (reg, J, sim) nas estruturas $ImProc[J]$,
$ImProc[J]RegProc[reg]$ e $ImProc[J].RegProc[reg].Termos[mx]$
Completar todas as entradas de $ImProc[J].RegProc[reg].Termos[mx]$
maiores a mx
Seja I a imagem da região Rc completa e maximal
$ImProc[I].RegProc[Rc].Sp \leftarrow Mp_{\Theta_j}(ImProc[I].RegProc[Rc])$
CurrentReg = Rc
CurrentIm = I
CurrentSim = $ImProc[J].RegProc[Rc].Sp$

$(CurrentReg, CurrentIm, CurrentSim)$ para seu uso quando a função $Head$ seja chamada. Uma próxima chamada a $Next$ desencadeia um novo processo de busca da região com maior similaridade. Este novo processo utiliza todas as estruturas do passo anterior, incluindo todas as regiões sendo processadas. Desta maneira, os resultados temporários de uma chamada mantém validade para a próxima.

A apresentação realizada nesta seção esteve centrada na determinação da próxima região de uma lista de similaridade intra-padrão. Por cada padrão existe uma lista deste tipo Sp_j e todas elas vão sendo processadas de acordo com a demanda. Como mencionado, a lista de similaridade global Sg combina os resultados de todas estas listas Sp_j em um resultado final. A próxima seção apresenta a implementação desta lista Sg.

5.5.4 Lista de similaridade global

As imagens que o algoritmo deve retornar como resultado da consulta são aquelas com maior valor do mínimo entre todas as similaridades intra-padrão (Equação 4.7) de todas as imagens do banco de dados.

Analisando a expressão na referida equação , vemos que para completar o cálculo da similaridade global de uma imagem, é necessário obter todos seus valores de similaridade intra-padrão para que o mínimo destes valores possa ser definido. Em outras palavras, para ter seu cálculo de similaridade global completo, uma imagem deve ter aparecido em todas as listas de similaridade intra-padrão para que todos os termos da expressão de similaridade sejam obtidos.

O fato de utilizar a função min neste nível do cálculo implica que não é possível fazer podas na árvore de similaridade pois precisamos de todos os valores em todas as listas para a realização do cálculo.

A lista de similaridade global Sg ordena as imagens do banco de dados segundo sua similaridade global. Como resultado, as N primeiras imagens desta lista correspondem à resposta à consulta.

Para esta lista existem duas operações fundamentais:

- $Head(Sg, \Theta) \rightarrow (\{regs\}, J, sim)$: devolve o conjunto de regiões $\{regs\}$ mais similares com cada padrão, da imagem J que estiver o início da lista juntamente com o seu valor de similaridade global Durante a construção por demanda desta lista, esta imagem é em cada momento a de maior similaridade global em todo o banco de dados.

- $Next(Sg)$: Atualiza o início da lista com a próxima imagem de maior similaridade global.

Para a implementação desta lista são utilizadas um conjunto de estruturas para armazenar as imagens e regiões sendo processadas em cada momento. Estas estruturas são:

- $ImProcG[J]$ é a estrutura que contém as imagens sendo processadas.

- $ImProcG[J].Termos[k]$ contém o valor de similaridade intra-padrão da imagem J para o padrão p_k

- $ImProcG[J].Reg[k]$ é a região da imagem onde a maior similaridade intra-padrão p_k é atingida.

A idéia do processamento da lista de Sg é extrair em cada momento a região da lista de similaridade intra-padrão cujo elemento inicial tem o maior valor de similaridade.

De maneira análoga à seção anterior, dizemos que uma entrada $ImProcG[J].Termos[k]$ é **definida** se o valor da similaridade intra-padrão correspondente foi determinado. Este valor é determinado se alguma região da imagem J foi processada e obtida da lista de similaridade intra-padrão Sp_k.

Quando todas as entradas para uma imagem estão definidas, então esta imagem é **completa** e pode ser selecionada como a imagem de maior similaridade global dentre todas as processadas. Como a similaridade global se define a partir da função mínimo, alguma outra imagem posteriormente completa não poderia atingir um valor de similaridade maior do que a imagem selecionada.

A figura 5.6 descreve esta estrutura de dados e processamento. A imagem processada é extraída como a associada à região de maior similaridade dentre todas as listas de similaridade intra-padrão. O valor de similaridade desta região é armazenado como valor de similaridade da imagem para o padrão correspondente.

Assim como a lista de similaridade intra-padrão, a lista de similaridade global Sg é implementada por demanda. Por esta razão, o processamento do procedimento $Next$ corresponde a determinar qual a próxima imagem de maior similaridade global.

O algoritmo 5.9 resume o processamento deste procedimento $Next$.

Note-se que cada chamada a $Next$ implica a determinação da próxima imagem de maior similaridade, juntamente com seu valor de similaridade e as regiões dentro dessa imagem selecionadas como as mais similares com cada um dos padrões. Esta informação é armazenada nas variáveis $CurrentJ, CurrentS, CurrentR$ para serem acessadas nas chamadas do procedimento $Head$.

A partir da definição da função $Next$ para a lista de similaridade global, pode ser construído o algoritmo final 5.10 de determinação das N imagens mais similares do banco de dados. Como vemos , a partir da abordagem apresentada o cálculo das N imagens mais similares se produz pela chamada N vezes da função $Head$ da lista de similaridade global (Algoritmo 5.9). Esta chamada desencadeia chamadas às outras listas de maneira

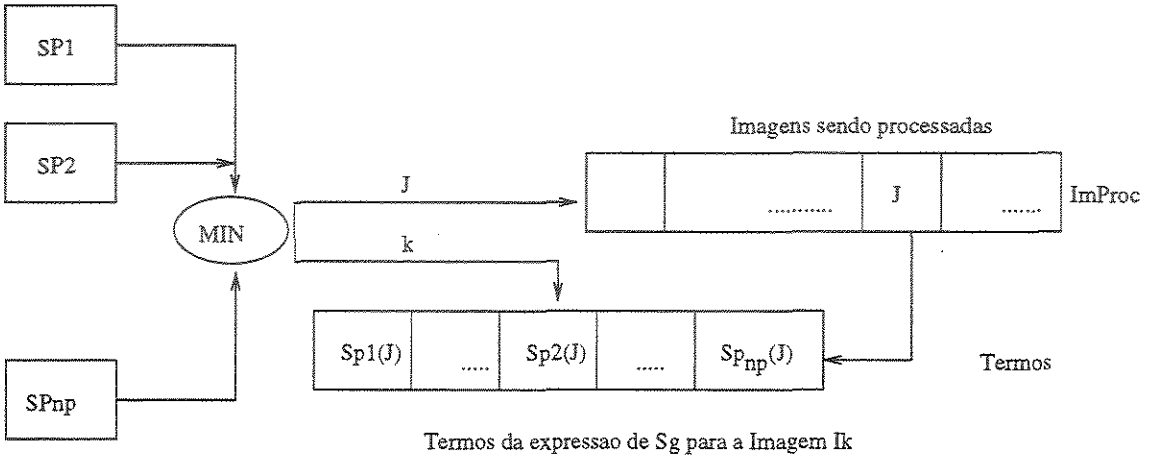


Figura 5.6: Processamento da similaridade global Lg. A região de maior similaridade entre as listas de similaridade intra-padrão é armazenada nas estruturas ImProc e Termos. Quando todos os termos de uma região são completados, esta imagem é candidata a ser a próxima na lista de similaridade global

---

Algoritmo 5.9 (Próxima imagem na Lista de similaridade global)

Forma Geral: $Next(Sg, \Theta)$
Entrada:
Sg: `Lista de similaridade Global`
Θ): `Pesos para o processamento`
Passos:
Enquanto `não existe uma imagem` J `completa`
`e seu valor de` $ImProc[j].min$ `maior que` $max(Head(Sp_i))$
$m = \max\{Head(Sp_1), Head(Sp_2), \cdots, Head(Sp_i), \cdots, Head(Sp_{NP})\}$
$(r_m, J_m, s_m) = Head(Sp_m)$
$Next(Sp_m)$
`Inserir/Atualizar` (r_m, J_m, s_m) `em` $ImProcG$
`Seja` J `a imagem completa`
$CurrentJ \leftarrow J$
$CurrentS \leftarrow ImProc[J].min$
$CurrentR = \{ImProc[J].reg[m]\}$ `/*todas as regiões de todos os padrões`

Algoritmo 5.10 (Cálculo das N imagens mais similares (versão final))

Forma Geral $Sg(P_C, N, RU_j, \{RS_j\}, \{RM_j^i\}, flag) \rightarrow (Jr, Sr, Rr, \{RS_j\}, \{RM_j^i\})$
Entrada:

P_C: parâmetros de consulta resultantes da etapa 1
N: número de imagens na resposta
RU_j: Conjunto das regiões relevantes para o usuário em cada padrão na iteração anterior
$\{RS_j\}$: Conjuntos das respostas do sistema para cada padrão na iteração anterior
$\{RM_j^i\}$: Conjuntos das regiões selecionadas por cada MR R_i como mais similares com cada padrão na iteração anterior
flag : Indica se esta chamada é a primeira iteração da consulta

Saída:

(Jr, Sr, Rr): Similar ao apresentado no algoritmo 5.2
$\{RS_j\}, \{RM_j^i\}$: Atualizações dos conjuntos de regiões selecionadas

Passos:
Se flag
$InicializarPesos(\Theta)$
senão
$AtualizarPesos(Selec, Relev, \Theta)$
Criar as listas Sg, $\{Sp_j\}$ e $\{X_j^i\}$
$InicializarSelec(\{RS_j\}, \{RM_j^i\})$
Repetir N vezes
$(\{R\}, J, S) \leftarrow Head(Sg, \Theta)$
Inserir J em Jr
Inserir S em Sr
Inserir {R} em Rr
Determinar os conjuntos $\{RS_j\}$ como o conjunto das regiões das imagens em Jr mais similares a cada padrão p_j.
Determinar os conjuntos $\{RM_j^i\}$ de cada MR R_i mais similares a cada padrão p_j.

return (Jr, Sr, Rr)

---

progressiva e por demanda. Esta característica garante que o processamento realizado é o estritamente necessário para obter as N imagens requisitadas na consulta.

5.6 Resumo

Este capítulo discutiu as idéias gerais da implementação do modelo de recuperação por conteúdo proposto. Inicialmente, apresentou uma arquitetura para o sistema, com os seus módulos principais. Posteriormente, foram apresentados os algoritmos básicos para o processamento de uma consulta. A implementação destes algoritmos baseia-se na utilização das estruturas de índice espacial em que são armazenados os descritores das regiões das imagens no banco de dados, assim como características da função de similaridade adotada. A partir das estruturas de índice, várias listas em vários níveis são construídas usando o principio de inserção por demanda. O algoritmo proposto permite diminuir significativamente a computação necessária para o processamento de uma consulta.

Capítulo 6

Ilustração do modelo e problema de normalização

6.1 Introdução

Este capítulo ilustra o modelo proposto, com a definição de dois Modelos de Representação e a solução do problema de normalização associado a eles. Para textura foi adotada a Transformada *Wavelet* de Haar e para a representação da cor, os momentos de histograma. São descritas as diferentes funções para estes MR, assim como o critério de normalização adotado para permitir a comparabilidade dos MR. Estes modelos foram parcialmente implementados. Como parte do capítulo a seção 6.2 apresenta um protótipo de interface.

6.2 Protótipo de Interface

Como parte do trabalho foi definido um protótipo de interface que oferece facilidades mínimas para a definição de três padrões.

A figura 6.1 mostra a interface proposta para o sistema. A lista da esquerda corresponde às imagens de consulta. Na região central da tela o usuário amplia imagens de consulta, selecionando amostras de regiões para padrões. A associação de uma região com um padrão é feita através de uma cor associada a cada padrão.

A figura 6.2 mostra como depois de realizada a consulta, o sistema devolve sua resposta na região da tela à direita. A região central é utilizada, neste caso, para ampliar as imagens da resposta, permitindo visualizar as regiões similares a cada padrão propostas pelo sistema. A priori, todas as regiões propostas pelo sistema são consideradas como confirmadas pelo usuário. Se o usuário marca alguma região, então o sistema a retira do conjunto das regiões do padrão associado. Uma vez explorada e analisada a resposta do

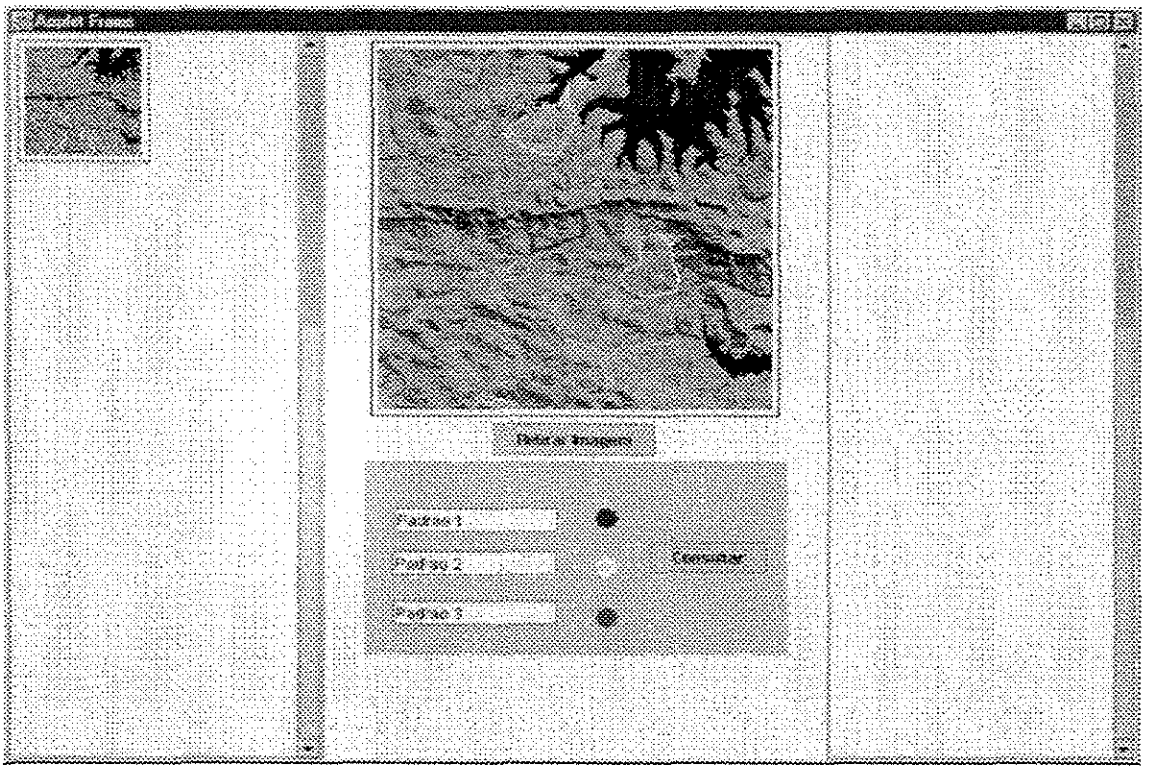


Figura 6.1: Interface do protótipo do sistema de recuperação por conteúdo durante uma consulta: A coluna da esquerda mostra uma imagem de consulta. O espaço central serve para as interações do usuário: aumentar uma imagem escolhida para marcar regiões na consulta.

sistema, o usuário solicita uma nova iteração no processo de consulta.

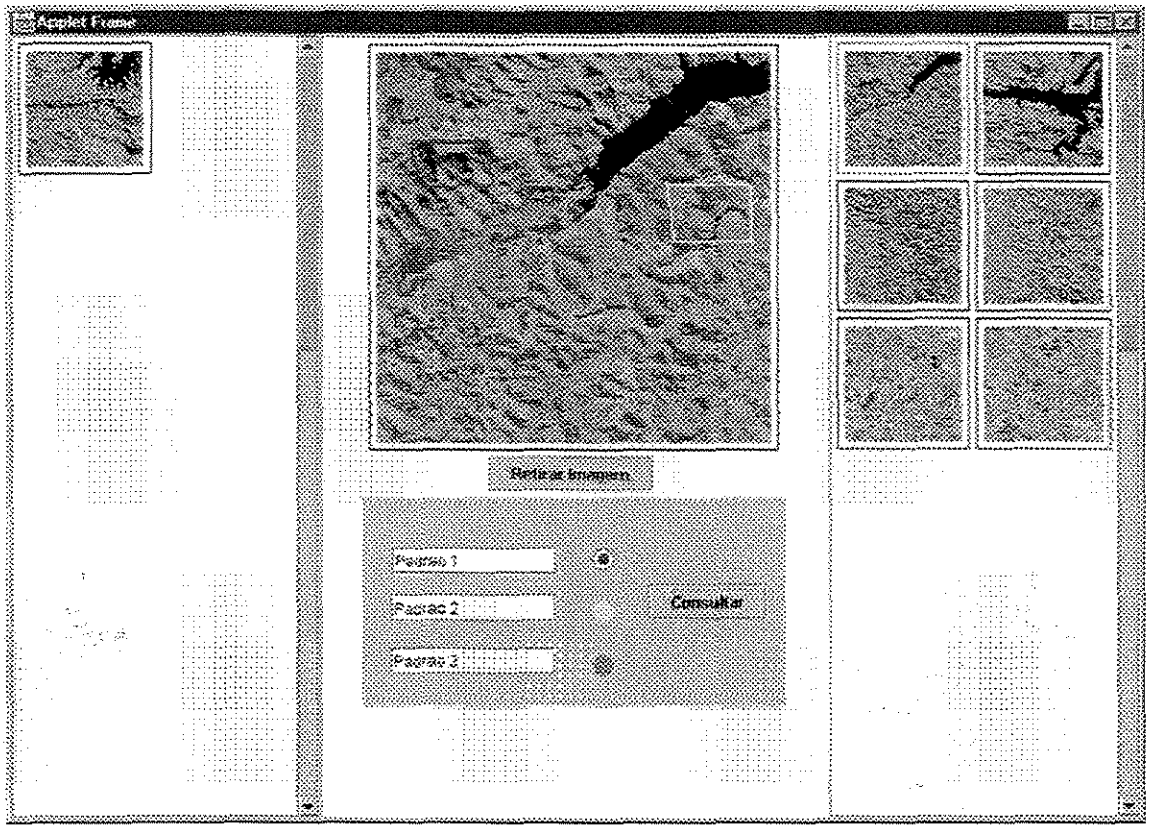


Figura 6.2: Interface do protótipo do sistema de recuperação por conteúdo durante a resposta do sistema: a coluna da direita mostra os resultados obtidos pelo sistema. O espaço central permite ao usuário avaliar a resposta, marcando aquelas regiões consideradas como incorretas.

6.3 Banco de dados utilizado

Para este exemplo foi utilizadas um conjunto de 150 ISR LANDSAT correspondentes às bandas 3, 4 e 5. A primeira corresponde à banda visual vermelha e as outras duas correspondem a intervalos no espectro infra-vermelho próximo. Estas imagens foram segmentadas manualmente em regiões com textura e cor similares, criando assim o banco de dados de regiões.

O processamento foi implementado utilizando um algoritmo de processamento de consulta que percorre todo o banco de dados. Esta implementação visou apenas constatar o funcionamento dos algoritmos, sem preocupação com desempenho.

Para cada MR adotado no protótipo, foi criado um arquivo contendo informações de cada região, com o número da imagem do banco de dados à qual pertence, seu retângulo mínimo envolvente e o vetor característico correspondente.

Para facilitar o processamento, o arquivo foi ordenado pelo número da imagem, o que permite um processamento seqüencial simultâneo nos dois arquivos dos MR.

6.4 Modelos de Representação

Para ilustrar o modelo proposto foram utilizados dois MR: a transformada Wavelet para textura e momentos dos histogramas para o caso de cores. Estes modelos possuem a propriedade de que o processamento das imagens envolvido é eficiente. A métrica de similaridade utilizada foi a distância euclidiana L_2. Ressalte-se que o intervalo de valores que pode tomar cada elemento do vetor característico pode ser diferente, e portanto, foi aplicado um processo de normalização descrito na seção 6.5.

Nas próximas seções são apresentadas as diferentes funções componentes segundo a definição de Modelo de Representação em 4.4.1.

6.4.1 Modelo de Representação 1: Transformada Wavelet

Como apresentado na seção 2.6.4, a *Transformada Wavelet* corresponde a uma aproximação da imagem em múltiplas escalas. Cada aplicação da transformada produz a imagem original reduzida em escala, e mais três imagens de detalhes com a informação filtrada correspondente a orientações diferentes em uma faixa de freqüência. Cada imagem contendo a filtragem em uma faixa de freqüência preserva a informação espacial e está associada a uma orientação: vertical, horizontal ou diagonal.

No exemplo aqui apresentado, foi considerada a transformada de Haar [46] aplicada com duas iterações em cada uma das três bandas das imagens, obtendo-se, assim, dois trios de imagens de detalhe. Dado que a transformada preserva a informação espacial, embora com redução de escala, é fácil determinar em cada imagem aproximada quais os pixels que correspondem a uma região da imagem original.

Para a nossa aplicação, a informação de orientação não é muito útil porque o mesmo padrão de textura pode aparecer em imagens diferentes com orientações diferentes. Neste caso, os valores em cada orientação seriam diferentes, considerando a sensibilidade da transformada à rotação. No entanto, todas as imagens de detalhe, de uma mesma aplicação da transformada, estão associadas ao mesmo intervalo de freqüência, embora as informações apareçam nas diferentes imagens. Por esta razão, processamos de maneira conjunta, em cada região, toda a informação em cada uma das imagens de filtragem.

A seguir são apresentadas as diferentes funções associadas a este modelo de representação.

Função de extração de características

A função de extração de características E_1 para o modelo da transformada *Wavelet* constrói um vetor que descreve a informação extraída para uma região. Inicialmente, consideramos duas iterações da transformada sobre a imagem original. Posteriormente, para cada região segmentada são processados os pixels contidos na região das três imagens de detalhe de cada iteração (filtragens nas direções vertical. horizontal e diagonal). São calculadas a média e variância dos valores nessas imagens. Este processo se repete para a segunda iteração da transformada. Como este processo é aplicado a cada banda da imagem obtém-se, finalmente, um vetor característico de 12 elementos, isto é, a média e variância das duas iterações da transformada, para cada banda da imagem. O vetor resultante para a imagem $r[I]$ tem a forma $W(r[I]) = (M_1^1, V_1^1, \cdots, , M_2^3, M_2^3)$, sendo M_j^i e V_j^i a média e variância dos valores a j-ésima iteração da Transformada para a i-ésima banda.

Função de distância

Como função de distância foi adotada a distancia L_2 após normalização da cada dimensão do vetor, como explicado posteriormente na seção 6.5.

Função de Parametrização

Seja $\{W_1, W_2, \cdots, W_n\}$ o conjunto de vetores característicos associados a um padrao p_j. A função de parametrização T_1, utilizada para o MR baseado em Wavelets é vetor W_M que possui a menor distancia média de todos os vetores da classe, isto é, aquele vetor mais próximo do centro de massa associado à coleção de vetores que descrevem o padrão.

Formalmente, W_M é calculado como

$$\min_{\{W_1, W_2, \cdots, W_n\}} \{L_2(W_i, W_C)\} \tag{6.1}$$

sendo W_C o centro de massa do conjunto $\{W_1, W_2, \cdots, W_n\}$.

6.4.2 Modelo de Representação 2: Momentos do Histograma

O outro MR utilizado é o conjunto de momentos do histograma. Uma região segmentada de um ISR, em geral, deve ser homogênea em relação à cor, tanto no momento da sua inserção no Banco de Dados, como da sua seleção como uma amostra de padrão. Esta

premissa implica em que o histograma associado a uma região deve possuir um pico bem definido ao redor da cor predominante. Considerando este fato, selecionamos os momentos do histograma por ser um modelo compacto para descrever a cor de uma região, no caso em que não existe grande variação de cores no histograma.

As funções associadas a este modelo de representação são apresentadas a seguir:

Função de extração de características

A função de extração de características gera um vetor composto por momentos associados a cada uma das bandas das imagens, considerando somente os pixels da região. Inicialmente, é construído o histograma em cada uma das bandas, para cada região processada. Posteriormente, para cada histograma, são calculados os momentos: média (M), variância (V) e simetria(S). Como resultado, para cada região é definido um vetor característico de 9 elementos do tipo. $H = (M^1, V^1, S^1, \cdots M^3, V^3, S^3)$

Função de distância

Para o caso do MR por momentos de histograma, utilizamos igualmente a métrica euclidiana L_2. Note-se que as diferentes elementos do vetor devem ser quantizadas de modo que a influência de cada um deles no resultado de distância seja proporcional ao intervalo de valores válidos para esse elemento.

Função de Parametrização

É natural entender o conjunto de vetores característicos das regiões de um padrão $\{H_1, \cdots, H_n\}$ como elementos de uma classe. Neste caso, seria possível considerar o uso de técnicas tradicionais de classificação para definir um vetor representativo da classe. Esta abordagem, pode introduzir valores artificiais, não resultantes de nenhuma região especifica. Por esta razão, a função de parametrização adotada usou uma abordagem similar ao caso do MR *Wavelet*. Selecionamos como vetor de parametrização, aquele H_M dentre os vetores do conjunto $\{H_1, H_2, \cdots, H_n\}$ cujos momentos de tipo média são os mais próximos aos valores médios dentro do conjunto. Em outras palavras selecionamos

$$\min_{\{H_1, H_2, \cdots, H_n\}} \{L_2((M_i^1, M_i^2, M_i^3), (MED^1, MED^2, MED^3))\} \tag{6.2}$$

sendo MED^j o valor médio do momento média na banda j em todos os vetores característicos do conjunto.

6.5 Função de Normalização

Como mencionado na seção (seção 4.4.2), para combinar as similaridades de diferentes MR é necessário normalizar suas funções. Esta normalização deve equalizar os valores de distância de cada MR para intervalo de similaridade $[0, 1]$. Uma distância igual a 0 é interpretada como máxima similaridade e portanto para efeito de normalização deve ser transformada num valor de similaridade 1. Adicionalmente, o vetor de características de cada MR possui dimensões com distintas distribuições de valores, implicando em intervalos de valores diferentes. Por exemplo, como resultado do processamento, médias e variâncias podem ter valores em intervalos diferentes, influenciando de maneira distinta o valor da métrica L_2 adotada.

Seja $X = (x_1, x_2, \cdots, x_n)$ um vetor característico de qualquer um dos MR adotados (wavelets ou momentos). Se cada uma das variáveis x_i do vetor é quantizada no intervalo $[0, 1]$, garante-se que o efeito de cada uma delas é aproximadamente igual sobre a métrica adotada.

Para garantir isto, adotamos a abordagem de transformar cada variável x_i do vetor característico em uma variável aleatória $\tilde{x}_i$ de média zero e variância 1, isto é, transformamos cada elemento do vetor x_i segundo :

$$\tilde{x}_i = \frac{x_i - \mu_i}{\sigma_i} \tag{6.3}$$

sendo μ_i e σ_i a média e o desvio quadrático médio da variável x_i.

Com esta transformação, é conhecido que existe 68% de probabilidade de que o valor de $\tilde{x}_i$ tenha seus valores no intervalo $[-1, 1]$. Uma transformação adicional da forma:

$$\tilde{x}_i = \frac{\frac{x_i - \mu_i}{3\sigma_i} + 1}{2} \tag{6.4}$$

garante, com 99% de probabilidade, que os valores da variável $\tilde{x}_i$. estão no intervalo $[0, 1]$ se supusermos que a variável está distribuída segundo a distribuição normal. Aqueles valores das variáveis que eventualmente estejam fora do intervalo $[0, 1]$ são truncados.

A determinação de μ_i e σ_i é feita da seguinte forma. Inicialmente, são processadas todas as regiões da amostra inicial do banco de dados, segundo os diferentes MR obtendo-se um vetor característico para cada região. Posteriormente, para cada dimensão do vetor característico, isto é, para cada variável x_i, são determinados os valores de μ_i e σ_i considerando os valores nessas variáveis dos vetores característicos de todas as regiões do banco de dados. Finalmente, é realizado o processo de normalização aplicando a Equação 6.4.

Como resultado deste processo, o vetor $\tilde{X}$ tem todas suas variáveis $\tilde{x}_i$ normalizadas no intervalo $[0, 1]$.

Como passo final do processo é necessária a normalização da métrica de cada modelo de representação para obter a função de distância s_j.

Como foi mencionado, a métrica adotada em ambos modelos foi L_2. Os valores de distância entre dois vetores normalizados utilizando esta métrica devem ser transformados no intervalo $[0, 1]$ sendo 1 o valor de maior similaridade.

Note-se que a partir da métrica L_2 adotada, o maior valor de distancia possível é aquele entre os vetores $(0, 0, \cdots, 0)$ e $(1, 1, \cdots, 1)$. Aplicando a métrica L_2 entre estes vetores, o resultado é $\sqrt{N}$ quando o processo de normalização deveria transformar em 0, isto é, a menor similaridade possível. Partindo deste fato, a função de distância normalizada s_j pode ser escrita como:

$$s_j(X_1, X_2) = 1 - (\frac{L_2(\tilde{X}_1, \tilde{X}_2)}{d_M}) \tag{6.5}$$

sendo $d_M = d_j((0, 0, \cdots, 0), (1, 1, \cdots, 1))$ a maior distancia entre vetores característicos. Note-se que esta transformação garante que uma distância 0 entre dois vetores se transforma em uma similaridade 1.

6.6 Resumo

Este capítulo apresentou um exemplo de implementação do modelo proposto nesta tese. Este protótipo permite ao usuário interagir com as imagens a partir de uma interface para apresentação de imagens de consulta e imagens resultantes. O protótipo inclui a definição de dois Modelos de Representação correspondentes à transformada *Wavelet* para descrição de textura e momentos de histograma para a cor. Adicionalmente, foi apresentado o critério de normalização das funções de similaridade adotado no protótipo.

Capítulo 7

Conclusões e Extensões

7.1 Conclusões

Esta tese apresentou um modelo de recuperação por conteúdo de Imagens de Sensoriamento Remoto (ISR). Este tipo de imagens possui características que justificam a busca de modelos específicos para seu tratamento.

Para a definição deste modelo foram analisados diferentes problemas da recuperação baseada em conteúdo, técnicas de processamento de imagens utilizadas nesta área, assim como as características particulares de vários sistemas existentes para este tipo de recuperação. A seguir foram estudadas as características das ISR e os problemas que estas apresentam para sua recuperação baseada em conteúdo. Constatou-se que estas características complicam o processo de recuperação.

Este estudo inicial permitiu identificar, dentre outras, três premissas básicas para a recuperação de ISR: (a) a ocorrência de um mesmo fenômeno em imagens de diferentes locais e épocas, ainda com alterações esperadas, pode conservar similaridade em textura e cor que permite ao especialista relacioná-las. (b) existe dificuldade em achar um modelo matemático universal de descrição de conteúdo de ISR . (c) as propriedades das ISR e do seu processamento fazem com que a similaridade entre padrões e objetos dentro delas seja dependente de cada especialista e mutável no tempo.

A partir destas premissas, foi apresentado o modelo de recuperação baseado em três elementos básicos: (1) padrões como conceitos de consulta, (2) uso de múltiplos modelos matemáticos de descrição do conteúdo e (3) um mecanismo de retroalimentação.

Um padrão foi definido como um conjunto de regiões de imagens associado a um mesmo conceito ou entidade do mundo real. A importância do padrão no modelo é que ele expressa a idéia de que um mesmo conceito pode ter várias formas de manifestação dentro das ISR. As consultas são realizadas buscando-se imagens que contenham regiões similares dentro desta definição.

O segundo elemento da proposta foi a utilização de múltiplos modelos matemáticos para representar o conteúdo das imagens. Isto implica em permitir o uso simultâneo de diversas técnicas conhecidas para a caracterização da textura e a cor em ISR. Se alguma dessas técnicas considera uma imagem relevante para o resultado da consulta, o sistema considera este resultado.

A vantagem de múltiplas representações é permitir o uso cooperativo de diversos resultados na área de processamento de imagens para caracterizar o conteúdo de ISR. Ao contrário de outros trabalhos na área de recuperação por conteúdo, a ênfase desta proposta não está em apresentar um modelo matemático específico para descrever o conteúdo, mas em permitir o uso simultâneo de vários deles.

O terceiro elemento importante da proposta é a introdução de um mecanismo de retroalimentação para a realização das consultas. A inclusão deste mecanismo permite definir uma consulta como um processo iterativo de refinamento do resultado. Para o usuário, a retroalimentação ocorre com a confirmação da validade dos resultados apresentados pelo sistema. Esta confirmação implica em uma redefinição de quais regiões são consideradas parte de um padrão. Para o sistema, a retroalimentação se produz com a realização de uma nova consulta utilizando uma padrões refinados e a associação de pesos aos diferentes modelos matemáticos de representação. A tese propõe um critério de cálculo dos pesos que visa a busca daqueles modelos mais próximos do senso de similaridade de um usuário específico.

Como parte da proposta foi introduzida uma função de similaridade que utiliza a abordagem de Fagin [17] de funções de pontuação. Esta função de similaridade constitui a base do processamento de uma consulta. A definição desta função combina os três elementos básicos do modelo mencionados anteriormente.

Como conseqüência direta do modelo, a tese propõe parâmetros para avaliar a evolução de uma consulta. Estes parâmetros oferecem uma descrição quantitativa do grau de convergência entre o senso de similaridade do usuário e a definição de similaridade antes mencionada. Por outro lado, também permitem detectar quantitativamente comportamentos inadequados do sistema ou do usuário durante uma consulta, como, por exemplo, confirmações não coerentes de um usuário nas diferentes iterações.

Um último resultado da tese é a proposta de um algoritmo que reduz a complexidade do processamento de consultas baseado em estruturas de índice espacial e nas propriedades da função de similaridade proposta. Este algoritmo foi apresentado e não foi implementado, mas mostra a viabilidade do modelo proposto para bancos de dados com número significativo de imagens.

Finalmente, a tese ilustra um caso particular do modelo apresentado e discute o problema de normalização para este caso.

Uma das dificuldades maiores para a aplicação do modelo é a falta de algoritmos

de segmentação automatizados que produzam resultados efetivos. As propostas mais prometedoras para este problema são algoritmos de segmentação supervisionados que sempre envolvem a participação humana. A aplicação do modelo em bancos de dados massivos sempre teria como limitação a capacidade humana de processar, do ponto de vista de segmentação, o número cada vez maior de ISR geradas.

As principais contribuições desta tese são, assim:

- Estudo dos problemas e particularidades das ISR para a recuperação por conteúdo.
- Definição de um modelo de recuperação por conteúdo para ISR que considere estas características, assim como as particularidades do processamento destas imagens pelos especialistas.
- Definição de um modelo e métrica de similaridade baseados no uso de múltiplas representações de conteúdo, refinado iterativamente a partir da retroalimentação (*feedback*) de cada especialista.
- Uma proposta de implementação do modelo que mostra a viabilidade de sua utilização.

7.2 Extensões

Existem várias extensões a este trabalho. dentre elas podemos citar:

- A validação prática do modelo não foi totalmente realizada. O modelo precisa ser implementado com bancos de dados reais considerando diferentes fontes de ISR.
- Uma aplicação direta do modelo proposto é sua integração a um Sistema de Informações Geográficas (SIG). No contexto de um SIG, um mecanismo de recuperação como o proposto pode ser uma ferramenta auxiliar de muita utilidade para qualquer aplicação sobre ISR. Os recursos de composição, filtragem e visualização de imagens do SIG podem ser utilizados diretamente como parte da interface do sistema de recuperação.
- É necessário explorar outros modelos de representação. Dada a dificuldade de achar modelos de representação únicos, adequados para ISR, um sistema baseado no modelo proposto pode servir como plataforma de análise de diferentes modelos de representação.

  Os parâmetros de precisão apresentados no capítulo 4 podem ser a fonte desta análise. A proposta considera basicamente valores máximo e mínimos, no entanto,

é possível manter o registro histórico desses parâmetros para todos os modelos de representação durante uma sessão de consulta.

Por exemplo, um modelo que para vários usuários não consegue a convergência com o usuário mostra dificuldades para caracterizar o conteúdo de ISR. Este modelo não contribui para o processo de consulta e poderia ser substituído. Outra aplicação destes parâmetros é a análise das series históricas que eles produzem. Esta análise permitiria traçar perfis de usuários, determinação de "outliers" durante o processo de consulta, dentre outros estudos possíveis.

- A definição de MR proposta no trabalho sugere um projeto e construção do sistema de recuperação, com flexibilidade para a integração rápida e fácil de novos MR. Arquivos de descrição dos MR podem ser idealizados de maneira tal que incluir um novo MR não implique na re-codificação e re-compilação do sistema, mas na sua reconfiguração.

- O modelo de similaridade adotado baseia-se na utilização de funções de máximo e mínimo para as similaridades intra-padrão e global, respectivamente. O uso destas funções é relevante para o algoritmo de processamento de consultas proposto. No entanto, outras alternativas de funções de pontuação poderiam ser aplicadas, permitindo estudar outras possibilidades de combinação de similaridades.

- A definição de modelo de representação adotada no capítulo 4 considera uma função de parametrização que devolve um vetor de descrição do padrão, que chamamos de vetor de parametrização. No entanto, pode acontecer que a descrição de um padrão, utilizando um único vetor, não seja sempre adequada. Em alguns casos, os vetores que se conformam a um padrão poderiam formar mais de um cluster indicando duas possibilidades: (a) o usuário está agrupando em um mesmo padrão entidades diferentes, (b) a descrição de uma entidade em um padrão corresponde a uma classe multimodal, isto é, descrita por vários clusters.

  Para considerar este caso, a definição de modelo de representação poderia ser estendida modificando a função de parametrização T. A nova função T devolveria como resultado um conjunto de vetores que descrevem o conteúdo de padrão. Obviamente, esta extensão implicaria sua redefinição da função de similaridade proposta e do algoritmo de processamento de consultas.

- A proposta de cálculos de pesos permite garantir que aqueles modelos de representação mais efetivos conseguem aumentar sua relevância na resposta do sistema. No entanto, deve ser estudado em que condições bons modelos de representação não conseguem incluir seus resultados nesta resposta. Modelos de representação com

tendência a associar baixos valores de similaridade (i.e. juizes muito rigorosos), poderiam ser excluídos da resposta do sistema, embora realizando uma caracterização boa do conteúdo das imagens.

Referências Bibliográficas

[1] M. Adiba. STORM: An objetc-oriented multimedia DBMS. In K Nwosu, B. Thuraisingham, and P. Bruce, editors, *Multimedia Database Systems*, pages 47–85. Kluwer Academic Publishers, 1996.

[2] S. Aksoy and R. Haralick. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, 22(5):563–582, 2001.

[3] J. Anderson and M. Stonebraker. SEQUOIA 2000 Metadata Schema for Satellite Images. *ACM SIGMOD RECORD*, 23(4):42–48, December 1994.

[4] E. Arkin, L. Huttenlocher, D. Kedem, and Mitchell J. An efficiently computable metric for comparing polygonal shapes. In *Proc. First ACM-SIAM Symposium on Discrete Algorithms*, 1990.

[5] J. Barros, J. French, W. Martin, and P. Kelly. System for indexing multi-spectral satellite images for efficient content-based retrieval. In W Niblack and R. Jain, editors, *Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 228–237. SPIE, February 1995.

[6] J. Barros, J. French, W. Martin, P. Kelly, and J. White. Indexing multispectral images for content-based retrieval. In *23rd AIPR Workshop on Image and Information Systems*, October 1994.

[7] T. Baumann. Management of Multidimensional Discrete Data. *VLDB Journal*, 3(4):401–444, 1994.

[8] N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger. The R*-tree: An efficient and robust access method for points and rectangles. In *Proceedings of the 1990 ACM SIGMOD Conference on Management of Data*, 1990.

[9] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color and texture - based image segmentation using EM and its application to content-based image retrieval. In *Proc. of IEEE Intern. Conference on Computer Vision*, 1998.

[10] C. Carson, M. Thomas, S. Belongie, J. Hellerstein, and J. Malik. Blobworld: A system for region-based image indexing and retrieval. In *Proc. of International Conference on Visual Information Systems*, 1999.

[11] T. Chang and C. Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Transactions on Image Processing*, 2(4):429–441, October 1993.

[12] Y. Chen and E. Dougherty. Gray-scale morphological granulometric texture classification. *Optical Engineering*, 33(8):2713–2722, August 1994.

[13] P. Ciaccia, M. Partella, and P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. In *Proceedings of the Twenty-third International Conference on Very Large Data Bases*, pages 426–435, August 1997.

[14] D. Clausi. *Texture segmentation of SAR Sea Ice Imagery*. PhD thesis, University of Waterloo, 1998.

[15] A.P. Crosta. *Processamento Digital de Imagens de Sensoramento Remoto*. UNICAMP, 1992.

[16] E. Dougherty, J. Newell, and J. Pelz. Morphological texture-based maximum-likelihood pixel classification based on local granulometric moments. *Pattern Recognition*, 25(10):1181–1198, 1992.

[17] R. Fagin. Allowing users to weight search terms in information retrieval. Technical Report RJ10108, IBM Almaden Research Center, March 1998.

[18] R. Fagin. Fuzzy queries in multimedia database systems. Technical Report RJ10106, IBM Almaden Research Center, March 1998.

[19] R. Fagin and L. Stockmeyer. Relaxing the triagle inequality in pattern matching. *International Journal of Computer Vision*, 28(3):219–231, 1998.

[20] R Fagin and E. Wimmers. Incorporating user preferences in multimedia queries. In *Proceedings of Intern. Conference on Database Theory*, number 1186 in Lecture Notes in Computer Science, pages 247–261, 1997.

[21] M. Flickner. Query by image and video content: The QBIC System. *IEEE Computer*, September 1995.

[22] D. Gabor. Theory of communications. *Journal of IEE*, (93):429–457, 1946.

[23] Volker Gaede and Oliver Guenther. Multidimensional access methods. Technical Report TR-96-043, International Computer Science Institute, Berkeley, CA, October 1996.

[24] A. Gersho and R. e. *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, 1992.

[25] R. Gonzalez and R. Woods. *Digital Image Processing.* Addison Wesley, 1992.

[26] C. Gotlieb and E. Kreyszig. Texture descriptors based on co-ocorrence matrices. *Computer Vision, Graphics and Image Processing*, 51, 1990.

[27] R. Gray. Content-based image retrieval: Color and edges. Technical report, Department of Computer Science, Darmouth College, 1995.

[28] V. Gudivada and V. Raghavan. Content-based image retrieval systems. *IEEE Computer*, September 1995.

[29] A. Guttman. R-trees: A dynamic index structure for spatial searching. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, pages 47–57, 1984.

[30] R. Haralick. Statistical and structural approaches to texture. *Proceeeding of IEEE*, 67(5):786–804, 1979.

[31] R. Haralick, K. Shanmugham, and I. Dinstein. Texture features for image classification. *IEEE Trans. Syst. Man Cybern*, 3(6):610–621, 1973.

[32] F. Huet and J. Mattioli. A textural analysis by mathematical morphology. In J. Serra and P. Soille, editors, *Mathematical Morphology and Its Applications to Image Processing*, pages 297–304. Kluwer Academic Press, 1994.

[33] A. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[34] A. Jain and G. Healey. A multiscale representation including opponent color features for texture recognition. *IEEE Transactions on Image Processing*, 7(1):124–128, January 1998.

[35] R. Jain, S. Jayaram, and P. Chen. Similarity measures for image databases. In *Proc. SPIE Storage and Retrieval for Image and Video Database*, 1995.

[36] B. Julesz. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological Cybernetics*, (41):131–138, 1981.

[37] I. Kamel and C. Faloutsos. Hilbert R-tree. In *Proceedings of the 20th VLDB Conference*, 1994.

[38] P. Kelly and M. Cannon. Query by Image Example: The CANDID approach. In *Proc. of SPIE Storage and Retrieval for Image and Video Databases III*, pages 238–248, 1995.

[39] P. Kelly and T. Cannon. CANDID: Comparison algorithm for navigating digital image database. In *Proc. of the Seventh Intern. Working Conference on Scientific and Statistical Database*, pages 252–258, September 1994.

[40] D. Lee and et. al. Complex queries for a query-by-context image database. Technical Report RJ-9919, IBM , Almaden Research Center, 1994.

[41] W. Li, V. Haese-Coat, and J. Ronsin. Residues of morphological filtering by reconstruction for texture classification. *Pattern Recognition*, 30(7):1081–1093, 1997.

[42] F. Liu and R. Picard. Periodicity, directionality and randomness: Wold features for image modeling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):722–733, July 1996.

[43] H. Lu, B. Ooi, and K. Tan. Efficient image retrieval by color contents. In *Proc. of the 1st Intern. Conf. on Appl. of Databases*, 1994.

[44] W. Ma, Y. Deng, and B. Manjunath. Tools for texture/color based search of images. In *Proc. of SPIE Intern. Conference on Human Vision and Electronic Imaging II*, pages 496–507, February 1997.

[45] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on PAMI*, 11(7):674–693, July 1989.

[46] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[47] B. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):837–841, August 1996.

[48] P. Maragos. Pattern spectrum and multiscale shape representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):701–716, July 1989.

[49] G. Matheron. *Elements pour une Theorie des Milieux Poreux.* Masson, 1967.

[50] C.B. Medeiros and F. Pires. Databases for GIS. *ACM SIGMOD Record*, 23(1):107–115, March 1994.

[51] NASA. Earth Observing System. http://eospso.gsfc.nasa.gov/, 1999.

[52] O. Nielsen. *Wavelets in Scientific Computing.* PhD thesis, Technical University of Denmark, 1996.

[53] V. Ogle and M. Stonebraker. Chabot: Retrieval from a relational database of images. *IEEE Computer*, September 1995.

[54] M. Ortega and et. al. Supporting ranked boolean similarity queries in MARS. *IEEE Transactions on Knowledge and Data Engineering*, 10(6):905–925, November 1998.

[55] D. Panjwani and G. Healey. Markov random field models for unsupervised segmentation of textured color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):939–954, October 1995.

[56] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. In *SPIE Storage and Retrieval Image and Video Databases II*, number 2185, 1994.

[57] E. Petrakis and C. Faloutsos. Similarity searching in large image databases. Technical Report CS-TR-3388, Department of Computer Science, University of Maryland, 1994.

[58] F. Rabitti and P. Savino. An information retrieval approach for image databases. In *Proceedings of the 18th VLDB Conference*, pages 574–584, 1992.

[59] H. Ramapriyan. Satallite imagery in earth science applications. In *Image Databases: search and retrieval of digital imagery*, chapter 3. John Wiley and Sons, 2002.

[60] T. Randen and J. Hakon. Filtering for texture classification: A comparative study. *IEEE Trans. on Pattern Analysis and machine Intelligence*, 21(4):291–310, April 1999.

[61] A. Rao and G. Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. In *Proc. IEEE Conf. Visualisation*, October 1993.

[62] J. Richards. *Remote Sensing Digital Image Analysis.* Springer-Verlag, 1986.

[63] J. Rocha. *The influence of Ground Survey Size on Accuracy of Area estimates from Satellite Images.* PhD thesis, Cranfield Institute of Technology, 1992.

[64] Y. Rui and T. Huang. Image retrieval: Past, present and future. In *Proc. Intern. Symp. on Multimedia Inform. Proc.essing*, December 1997.

[65] Y. Rui et. al. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Trans. on Circuits and Sys. for Video Tech.*, 8(5):644–655, 1998.

[66] H. Samet. *The Design and Analysis of Spatial Data Structures.* Addison Wesley, 1989.

[67] S. Santini and R. Jain. Similarity queries in image databases. In *Proceedings of IEEE Intern. Conference on Computer Vision and Pattern Recognition*, 1996.

[68] S. Santini and R. Jain. Similarity is a geometer. *Multimedia Tools and Applications*, 5(3):306–3775, November 1997.

[69] S. Santini and R. Jain. Beyond query by example. In *Proc. of Sixth ACM Intern. Multimedia Conference*, September 1998.

[70] S. Santini and R. Jain. The "El Niño"Image Database System. In *Proceedings of IEEE Conf. on Multimedia Computing and Systems, Florence, Italy*, 1999.

[71] S. Santini and R. Jain. Similarity matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence (in press)*, 1999.

[72] R. Santos, A. Traina, and C. Traina. Uma linguagem de definição e recuperação de imagens baseada em conteúdo em um banco de dados orientada a objetos. In *Proceedings XI Simposio Brasileiro de Bancos de Dados*, October 1996.

[73] M. Schroder, H. Rehrauer, K. Seidel, and M. Datcu. Spatial information retrieval from remote sensing images: Part B: Gibbs Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 1998.

[74] M. Schroder, K. Seidel, and M. Datcu. Gibbs random field models for image content characterization. In *Proc. IEEE Intern. Geoscience amd Remote Sensing Symposium (IGARSS'97)*, 1997.

[75] T. Sellis, N. Roussopoulos, and C. Faloutsos. The R+ tree: A dynamic index for multidimensional objects. In *Proceedings of the 13th VLDB Conference*, 1987.

[76] J. Serra. *Image Analysis and Mathematical Morphology.* Academic Press, 1982.

[77] G. Sheikholeslami, J. Guo, A. Zhang, and E. Remias. Image decomposition and representation in large image database systems. *Journal of Visual and Image Representation*, 8(2):167–181, June 1997.

[78] G. Sheikholeslami and A. Zhang. A clustering approah for large visual databases. In *SPIE Conference on Visual Data Exploration and Analysis IV,*, volume 3017, pages 322–333, February 1997.

[79] G. Sheikholeslami and A. Zhang. Features visualization and analysis image classification and retrieval. In *2nd International Conference on Visual Information Systems*, December 1997.

[80] G. Sheikholeslami, A. Zhang, and L. Bian. Geographical image classification and retrieval. In *ACM Workshop in GIS*, 1997.

[81] G. Sheikholeslami, A. Zhang, and L. Bian. A multi-resolution content-based retrieval approach for geographic images. *GEOINFORMATICA*, 3(2):109–139, June 1999.

[82] J. Smith and S. Chang. Tools and techniques for color image retrieval. In *Symposium on SPiE Storage and Retrieval for Image and Video Database IV*, volume 2670, February 1996.

[83] J. Smith and S. Chang. VisualSEEK: a fully automated content-based image query system. In *Proceedings of ACM Multimedia'96*, pages 87–98, 1996.

[84] M. Stricker and M. Orengo. Similarity of color images. In *Proc. SPIE - Storage and Retrieval for Image and video Databases III*, pages 381–393, February 1995.

[85] M. Swain and D. Ballard. Color indexing. *Intern. Journal of Computer Vision*, 7(1):11–32, 1991.

[86] J. Tajima. Uniform color scale applications to computer graphics. *Computer Vision, graphics, and Image Processing*, 21:305–325, 1983.

[87] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-8(6):460–473, June 1978.

[88] A. Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.

[89] A. Tversky and I. Gati. Similarity, separability and the triangle inequality. *Psichological Review*, 89:123–154, 1982.

[90] A. Vellaikal, S. Dao, and C. Jay Kuo. Content-based retrieval of remote sensed images using a feature-based approach. In *Science Information Management and Data Compression Workshop*, pages 103–114. NASA Goddard Space Flight Center, October 2995.

[91] A. Vellaikal and C. Jay Kuo. Content-based image retrieval using multiresolution histogram representation. In *SPIE Digital Image Storage and Archiving Systems*, pages 312–323, October 1995.

[92] A. Vellaikal, C. Jay Kuo, and S. Dao. Content-based retrieval of color and multispectral images using join spatial-spectral indexing. In *SPIE Digital Image Storage and Archiving Systems*, pages 312–323, October 1995.

[93] A. Vellaikal, C. Jay Kuo, and S. Dao. Content-based retrieval of remote sensed images using vector quantization. In *SPIE Visual Information Processing IV*,, pages 179–189, April 1995.