



**FACULDADE DE COMPUTAÇÃO
UNIVERSIDADE FEDERAL DE UBERLÂNDIA**



USO DE SALIÊNCIAS DO CONTORNO VIA ESQUELETIZAÇÃO PARA RECUPERAÇÃO DE IMAGENS

Cristiane de Fátima dos Santos

Orientadora:

PROF^A DR^A CÉLIA APARECIDA ZORZO BARCELOS

Janeiro de 2007



**FACULDADE DE COMPUTAÇÃO
UNIVERSIDADE FEDERAL DE UBERLÂNDIA**



USO DE SALIÊNCIAS DO CONTORNO VIA ESQUELETIZAÇÃO PARA RECUPERAÇÃO DE IMAGENS

Cristiane de Fátima dos Santos

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

Orientadora:

PROF^A DR^A CÉLIA APARECIDA ZORZO BARCELOS

Janeiro de 2007

Dados Internacionais de Catalogação na Publicação (CIP)

- S237u Santos, Cristiane de Fátima dos, 1982-
Uso de saliências do contorno via esqueletização para recuperação de
imagens / Cristiane de Fátima dos Santos. - 2007.
124 f. : il.
- Orientadora: Célia Aparecida Zorzo Barcelos.
Dissertação (mestrado) – Universidade Federal de Uberlândia, Progra-
ma de Pós-Graduação em Ciência da Computação.
Inclui bibliografia.
1. Processamento de imagens - Técnicas digitais - Teses. I. Barcelos,
Celia Aparecida Zorzo. II. Universidade Federal de Uberlândia. Programa
de Pós-Graduação em Ciência da Computação. III. Título.

CDU: 621.397.331

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

O presente documento assinado certifica que foi lido e recomendado à Faculdade de Computação - FACOM a aceitação da dissertação intitulada “**Uso de Saliências do Contorno via Esqueletização para Recuperação de Imagens**” de **Cristiane de Fátima dos Santos** como requerimento parcial do grau de **Mestre**.

Data: Janeiro de 2007

Orientador:

Prof^a Dr^a Célia Aparecida Zorzo Barcelos

Avaliador 1:

João do Espírito Santo Batista Neto

Avaliador 2:

Ilmério Reis da Silva

Ser mãe é matemática,
temos que dividir atenção, pois nossos
filhos precisam muito dela, temos que
multiplicar afeto pois ele é imprescindível,
temos que somar alegria e subtrair can-
saço, pois haja fôlego. Mas ser mãe é
mágico também, por isso dedico
essa dissertação a minha
amada filha Laura



Agradecimentos

Agradeço a todos aqueles que contribuíram de alguma forma para a escrita dessa dissertação, seja direta ou indiretamente, em especial, agradeço a ...

Luciano da Fontoura Costa e **Ricardo Picado**, pela sua solicitude em atender os pedidos e gentilmente fornecer material.

Célia Aparecida Zorzo Barcelos acima de tudo pela paciência, compreensão e carinho que dedicou à mim ao longo desta jornada.

Marcos Aurélio Batista meu maior motivador, sempre acreditou em mim incondicionalmente, me defendeu e protegeu quando preciso. Além disso a sua contribuição nessa dissertação é gigantesca: quando tudo parecia sem solução, lá estava ele concebendo maravilhosas idéias.

Jucelino Cardoso meu companheiro que sempre me deu o suporte necessário, acalmando, consolando, motivando, aplaudindo, e tudo isso com todo o amor que se pode ter.

Maria de Fátima Marciano, **Cristina** Cardoso e **Juslene** Cardoso, sogra e cunhadas que me trouxeram sossego ao cuidar de minha filha. Sem elas, certamente não teria chegado até aqui.

Maria de Fátima Martins e **Vanice** que foram meus anjos da guarda, me oferecendo moradia e auxílio.

Sérgio Francisco da Silva, **Ivan** Lopes, **Alexandre** Fieno, e **Liliane** Nascimento pela ajuda sob diversos aspectos e pela amizade.

Agradeço também ao **povo brasileiro** pelo auxílio financeiro prestado através do CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico - na forma de bolsa de mestrado. A todos o meu muito obrigado!

Resumo

A recuperação de imagens é um ramo oriundo da recuperação de informação que vem recebendo grande enfoque atualmente. E dentro deste novo ramo a recuperação baseada em formas é uma das tarefas mais difíceis. Assim, esse trabalho explora os principais aspectos de recuperação baseada em forma: a representação e a descrição de formas, que visam extrair as principais características da forma, e também medidas de similaridade, que são utilizadas na comparação entre dois vetores de característica resultantes da extração de características. Diversos descritores de diferentes classes são abordados dando uma visão geral de como pode ser realizada a recuperação baseada em forma.

Considerando ainda os descritores, importantes aspectos podem ser apontados como sendo desejáveis, sendo a invariância a transformações geométricas um dos principais. Assim, os descritores são analisados quanto à sua invariância a transformações, e algumas normalizações no sentido de tornar os descritores invariantes também são analisadas.

Um descritor novo e bastante promissor, por ser invariante a tais transformações, é o descritor saliências do contorno. Por isso este é amplamente discutido nesse trabalho. Sua obtenção é realizada através do uso de esqueletos sendo que cada saliência é associada à uma extremidade do mesmo, e este por sua vez é obtido através da transformada imagem floresta. Um CBIR baseado em forma utilizando saliências é implementado e seus resultados discutidos.

Palavras-chave:

Recuperação, Imagem, Recuperação de Informação, Transformada Imagem Floresta, Saliências, Esqueletos, Processamento de Imagens Digitais.

Abstract

Image retrieval is an offshoot of information retrieval, which has received growing attention. Within this new offshoot the area of shape-based retrieval has shown itself to be one of the most difficult tasks to perform successfully. It is therefore the purpose of this work to explore the main aspects of shape-based retrieval: the representation and the description of shapes, which aim at extracting the main shape forms, as well as the similarity measurements which are used in the comparison between two characteristic vectors which are the result of characteristic extraction. Various descriptors of different classes are covered giving a general view of how shape based retrieval can be carried out.

When given further consideration these descriptors are able to point out some desirable aspects, with the invariance to geometric transformations being one of the principal ones. Therefore the descriptors are analyzed to their degree of invariance to transformations, also normalizations in the sense of making the descriptors invariant are analyzed.

One new descriptor, contour saliences, in particular shows promising results as it is invariant to such transformations. For this reason the subject is widely discussed in this work, its retrieval is carried out through the use of skeletons being that each saliency is associated to an adjoining extremity, and this is obtained through the use of the image forest transform. A CBIR based on shape that uses salience is implemented and the final results are discussed.

Keywords:

Retrieval, Image, Information Retrieval, Image Forest Transform, Saliences, Skeletons, Digital Image Processing.

Sumário

Agradecimentos	v
Resumo	vii
Abstract	ix
Sumário	xi
Lista de Figuras	xv
Introdução	1
1 Conceitos Básicos	5
1.1 Imagens	5
1.1.1 Características de uma imagem digital	7
1.1.2 Domínio Espacial e Domínio da Frequência	8
1.2 Estatística	9
1.2.1 Histograma	9
1.2.2 Variáveis Aleatórias (VA)	10
1.2.3 Momentos	11

2	CBIRs	12
2.1	Introdução	12
2.2	Recuperação de Informação	13
2.3	Avaliação dos Sistemas de Recuperação de Informação	14
2.3.1	Métricas para avaliação de Sistemas de Recuperação de In- formação	15
2.4	Recuperação de Imagem Baseada em Conteúdo	18
2.4.1	Caracterização de CBIRs	21
2.4.2	Exemplos de CBIRs	22
2.5	Considerações Finais	23
3	Recuperação Baseada em Forma	24
3.1	Introdução	24
3.2	Descrição	25
3.3	Métodos Globais Baseados em Contorno	27
3.3.1	Simple Descritores de Forma	27
3.3.2	Assinatura da forma	28
3.3.3	Método Estocástico	29
3.3.4	Método Scale Space	30
3.3.5	Transformadas Espectrais	33
3.4	Métodos Estruturais Baseados em Contorno	37
3.4.1	Código da Cadeia	37
3.4.2	Aproximação por Polígonos	39
3.4.3	Análise Sintática	41
3.5	Métodos Globais Baseados em Região	43
3.5.1	Descritores de Região Simples	43
3.5.2	Momentos Invariantes Geométricos	44

3.5.3	Momentos Invariantes Algébricos	47
3.5.4	Momentos Ortogonais	47
3.6	Métodos Estruturais Baseados em Região	49
3.6.1	Fecho Convexo	49
3.6.2	Eixo Medial	50
3.7	Descritores Relacionais	54
3.8	Medidas de similaridade	56
3.8.1	Distância Euclidiana	57
3.8.2	Distância de Minkowski	57
3.8.3	Distância Bottleneck	58
3.8.4	Distância de Hausdorff	58
3.8.5	Similaridade da Função Assinatura	59
3.8.6	Distância de Fréchet	60
3.8.7	Distância City-Block	61
3.8.8	Distância Chessboard	61
3.8.9	Casamento Elástico	62
3.9	Considerações Finais	65
4	Saliências do Contorno	67
4.1	Introdução	67
4.2	Transformada Imagem-Floresta	68
4.3	Rotulação dos Pixels do Contorno	72
4.4	Saliências Através da Área de Influência	77
4.5	Esqueletos	79
4.6	Geração do conjunto de características	82
4.6.1	O Algoritmo de Casamento	83
4.7	Considerações finais	87

5	A Implementação	89
5.1	CBIR	89
5.1.1	Módulo de Extração de Características	91
5.1.2	Módulo de Busca	92
5.2	Ferramentas e linguagens	92
5.3	Interface	93
6	Resultados	98
6.1	Introdução	98
6.2	As Consultas	99
6.3	Análise dos Gráficos Precisão x Revocação	99
6.4	Discussão	106
6.4.1	Normalizações no Algoritmo de Casamento	106
6.4.2	Casamentos Parciais	113
6.4.3	Casamentos Exatos de Saliências	115
6.4.4	Valores Repetidos de Dissimilaridade	115
6.4.5	Outras Dificuldades	116
6.5	Considerações Finais	116
	Conclusão e Trabalhos Futuros	118
	Referências Bibliográficas	120

Lista de Figuras

1.1	Histograma	10
2.1	Conjuntos de documentos na Recuperação de Informação	16
2.2	Etapas na Recuperação Baseada em Conteúdo	20
3.1	Classificação das técnicas de representação e descrição de formas . .	26
3.2	Uma maçã e sua assinatura	28
3.3	Obtenção do descritor CSS	33
3.4	Código direcional	37
3.5	Código da cadeia	38
3.6	Aproximações por polígonos	40
3.7	Descrição estrutural de uma forma	41
3.8	Fecho convexo	50
3.9	Eixo medial	51
3.10	Círculo de raio 3 representado dentro de uma grade 9x9	51
3.11	Estrutura de representação SEDR	52
3.12	Estrutura de elementos primitivos em escada	54
3.13	Descrição relacional	56
3.14	Perceptron	63
4.1	Casos ambíguos	73

4.2	Relações na vizinhança do pixel	74
4.3	Rotulação de casos ambíguos	75
4.4	Área de influência interna e externa do ponto	78
4.5	Relação entre o esqueleto e a saliência	82
4.6	Conjunto de características	84
4.7	Saliências através de IFT	85
5.1	Esquema seguido na recuperação	90
5.2	Extração de características	91
5.3	Módulo de busca	92
5.4	Interface do programa mostrando como adquirir imagem consulta	94
5.5	Interface do programa mostrando resultado da consulta	94
5.6	Gráfico de características da imagem consulta	95
5.7	Escolha da imagem para sobrepor características	96
5.8	Características sobrepostas: imagem consulta e imagem no ranking	96
5.9	Saliências de imagem no ranking	97
6.1	Avaliação de consulta: óculos	100
6.2	Avaliação de consulta: taça	101
6.3	Avaliação de consulta: coração	102
6.4	Avaliação de consulta: pássaro	102
6.5	Avaliação de consulta: pássaro	103
6.6	Avaliação de consulta: coelho	104
6.7	Avaliação de consulta: osso	104
6.8	Avaliação de consulta: peixe	105
6.9	Avaliação de consulta: máscara	106
6.10	Avaliação de consulta: tijolo	107
6.11	Ranking para imagem consulta óculos	108

6.12	Ranking para imagem consulta taça	108
6.13	Ranking para imagem consulta coração	109
6.14	Ranking para imagem consulta pássaro07	109
6.15	Ranking para imagem consulta pássaro02	110
6.16	Ranking para imagem consulta coelho	110
6.17	Ranking para imagem consulta osso	111
6.18	Ranking para imagem consulta peixe	111
6.19	Ranking para imagem consulta máscara	112
6.20	Ranking para imagem consulta tijolo	112
6.21	Casamentos Parciais entre imagens bastante distintas	114

Introdução

O uso de imagens é antigo, pré-histórico na verdade: muitos povos antigos costumavam desenhar cenas de seu cotidiano em paredes de cavernas. O que surpreende é que hoje, milhares de anos depois, qualquer ser humano é capaz de compreender fatos da vida de pessoas que viveram há tanto tempo. Uma imagem tem a propriedade de ser universal, no sentido de que qualquer cidadão do planeta e em qualquer época pode compreender o significado dela; exatamente por isso pode-se compreender hoje o que antepassados do *homo sapiens* deixaram inscrito em cavernas há milhares de anos. Uma imagem transpõe qualquer barreira seja ela temporal ou idiomática e não exige conhecimento prévio como a escrita, que pode até mesmo ser perdida com o tempo caso este conhecimento não seja passado para gerações futuras. Por isso, há tempos, sabe-se que uma imagem pode expressar idéias, ações, sentimentos, até melhor do que palavras: “uma imagem vale mais que mil palavras”, é um velho provérbio incansavelmente repetido em várias partes do mundo e por povos de várias culturas.

As técnicas de processamento de imagens começaram a ser utilizadas por jornais no início da década de 20 quando ocorreu a primeira transmissão de imagem via cabo submarino, entre Londres e Nova Iorque. Como a qualidade da imagem era comprometida pela transmissão, havia a necessidade de melhoramentos. Desde então, o processamento de imagem passou a ser amplamente estudado, pois a partir daquele momento surgiu a necessidade de codificar e decodificar a imagem

e muitas vezes melhorá-la, inserir certos efeitos, modificar o tamanho, etc. [22].

Com o tempo e com as melhorias em termos de equipamentos as imagens se tornaram cada vez mais importantes, algumas áreas incorporaram o seu uso, algumas surgiram em função do uso de imagens, e outras passaram a fazer uso massivo delas. Como pode ser percebido, imagens são usadas para diversos fins: na comunicação as imagens dão dinamismo à linguagem, na medicina são de fundamental importância para os exames médicos, na segurança garantem o uso de senhas confiáveis, além de outras aplicações, tais como sensoriamento remoto, construção de protótipos, etc. Por isso, a quantidade de informação visual circulando e sendo usada é gigantesca, o que torna necessário o uso de maneiras eficientes de armazenar, manter e recuperar imagens.

A recuperação de imagens, ou recuperação baseada em conteúdo como também é chamada, é algo bastante recente posto que surgiu da recuperação de informação. Até porque deve se levar em consideração que há algum tempo atrás a preocupação dos cientistas era em otimizar e conseguir espaço para armazenamento. A partir do momento que isso deixou de ser a principal preocupação, o acúmulo de informação se tornou maior, fazendo com que surgisse a “Recuperação” pois era necessário encontrar informação de forma rápida em uma grande coleção de dados. E como não poderia deixar de ser, a recuperação de imagens usou a experiência que se tinha até então em áreas similares: a própria recuperação de informação, em Sistemas Gerenciadores de Banco de Dados e reconhecimento de imagens.

Com base em experiências e com mais pesquisas surgiram o CBIRs (CBIR - *Context Based Image Retrieval*), que são os sistemas que realizam a recuperação de imagens. Diante destes, algumas observações podem ser feitas: os elementos usados na recuperação, a maneira como as imagens são descritas, a maneira como a consulta será submetida, a maneira utilizada para comparar imagens, etc. A recuperação baseada em conteúdo pode ser feita considerando principalmente três

aspectos: cor, forma e textura, sendo que se pode utilizar apenas um destes ou mesmo uma combinação deles. Em algumas situações, a forma é o aspecto mais importante [47], sendo que invariância quanto à transformações é imprescindível: as características de uma forma devem se manter constantes à medida que a imagem é rotacionada, transladada, etc. Bons descritores devem ser empregados na representação das imagens garantindo invariância. Alguns sistemas permitem que o usuário faça um esboço ou especifique através de palavras o que se deseja buscar, porém o mais comum é utilizar imagem consulta. Sobre a maneira de comparar imagens, é fundamental que seja adequada ao tipo de descritor em uso e que seu cálculo não demande muito tempo.

Contudo, pode se dizer que o coração de um sistema de recuperação de imagens é o descritor empregado, afinal, uma imagem deve ter suas características representadas adequada e eficientemente para permitir boa recuperação. Assim o estudo de novos e bons descritores é fundamental para garantir que haja o desenvolvimento da recuperação de imagens. O descritor saliências do contorno, que é recente e possui invariância à transformações geométricas, é bastante promissor e por isso será tratado neste trabalho [47; 46; 45].

É importante ressaltar que invariância sozinha não garante a qualidade do descritor, outros aspectos devem ser analisados e não está claro na literatura se existem problemas na descrição e na recuperação de imagens usando o descritor saliências. Com isso, é objetivo deste trabalho a aplicação desse descritor em uma base de dados heterogênea obtendo as possíveis dificuldades para que possam ser posteriormente resolvidas.

Quanto a estruturação deste trabalho seguem abaixo as principais considerações.

O primeiro capítulo trata de conceitos básicos abordando os principais assuntos que serão necessários na compreensão das técnicas apresentadas nos capítulos

seguintes.

O segundo capítulo trata dos CBIRs, assim, a recuperação de informação também é discutida, já que desta nasceu a recuperação baseada em conteúdo. E dentro da recuperação de informação algumas métricas para medir a eficiência dos sistemas de recuperação de informação são apresentadas.

O terceiro capítulo, visa dar uma idéia geral sobre a recuperação baseada em forma, diversos descritores de forma são apresentados, bem como normalizações para torná-los invariantes a transformações geométricas. Medidas de similaridade são apresentadas ao final do capítulo.

O quarto capítulo apresenta o descritor saliências do contorno, que é obtido via esqueletização. Um método bastante robusto chamado transformada imagem floresta (IFT) é utilizado para extrair os esqueletos das formas, e por isso também é exposto neste capítulo.

O quinto capítulo trata de aspectos da implementação de um sistema de recuperação baseada em conteúdo usando saliências, tais como os módulos utilizados, as ferramentas, etc.

O sexto e último capítulo discute os resultado obtidos no sistema implementado.

Capítulo 1

Conceitos Básicos

Alguns dos conceitos fundamentais na área de processamento de imagens, assim como conceitos mais complexos merecem um capítulo à parte para que não desviem o foco de atenção no momento em que forem necessários, alguns desses conceitos são apresentados neste capítulo.

1.1 Imagens

Vários conceitos devem ser compreendidos e algumas observações realizadas quando se trabalha com imagens, uma dessas observações, e que antecede a qualquer processamento é a captura, que pode ser feita através de máquinas fotográficas, câmeras de televisão, etc. Imagens são de natureza analógica, e para o processamento digital, obviamente necessita-se de imagens digitais. Realiza-se uma conversão analógico-digital para a obtenção desta. Em alguns casos o dispositivo utilizado na captura efetua essa conversão, por exemplo câmeras digitais. Do contrário, algum dispositivo digitalizador, tal qual um *scanner*, pode ser utilizado. Todavia, existe também a opção de sintetizar a imagem em um computador.

A partir do momento que se possui uma imagem digital “em mãos”, é necessário

compreendê-la em sua essência antes de se realizar qualquer processamento: como é armazenada, como é representada ou constituída, como alterá-la, etc, são questionamentos que devem ser feitos para que se chegue de fato nessa compreensão. Uma imagem é constituída de pequenas unidades chamadas *pixels*. Do ponto de vista de representação cada *pixel* é considerado um ponto e uma imagem bidimensional é uma função representada sob forma de matriz cujos elementos são pontos. Computacionalmente, uma imagem possui um determinado formato e um *pixel* não é um ponto mas sim a junção de três pontos de cores distintas que ao final produzem uma cor desejada. Normalmente, utiliza-se o sistema RGB, no qual um *pixel* é a união de um ponto vermelho, um verde e um azul (*Red, Green e Blue*), a variação na intensidade de brilho de cada uma destas cores irá produzir diferentes tonalidades no *pixel* resultante. No caso de imagens monocromáticas um *pixel* é apenas um ponto com um determinado nível de intensidade de brilho, comumente se utiliza uma variação de 0 a 255 nestes níveis.

Computacionalmente, há ainda outra grande preocupação: o armazenamento. Uma imagem pode ser armazenada sob vários formatos, cada qual com uma aplicação considerada mais adequada. Por exemplo, imagens como fotografias, que são constituídas essencialmente de sombras e grandes variações de cor e brilho, necessitam de formatos que permitem boa visualização das inúmeras cores e que não ocupe muito espaço em disco quando armazenadas. Portanto o formato a ser utilizado deverá “encolher”, ou seja, comprimir a imagem sem permitir a perda excessiva de qualidade, já que pequenas perdas não são percebidas pelo olho humano. Já imagens mais simples, com poucas cores, poderão ser armazenadas como um mapeamento de bits, cuja principal vantagem é a facilidade com que o hardware irá interpretar este formato, pois não há a necessidade de descomprimir a imagem antes de visualizá-la.

Uma grande restrição no uso da compressão é que quanto maior a sua taxa em

uma imagem maior será degradação da qualidade, contudo, alguns formatos especiais permitem excelentes taxas de compressão sem grandes perdas de qualidade, o que torna a compressão um grande benefício.

A partir do que foi expostos algumas definições devem ser feitas:

Definição 1.1 (Imagem). *Arte de representar uma pessoa, um objeto, uma paisagem, uma figura ou uma abstração, através de desenhos, traços, figuras geométricas, modelagem ou fotografia, [5].*

Definição 1.2 (Imagem Digital). *Discretização de uma imagem de tonalidade cinza ou colorida, em duas ou três dimensões, onde é atribuído a cada coordenada, um ou três valores numéricos, os quais representam o tom de cinza ou os tons de vermelho, verde e azul da respectiva coordenada ¹ [5].*

Definição 1.3 (Elemento da Imagem). *Menor informação de uma imagem digital. Suas características são suas coordenadas e seu(s) valor(es) numérico(s), [5].*

Encontra-se na literatura, como sinônimo para “elemento da imagem”, o termo “elemento da figura” e as abreviaturas de “*picture element*” - “*pixel*” e “*pel*”, [5].

Definição 1.4 (Imagem Monocromática). *Imagem digital bidimensional com um tom de cinza associado a cada pixel, [5].*

1.1.1 Características de uma imagem digital

Definição 1.5 (Região). *É um subconjunto de uma imagem, [5].*

Definição 1.6 (Vizinhança-de-4). *Dentro de uma região quadrada contendo nove elementos, o elemento central é tomado como base e os 4 elementos alinhados horizontalmente e verticalmente com tal elemento definem sua vizinhança-de-4.*

¹Esta definição expressa as características do modelo RGB de cores e é facilmente adaptável a outros modelos.

Definição 1.7 (Vizinhança-de-8). *Uma região quadrada contendo nove, e não mais que nove, elementos de uma imagem, menos o elemento central, é dita vizinhança-de-8 do elemento no centro desta região, [5].*

Definição 1.8 (Borda). *A borda de uma região é composta pelos elementos pertencentes a esta região, cuja vizinhança-de-8 contenha pelo menos um elemento fora da mesma, [5].*

Definição 1.9 (Contorno). *O contorno de uma região é composto pelos elementos pertencentes ao complementar desta região, cuja vizinhança-de-8 contenha pelo menos um elemento dentro da mesma, [5].*

Definição 1.10 (Fronteira). *A fronteira é composta pelos elementos que dividem partes distintas de uma região, assim, um contorno ou uma borda podem ser chamados de fronteira, pois separam a parte interna da parte externa de um objeto.*

Definição 1.11 (Ruído). *Tonalidade alterada na captação e/ou na transmissão/recepção de uma imagem, [5].*

1.1.2 Domínio Espacial e Domínio da Frequência

As técnicas de processamento de imagens podem ser divididas em duas grandes categorias: métodos no domínio do espaço e métodos no domínio da frequência. Na primeira categoria tem-se técnicas que trabalham no próprio domínio da imagem e fazem a manipulação direta dos *pixels*. Na segunda categoria, tem-se técnicas que são baseadas em transformadas como a de Fourier [22].

Em métodos baseados no domínio do espaço é comum a utilização de janelas de *pixels*, sendo que estas janelas são sobrepostas à imagem a ser processada e possuem diversos tamanhos (1x1, 3x3, 5x5, 7x7, 9x9, etc). Além disso, os elementos em uma janela são pesos que devem ser aplicados a cada *pixel* na imagem, normalmente

esse tipo de janela é denominada máscara. O valor dos pesos é o que determina a natureza do processamento.

No domínio da frequência surgem novas interpretações para elementos da imagem, bordas e mudanças abruptas nos níveis de cinza da imagem contribuem significativamente para o conteúdo de alta frequência. Com isso meios podem ser encontrados para atenuar, aguçar, retirar ou inserir esse tipo de informação. Muitos filtros são definidos no domínio da frequência pois a interpretação é mais intuitiva, por exemplo, filtros de suavização trabalham atenuando os componentes de alta frequência.

1.2 Estatística

Propriedades estatísticas de imagens são de grande importância em processamento de imagens por garantir a possibilidade de análise de aspectos globais e locais nas mesmas. Os conceitos de aspectos estatísticos explorados neste trabalho seguem abaixo.

1.2.1 Histograma

Um histograma é considerado uma função discreta:

$$p(r_k) = \frac{n_k}{n} \quad (1.1)$$

Onde r_k é o k -ésimo nível cinza existente na imagem, n_k é a quantidade de *pixels* cujo nível de intensidade é r_k , e n é a quantidade total de *pixels* na imagem. Portanto, $p(r_k)$ é na verdade uma estimativa da probabilidade de ocorrência do nível cinza r_k .

A figura 1.1 ilustra o histograma de uma imagem. Como cada barra representa

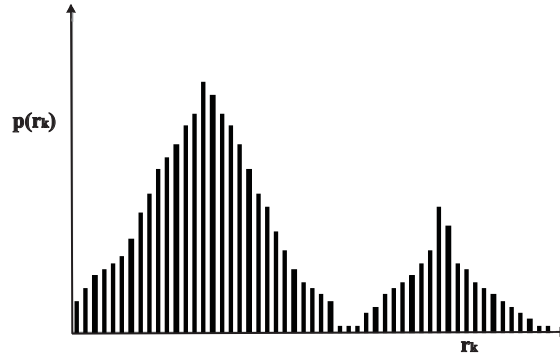


Figura 1.1: Histograma

uma proporção da quantidade de pixels que possuem o nível cinza r_k , observa-se que predominam dois níveis de cinza por causa da formação de dois picos. Outra observação a ser feita no histograma é o eixo vertical que traz informação sobre a frequência de ocorrência do nível r_k .

1.2.2 Variáveis Aleatórias (VA)

Seja E um experimento aleatório, e S um conjunto contendo amostras de possíveis resultados para o experimento, sendo designado de espaço amostral associado a E . Uma função X que associe a cada elemento $s \in S$ um número real é denominada uma variável aleatória. Uma variável aleatória é contínua se existe uma função $f(x)$, denominada função de densidade de probabilidade, que satisfaça as seguintes condições:

- a) $f(x) \geq 0$
- b) $\int_{-\infty}^{+\infty} f(x)dx = 1$
- c) $P(a \leq x \leq b) = \int_a^b f(x)dx$

Na prática, se o intervalo de valores associados a um experimento for enumerável, tem-se uma variável aleatória discreta, do contrário a VA é contínua, e neste

caso o espaço amostral é descrito através de sua probabilidade por uma função: a função de densidade de probabilidade.

1.2.3 Momentos

Momento é uma característica numérica de uma distribuição de frequência, e é denotado por:

$$\mu_t(x) = \sum_{i=1}^n \frac{(x_i - \bar{x})^t f_i}{n} \quad (1.2)$$

Neste caso tem se um momento centrado em t , em que x_i é um valor associado a uma variável aleatória, $\bar{x}$ é uma média do conjunto de valores do espaço amostral, e f_i é a frequência com que x_i aparece na amostra, sendo que a quantidade de elementos na mesma é n . Outra forma de se escrever a equação 1.2 é:

$$\mu_t(x) = \sum_{i=1}^n (x_i - \bar{x})^t p(x_i) \quad (1.3)$$

Onde $p(x)$ é a função de distribuição de probabilidade.

Capítulo 2

CBIRs

2.1 Introdução

O volume de informação de conteúdo visual gerado atualmente é gigantesco, a *World Wide Web* juntamente com as inúmeras áreas do conhecimento humano vem se apoiando cada vez mais no uso de imagens e daí surge a necessidade de armazenar, manter e consultar grandes bases de dados de imagens eficientemente [8; 10; 11; 22; 50]. Dentre as áreas em que essa necessidade é maior pode-se citar: a biologia (genes, espécies, proteínas, etc), aplicações médicas, os sistemas biométricos (reconhecimento de face, digital, íris, etc.), dados espaciais e geográficos, bibliotecas digitais, etc. A Recuperação de Imagem baseada em Conteúdo é área que visa a recuperação de imagens em bases de dados de imagens e é o principal interesse dentro deste trabalho.

2.2 Recuperação de Informação

Recuperar informação significa obter documentos relevantes a uma necessidade de informação especificada inicialmente por meio de uma consulta, sendo que cada documento na base de dados é analisado frente a esta consulta e um grau de importância lhe é atribuído. Esse grau de importância pode ser constituído de valores binários ou não, neste primeiro caso o usuário do SRI (Sistema de Recuperação de Informação) não pode inferir muito sobre os documentos recuperados, sabendo apenas se um determinado documento retornado é relevante ou não. Já com o uso de valores não binários, um *ranking* é construído, possibilitando ao usuário do sistema saber quais são os documentos mais importantes frente à sua consulta [50].

Considerando a recuperação de informação textual, várias maneiras de se efetuarla podem ser estabelecidas, sendo que essas variações podem ir desde a escrita da consulta até a forma tal qual os dados são apresentados. Com isso torna-se conveniente tratar em separado as diversas fases do processo de recuperação [50]:

- 1) Modelagem: internamente, um SRI deve possuir uma estrutura para que haja a recuperação. Tal estrutura visa facilitar o acesso, o armazenamento dos dados e a comparação dos documentos com a consulta. O tipo de estrutura é que define o modelo, que pode ser booleano, vetorial, probabilístico, etc.

- 2) Consulta: com relação à consulta deve-se definir linguagens em que pode ser especificada, bem como operações possíveis de serem realizadas.

- 3) Operações de texto: quando o texto inteiro não é adequado à busca devido a problemas de imprecisão ou redução de tempo de transmissão de dados, operações tais como retirada de *stopwords* (palavras que aparecem muito e não tem nenhum significado semântico, tal como artigos, preposições, etc.), *stemming* (desconsideração de flexões de número e gênero, apenas a raiz da palavra é utilizada), compressão, etc, são necessárias.

4) Indexação e busca: quanto melhor a estrutura de indexação, mais rápida a busca. Além disso, diversos algoritmos de casamento de padrões podem ser utilizados para encontrar um certo padrão de texto nos documentos.

5) Ranking: as informações devem ser apresentadas ao usuário de alguma maneira. Levando em consideração que é altamente conveniente que essas informações estejam em ordem de relevância, um *ranking* deve ser provido.

2.3 Avaliação dos Sistemas de Recuperação de Informação

Diante de um sistema de recuperação de informação, inúmeras características podem ser observadas. Para o usuário algumas dessas características são de fundamental importância e se, de um modo geral, o sistema agrada a este, diz-se que o sistema gera satisfação subjetiva [41]. Contudo, testes de satisfação subjetiva não são fáceis de serem realizados [41]. Torna-se então, conveniente estimar o quanto um sistema poderá agradar o usuário.

Um usuário recorre a um SRI quando possui uma necessidade de informação (que saberá expressar bem ou mal). No momento em que submete essa necessidade o usuário espera que o sistema recupere apenas documentos que sejam do seu interesse (documentos relevantes). É possível verificar se um SRI gera satisfação subjetiva, se o mesmo está recuperando uma grande quantidade de documentos relevantes e também se possui precisão na recuperação, ou seja, não precisa fornecer muitos documentos irrelevantes para recuperar uma quantidade satisfatória de documentos relevantes.

A partir das constatações acima é possível estabelecer duas métricas para a avaliação de SRIs [50]:

- Precisão (*precision*) - quantos dos documentos recuperados são relevantes.
- Revocação (*recall*)- quantos dos documentos relevantes foram recuperados.

O que se busca na realidade é estimar se o SRI está obtendo os documentos pertinentes à consulta. Dessa forma, é preciso saber de antemão quais são os documentos considerados relevantes pelo usuário. Para estabelecer tal controle sobre os dados utiliza-se coleções de documentos em que o conjunto de relevantes é dado através da análise de especialistas humanos. Como exemplo de coleções de referência pode-se citar: a TREC (Text REtrieval Conference), ISI (Institute of Scientific Information), coleção de fibrose cística e a WBR [50].

2.3.1 Métricas para avaliação de Sistemas de Recuperação de Informação

Considerando uma coleção à qual se submete uma consulta, vários conjuntos de documentos estão envolvidos. Dentre eles os mais importantes são: o conjunto de documentos pertinentes à consulta - conjunto de documentos relevantes R e o conjunto de documentos recuperados - conjunto resposta A . Dentro desse último tem-se documentos relevantes e não relevantes, sendo que o conjunto de maior interesse é o de documentos relevantes recuperados Ra . A figura 2.1 ilustra a relação entre estes conjuntos.

Outros conjuntos de documentos também podem ser definidos quando se tem o ponto de vista do usuário: documentos relevantes conhecidos pelo usuário U , documentos relevantes conhecidos pelo usuário que foram recuperados Rk , documentos relevantes previamente desconhecidos pelo usuário que foram recuperados Ru .

Dentre as diversas métricas para avaliação, precisão e revocação são as principais. Ambas estão intimamente ligadas, pois dependem da quantidade de do-

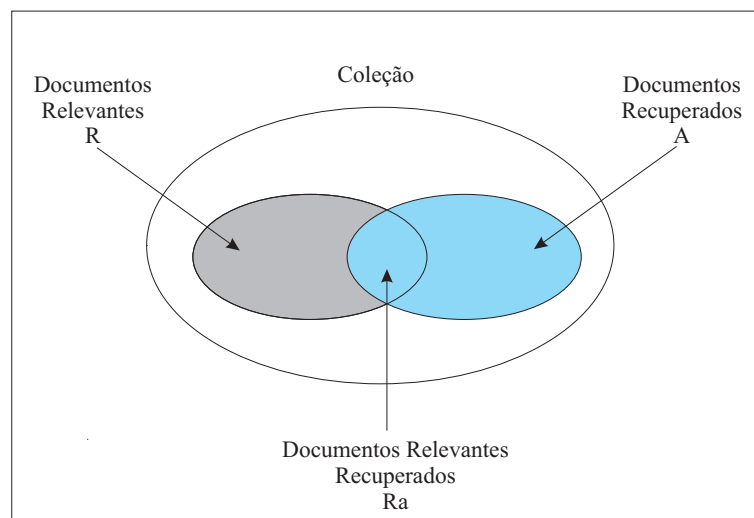


Figura 2.1: Conjuntos de documentos na Recuperação de Informação

documentos relevantes recuperados, $|Ra|$. Um exemplo prático desta ligação é o seguinte: em geral, aumentando a revocação a precisão diminui. Isso ocorre porque à medida que documentos são inseridos na resposta a uma dada consulta aumenta-se a quantidade de documentos relevantes recuperados, contudo, juntamente com os documentos relevantes, documentos não relevantes são inseridos em maior proporção que os relevantes, produzindo o efeito de reduzir a precisão uma vez que a resposta perde em qualidade.

Assim, é importante buscar um equilíbrio entre as duas medidas, uma maneira de facilitar a análise entre as mesmas é traçar o gráfico da Precisão x Revocação. Como a revocação é aumentada à medida que acrescentamos documentos ao resultado da busca, utilizam-se níveis de revocação para análise da performance. Assim, o primeiro nível dá uma revocação teórica de 0%, o segundo 10%, o terceiro 20%, até chegar o décimo primeiro nível que equivale a revocação 100%.

As principais medidas com a finalidade de avaliar os SRIs, estão listadas a seguir:

- Revocação - fração de documentos relevantes que foram recuperados.

$$\text{Revocação} = \frac{|Ra|}{|R|} \quad (2.1)$$

- Precisão - fração de documentos recuperados que são relevantes

$$\text{Precisão} = \frac{|Ra|}{|A|} \quad (2.2)$$

- Precisão média - como a revocação pode variar de 0% a 100% (nível 1 ao 11), é possível fixar um certo nível e para diferentes consultas verificar a precisão média para o nível de revocação dado. Sendo N o número de consultas a média é dada por:

$$\text{Precisão média} = \frac{\text{precisão}_1 + \text{precisão}_2 + \dots + \text{precisão}_N}{N} \quad (2.3)$$

- Precisão-R - sendo R o número de documentos relevantes a uma consulta, uma precisão de 100% seria alcançada quando apenas R documentos são recuperados sem que qualquer outro não relevante o seja, assim a precisão-R mostra o quão perto de uma precisão 100% se tem. Ela é dada pelo simples cálculo da precisão quando R documentos já foram recuperados.
- Cobertura - é uma medida orientada ao usuário baseada no que se espera da recuperação. No momento em que submete uma consulta o usuário espera encontrar alguns documentos. Se todos esses documentos são recuperados, então tem-se uma boa cobertura. Portanto a cobertura, que por definição trata dos documentos relevantes que foram recuperados e que o usuário esperava ver é dada por:

$$\text{cobertura} = \frac{|Rk|}{|U|} \quad (2.4)$$

- Novidade - também é orientada ao usuário, trata-se dos documentos relevantes que foram recuperados e que eram desconhecidos pelo usuário.

$$\text{Novidade} = \frac{|Ru|}{|Ru| + |Rk|} \quad (2.5)$$

2.4 Recuperação de Imagem Baseada em Conteúdo

Em pouco tempo do surgimento da Recuperação de Informação a mesma foi estendida para conteúdo visual. Afinal, muitas vezes o objeto de busca é um conjunto de imagens que sejam semelhantes a uma imagem *query* (imagem consulta, ou seja a informação inicial) [8; 50]. Devido à peculiaridades das imagens, ocorre variações na forma de realizar a recuperação considerando o modelo direcionado a texto. Ainda assim, o esquema básico se aproxima bastante do padrão usada na recuperação textual.

Uma das possibilidades é que a recuperação pode ser feita a partir de informação textual associada a cada documento de imagem descrevendo sua semântica. Contudo, diante do tamanho das coleções, da dificuldade de descrever os objetos da base de dados e da necessidade de processo totalmente automático é necessário criar maneiras de inferir informação do próprio documento: para isto utilizam-se técnicas de análise do conteúdo das imagens, o que originou um novo ramo da Recuperação de Informação que será abordado a seguir.

A Recuperação de Imagem Baseada em Conteúdo é uma sub-área da Recuperação de informação relativamente recente, e sistemas que realizam a recuperação baseada em conteúdo (CBIR - *Context Based Image Retrieval*) têm merecido amplo destaque entre a comunidade científica. Em geral, delimitam-se os aspectos a serem observados a fim de que a recuperação se dê de forma mais rápida. A cor, a forma e a textura são os principais elementos [22; 47; 46]. Alguns dos sistemas

de recuperação de informação trabalham com esses três elementos, enquanto que outros utilizam somente um [22; 47].

Definido o subconjunto de elementos a ser utilizado é necessário escolher ou criar um esquema adequado de representação e descrição. Afinal, os dados sendo considerados durante a recuperação devem ser compactos, além de representarem com certa fidelidade o objeto de imagem. O benefício que isso traz à busca é imenso pois cada objeto passa a ser representado em termos de suas características em um espaço de características definido. Não se tem uma mera descrição informal com base no ponto de vista de um observador como é o caso de descrição textual, mas sim uma maneira bem definida e sucinta de descrição que facilitará a comparação durante as consultas. O vetor de características sintetiza as propriedades da imagem de acordo com os elementos escolhidos, enquanto que a distância métrica mede a dissimilaridade entre imagens.

Além da representação e da descrição serem adequadas deve-se acrescentar o uso de boas características, sendo que uma boa característica é aquela que possui compactabilidade e separabilidade. Isso quer dizer que a característica se comporta relativamente constante dentro de uma classe de objetos enquanto que separa nitidamente os objetos de classes diferentes [47].

As descobertas decorrentes das pesquisas em reconhecimento de imagens, que é uma sub-área da análise, foram fundamentais para os primeiros passos na recuperação baseada em conteúdo. O reconhecimento pode ser definido como a atribuição de uma imagem à uma classe que possua imagens semelhantes. Descobriu-se que seguindo um conjunto de passos bem definidos que tentam simular o que é feito pelo sistema visual humano, é possível obter o reconhecimento de imagens por máquinas. Seqüencialmente, esses passos são os seguintes: aquisição, pré-processamento, segmentação, obtenção de características (representação e descrição), e finalmente o reconhecimento. A fase de aquisição é dependente das peculiaridades da apli-



Figura 2.2: Etapas na Recuperação Baseada em Conteúdo

cação. A fase de pré-processamento engloba a aplicação de técnicas de realce de imagens, remoção de ruído, etc.. A segmentação visa dividir a imagem em objetos constituintes. A fase de representação e descrição consiste basicamente em adotar um esquema de representação de acordo com as características que se esteja buscando, e então descrever as informações, e por fim o reconhecimento que consiste em varrer um conjunto de imagens separadas ou não em classes definindo a que classe pertence a imagem dada inicialmente. Na recuperação de informação com base no conteúdo isso é equivalente a encontrar o conjunto de imagens mais similares que serão retornadas como resultado da consulta [14; 21; 22]. Muitos CBIRs utilizam apenas algumas dessas fases pois normalmente a base de dados já está formada e com qualidade adequada à aplicação. Assim, a aquisição, o pré-processamento e a segmentação podem ser descartados. Um esquema completo das etapas a serem percorridas na recuperação é ilustrado na figura 2.2.

As técnicas de reconhecimento e análise de imagens foram primeiro aplicadas a Sistemas Gerenciadores de Banco de Dados (SGBDs), e estes foram o ponto de partida para o uso dessas técnicas em recuperação de informação [50]

Dois requisitos importantes dentro de um CBIR são [45; 50]:

1. Efetividade - boa recuperação dos itens relevantes a consulta;
2. Eficiência - recuperação tão rápida quando possível dos itens relevantes.

A eficiência depende principalmente da estrutura de indexação utilizada. Enquanto que bons descritores colaboram para a efetividade do sistema e consequentemente para sua qualidade.

A qualidade de um sistema de recuperação de informação pode ser medida em termos de precisão e revocação. Assim, uma representação ruim, em que a simples mudança de tamanho da forma ou ainda a rotação e a translação da mesma, impeçam aquela forma da base de dados de ser reconhecida como relevante, afeta a medida revocação – documentos deveriam ter sido recuperados e não foram, ao passo que se acontecer do esquema de representação não fornecer uma clara distinção entre as formas, várias delas podem ser consideradas relevantes, quando na realidade não são, com isso a precisão do sistema é afetada – documentos não relevantes estão entre os recuperados. Obviamente deve haver um equilíbrio nestas medidas. Nem sempre uma precisão de 100 % é adequada, pode ser que apenas um documento relevante dentre inúmeros tenha sido recuperado, ou seja a revocação é muito baixa, e por outro lado a revocação 100% é muito pouco proveitosa se junto a todos os documentos relevantes recuperados existe uma grande quantidade de documentos não relevantes. Nesse caso a precisão é que deixa a desejar.

2.4.1 Caracterização de CBIRs

Segundo Moura et al. [34] CBIRs podem ser classificados de acordo com o sistema de busca que utilizam:

1. esboço– o usuário cria um desenho da imagem que deseja recuperar. Esse método é bem difícil para a grande maioria dos usuários, já que nem todos possuem dotes artísticos;
2. diretório– as imagens são classificadas manualmente em diretórios e recuperadas através dos mesmos. A classificação manual é o grande problema desse sistema;
3. imagem exemplo– uma *imagem query* é submetida e o sistema recupera as mais similares. Este é o sistema mais utilizado pelos CBIRs;

4. contexto – a imagem é recuperada através do contexto de seus objetos constituintes.

Segundo Erpen [13], um CBIR é constituído por:

1. Uma base de dados – todo item no conjunto resposta pertence a esta base;
2. Um descritor – constituído por vetor de característica e distância métrica;
3. Uma estrutura de dados para indexação – visando dar mais rapidez à busca.

2.4.2 Exemplos de CBIRs

Um dos primeiros CBIRs foi desenvolvido por Hirata e Kato [24]. Trata-se de um CBIR baseado em exemplos visuais no qual arestas são extraídas da consulta e o casamento com as arestas das imagens da base de dados é obtido através de deslocamentos e deformações.

Algum tempo depois do sistema de Hirata e Kato, foi desenvolvido o QBIC (*Query By Image Content*) [35], da IBM, um dos mais avançados e conhecidos sistemas de recuperação baseado em conteúdo de sua época. O QBIC utiliza os principais aspectos da imagem: cor, forma e textura para efetuar a recuperação. Histograma de cores é utilizado na recuperação baseada em cor e descritores simples e momentos são empregados na recuperação baseada em forma e textura. Além disso, uma estrutura de indexação é implementada dando maior rapidez à busca.

Outro CBIR que merece destaque é o sistema desenvolvido por Jacobs et al. [27]. Nele, informação espacial é extraída da imagem e sua representação é feita através de *wavelets*. A informação espacial neste contexto é a disposição física das regiões.

VisualSeek é o sistema desenvolvido por Smith e Chang [42] que permite busca por cor, região e *layout* espacial. Assim como o QBIC, implementa uma estrutura

de indexação.

2.5 Considerações Finais

A recuperação de informação constitui uma das áreas que tem recebido grande atenção da comunidade científica por ser algo recente e de grande potencial de aplicação. Inicialmente foi desenvolvida para texto, e ao passar do tempo agregou técnicas de reconhecimento de imagens com base em experiências de SGBD's e surgiu então a Recuperação de Imagens baseada em Conteúdo. O surgimento dessa nova área foi marcado pela criação dos primeiros CBIR's, sendo que um dos mais conhecidos é o QBIC.

Um dos aspectos mais importantes do CBIR é a informação de imagem que se utiliza para efetuar a recuperação. A qual pode ser feita com base em textura, forma ou cor. Alguns CBIRs utilizam apenas um desses elementos e outros utilizam mais de um ou até os três elementos. Contudo, o que se percebe é que a recuperação baseada apenas na forma está muito longe de atingir um nível ótimo nesses sistemas, tanto que é considerada por alguns pesquisadores [48] uma das tarefas mais difíceis em recuperação.

Os CBIR's, desde que foram criados, vêm sendo aprimorados, mas ainda há o que ser melhorado, seja em termos de eficiência seja em termos de eficácia. É fundamental que haja desenvolvimento e melhorias nas técnicas de representação e descrição de imagens assim como nas métricas utilizadas para comparar imagens que tenham sido descritas. A necessidade de melhorias pode ser comprovada pelo fato de que mesmo diante das inúmeras técnicas de extração de características existentes, novas técnicas tem surgido, como por exemplo a recuperação baseada em saliências, que será abordada nesse trabalho. Técnicas conhecidas e tradicionais de extração de características serão abordadas no capítulo seguinte.

Capítulo 3

Recuperação Baseada em Forma

3.1 Introdução

Recuperação de imagens pode ser baseada em intensidade de cor ou em características geométricas. No primeiro caso tem-se busca por cor e/ou textura, já no segundo buscam-se formas [48; 49]. Em muitos problemas reais, tais como a recuperação em banco de dados biológicos, o reconhecimento de caracteres, o reconhecimento de digitais, etc. [10; 11; 48; 49] as formas são o objeto de busca. Alguns autores até consideram a forma o principal elemento discriminante entre imagens [47; 46; 48; 49; 51].

Apesar da importância, a recuperação baseada em forma não é simples, pelo contrário, segundo Veltkamp et. al [48; 49] esta é uma das tarefas mais difíceis em busca baseada em conteúdo, por exemplo o QBIC não consegue realizá-la eficientemente. Esse problema se deve a própria natureza da imagem digital, afinal, formas são objetos 3-D e durante a sua projeção no plano 2-D informações podem ser perdidas. Somam-se ainda as oclusões, os ruídos, os defeitos, as distorções, etc. [51]. Outro grande problema na recuperação baseada em forma é a dificuldade em

representar e descrever as mesmas.

A representação é feita com base em um esquema, dentre eles pode-se destacar: código da cadeia, aproximações por polígonos, assinaturas e esqueletos [22; 51]. A partir do esquema de representação adotado, um ou mais descritores podem ser utilizados para descrever a forma e transformá-la em um padrão. A tarefa de recuperar se transforma em casamento de padrões.

Na próxima seção a descrição é discutida. Nas seções seguintes os métodos de representação e descrição são abordados de acordo com uma divisão em categorias: na seção 3.3 tem-se métodos globais baseados em contorno; em 3.4 tem-se métodos estruturais baseados em contorno; em 3.5 tem-se métodos globais baseados em região e em 3.6 tem-se métodos estruturais baseados em região. E finalizando o capítulo, descritores relacionais são abordados na seção 3.7, e medidas de similaridade na seção 3.8.

3.2 Descrição

Para que a forma seja transformada em um vetor de características é necessário representá-la e descrevê-la. Portanto, esse processo visa simplificar a forma criando um padrão que a represente. Dois enfoques distintos podem ser adotados nesse processo: o contorno ou a região. Algumas vezes o termo forma é empregado como sinônimo de contorno caso o objeto de interesse seja a parte externa da forma, ou seja o contorno. Dentro dessas duas abordagens uma nova divisão pode ser feita com base no modo como a forma é representada: se primitivas tais como segmentos e figuras geométricas, são utilizadas, considera-se uma abordagem estrutural. Se a forma é tratada como um todo a abordagem é considerada global. Dentro de cada uma dessas abordagens vários descritores podem ser enumerados. Zhang e Lu [51] fazem uma divisão bastante coerente quanto a esses aspectos na representação e

descrição de formas. A figura 3.1, ilustra essa divisão.

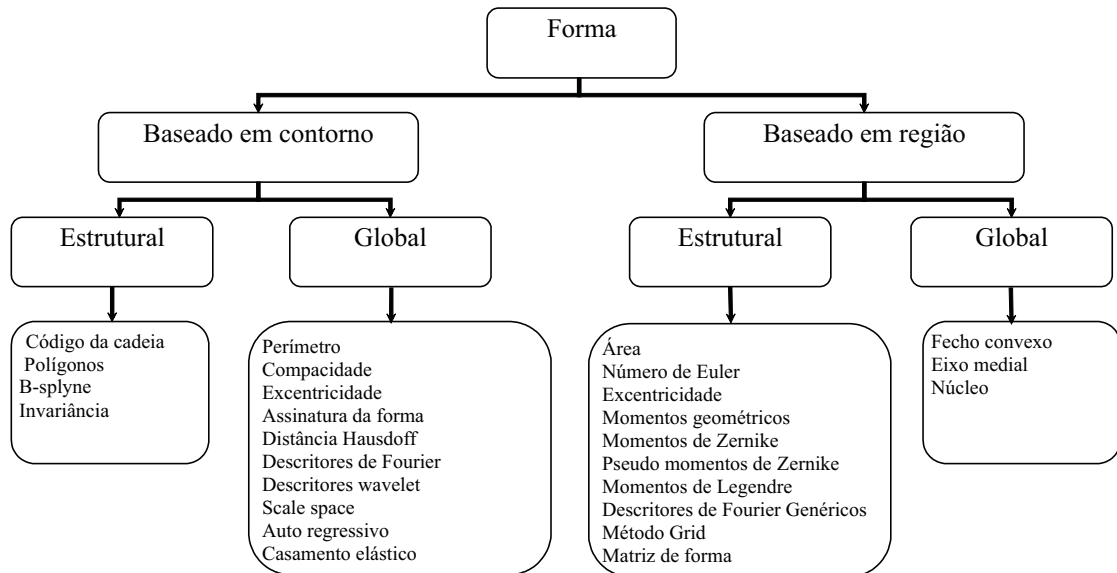


Figura 3.1: Classificação das técnicas de representação e descrição de formas. Retirado de [51]

Inúmeras características podem ser enumeradas quando se deseja projetar um bom descritor. Dentre elas destaca-se que o mesmo deve representar bem a forma, ser conciso, fácil de calcular e comparar e ainda possuir invariância quanto a mudanças geométricas. Na prática é difícil encontrar descritores que tenham pelo menos uma boa parte dessas características. Alguns autores também consideram de grande importância a associação do descritor a um significado físico [51]. Quanto às mudanças geométricas, pode-se destacar:

Rotação - o objeto é deslocado ao longo de seu eixo

Mudança de escala - o objeto é aumentado ou diminuído

Translação - o objeto é transportado de um lugar na imagem para outro

Mudança do ponto inicial - alguns descritores dependem do ponto onde se iniciou a análise do objeto, e pode ocorrer mudança nesse ponto.

3.3 Métodos Globais Baseados em Contorno

O contorno pode ser explorado de duas maneiras distintas: abordagem global e abordagem estrutural, sendo que a primeira é dita contínua e a segunda discreta. A diferença fundamental entre elas é que em abordagens contínuas não se divide a forma em sub-partes, enquanto que na abordagem discreta a borda da forma é dividida em primitivas, e cada método determina um tipo diferente de primitiva. A representação final na abordagem global é uma string e na estrutural, uma string ou um grafo. Em métodos globais, devido à natureza dos dados gerados, maneiras diretas podem ser usadas para efetuar o casamento entre as formas. Utilizam-se por exemplo, métricas como distância Euclidiana, *City Block*, etc. A seção 3.8 aborda algumas métricas existentes. Enquanto isso, em métodos estruturais a métrica pode depender fortemente do tipo de primitiva utilizada.

3.3.1 Simples Descritores de Forma

Alguns descritores são bastante simples e visam somente quantificar a forma sob algum aspecto [22; 51]. Dentre eles pode-se citar:

Perímetro consiste da contagem dos *pixels* ao longo do contorno;

Compacidade é definido por $(\text{perímetro})^2/\text{área}$.

Eixo maior é o diâmetro máximo da forma;

Orientação do eixo maior é a orientação do diâmetro máximo da forma;

Excentricidade é definida por $(\text{eixo maior})/(\text{eixo menor})$;

Curvatura é definida pela taxa de mudança da inclinação do contorno.

Tais descritores só discriminam formas muito diferentes entre si e podem ser utilizados para evitar falsos acertos ou em associação a descritores mais robustos [51]. Um valor isolado de área ou compacidade não diz muito a respeito da forma já que formas muito distintas podem ter o mesmo valor para estas medidas.

3.3.2 Assinatura da forma

A assinatura de um contorno é uma representação unidimensional da mesma. Como uma representação unidimensional é mais simples que a própria forma, justifica-se o uso dessa abordagem.

Diversas maneiras podem ser utilizadas para a obtenção da assinatura, como por exemplo: perfil centroidal, coordenadas complexas, ângulo tangente, ângulo cumulativo, curvatura e área. Contudo, uma maneira bastante simples é através da distância centroidal: dado um centróide dentro do contorno, realizar uma varredura de 360° na mesma e mapear o raio de cada ângulo distinto. O raio nada mais é que a distância de determinado ponto do contorno até o centróide. A partir desta varredura gera-se um gráfico, no qual em um dos eixos são representados os ângulos e no outro eixo o raio, sendo que este ângulo é formado entre o ponto que está sendo mapeado e um ponto de origem localizado no contorno. Na figura 3.2 tem-se uma forma e sua respectiva assinatura.

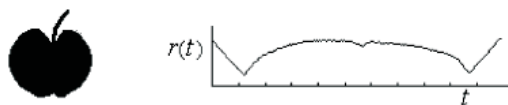


Figura 3.2: Uma maçã e sua assinatura. (Retirado de [51])

A assinatura da forma é invariante a translação, mas encontra problemas na presença de transformações geométricas, tais como rotação e mudança de escala. Normalização quanto à rotação pode ser obtida através da seleção do mesmo ponto

inicial para as assinaturas. Esse ponto inicial pode ser definido como sendo o ponto mais distante do centróide, se as formas forem similares começarão pelo mesmo ponto. Uma normalização do intervalo de amplitude da assinatura torna a assinatura robusta a mudança de escala. Uma maneira de se obter essa normalização é dividir cada amostra pela variância da assinatura [22]. O uso da variância leva a um fator de mudança de escala variável, já que esse fator muda conforme o contorno sendo descrito.

Uma outra alternativa é construir um histograma a partir da assinatura de uma forma, tomando por base os ângulos obtidos. Tal abordagem possibilita invariância à rotação e facilita o casamento entre as formas, contudo perde em qualidade.

3.3.3 Método Estocástico

Um padrão estocástico é aquele que surge através de eventos aleatórios. Uma função unidimensional representando a forma, assinatura por exemplo, pode ser modelada de maneira estocástica e um modelo auto-regressivo (AR) expressa o valor da função como uma combinação linear de um certo número de valores precedentes. Dessa forma, cada valor de função na sequência gerada possui alguma relação com valores prévios de função e por isso podem ser determinados através de M observações de valores prévios de funções. Assim, o modelo AR é um simples preditor do raio corrente por uma combinação linear de M raios prévios mais um termo constante e um termo de erro:

$$f_t = \alpha + \sum_{j=1}^m \theta_j f_{t-j} + \sqrt{\beta \omega_t} \quad (3.1)$$

onde θ_j são os coeficientes do modelo AR, m mostra quantos valores de função precedentes o modelo usa, $\sqrt{\beta \omega_t}$ é o termo de erro corrente e reflete a precisão da predição e α é um valor proporcional à média dos valores de função.

Os parâmetros $\alpha, \theta_1, \dots, \theta_m, \theta$ são estimados usando o critério quadrado mínimo (*Least Square- LS*) [12; 28]. Cada um desses parâmetros pode ou não ser invariante às transformações: θ_j é invariante a rotação, translação e mudança de escala, α e β não são invariantes a escala mas o quociente $\alpha/\sqrt{\beta}$ é considerado invariante. Com isso o vetor de características $[\theta_1, \dots, \theta_m, \alpha/\sqrt{\beta}]^T$ é usado como descritor de formas.

Embora o descritor consiga garantir invariância quanto a transformações, ele possui desvantagens decorrentes do método AR que tem dificuldade em representar bordas complexas através de um pequeno número de parâmetros. Além disso, a escolha de m não é algo simples, quase sempre é feita empiricamente [51].

3.3.4 Método Scale Space

Scale Space visa descrever formas através de propriedades de seu contorno, tais como curvatura e pontos de inflexão, trata-se de um método robusto e eficiente [1]. A representação *scale space* de uma forma é criada traçando os pontos de inflexão das bordas da forma filtrada por um filtro gaussiano de largura variável. É utilizado um filtro gaussiano passa-baixa. Como a largura do filtro cresce, inflexões insignificantes da borda são eliminadas tornando-a mais suave.

Inicialmente, deve ser obtida a curvatura do contorno, que é dada por:

$$k(s) = \lim_{h \rightarrow 0} \frac{\phi}{h} \quad (3.2)$$

onde ϕ é o ângulo entre $t(s)$ e $t(s+h)$, t representa o vetor tangente e s é o comprimento do arco. Cruzamentos por zero são pontos onde o sinal da curvatura muda.

A curvatura deve ser obtida à medida que a forma é suavizada pelo filtro gaussiano. Assim, faz-se necessário realizar uma convolução entre a equação 3.2 e um kernel Gaussiano. Seja:

$$\Gamma(u) = (x(u), y(u)) \quad (3.3)$$

onde u é um parâmetro arbitrário. Em seguida a convolução dos componentes de Γ com o kernel gaussiano de largura σ é efetuada:

$$\begin{aligned} X(u, \sigma) &= x(u) \star g(u, \sigma) \\ Y(u, \sigma) &= y(u) \star g(u, \sigma) \end{aligned} \quad (3.4)$$

Finalmente a equação 3.2 pode ser reescrita a partir das derivadas dos componentes em 3.4:

$$k(s) = \frac{X_u(u, \sigma)Y_{uu}(u, \sigma) - X_{uu}(u, \sigma)Y_u(u, \sigma)}{(X_u(u, \sigma)^2 + Y_{uu}(u, \sigma)^2)^{3/2}} \quad (3.5)$$

onde X_u , X_{uu} , Y_u e Y_{uu} são as derivadas de primeira e segunda ordem de cada componente. Tais derivadas são obtidas a partir das propriedades de convolução:

$$\begin{aligned} X(u, \sigma) &= x(u) \star g_u(u, \sigma) \\ Y(u, \sigma) &= y(u) \star g_{uu}(u, \sigma) \end{aligned} \quad (3.6)$$

Cada objeto é representado pelas coordenadas x e y de seus pontos de borda. Normalmente um número fixo de pontos igualmente distantes na borda são escolhidos, Abbasi et al. [2; 1] sugerem utilizar 200 pontos para formas contendo entre 400 e 1200 pontos de borda. Só então é feita a convolução com o filtro gaussiano. A curvatura é obtida a partir da equação 3.6, sendo que a largura de banda σ do filtro gaussiano é incrementada gerando assim vários valores diferentes para Γ . A

medida que σ cresce, apenas os pontos mais significativos permanecem, quando todos os pontos sumirem e uma borda convexa for obtida não mais se realizará incrementos em σ .

Determinando a localização dos cruzamentos por zero de curvatura os pontos obtidos podem ser plotados em um plano (u, σ) onde u é o índice do ponto com relação aos pontos de contorno, $(u = 1, 2, \dots, 200)$, e σ a largura do kernel gaussiano. Veja figura 3.3(c).

Como pode ser observado na figura 3.3, espera-se que os pontos de inflexão da borda que permanecem na representação sejam as características mais significativas do objeto. Em 3.3(a) a aplicação do filtro gaussiano com várias larguras pode ser visualizada, sendo que os pontos marcados são pontos de inflexão da borda. Em 3.3(b) verifica-se a árvore de intervalo, que é o produto final do processo de suavização: localização dos cruzamentos por zero. Ela guarda informações sobre os pontos de inflexão e muitas vezes é chamada de mapa do contorno *curvature scale space*. Em 3.3(c) são plotados os cumes da árvore de intervalo.

Essa aproximação possui como desvantagem a dificuldade na interpretação dos resultados finais. Algumas tentativas podem ser encontradas na literatura. Dentre elas, uma das primeiras é a de Asada e Brady [4], cuja árvore de intervalo é baseada na detecção de cumes dos ramos de alta escala a baixa escala. A partir dos cumes eventos primitivos de alto nível são detectados, tais como cantos, junções suaves, terminações, curvas, etc.

Mokhtarian and Mackworth [33] fizeram alterações sobre o modelo de Asada e Brady. No lugar de interpretar cumes de ramos individuais como eventos primitivos de alto nível, eles aplicaram um modelo de casamento que tenta encontrar o melhor casamento entre ramos do contorno para duas árvores de intervalo. Para cada casamento individual, o modelo de ramo do contorno tem que sofrer mudança de escala e ser deslocado para garantir invariância às respectivas transformações.

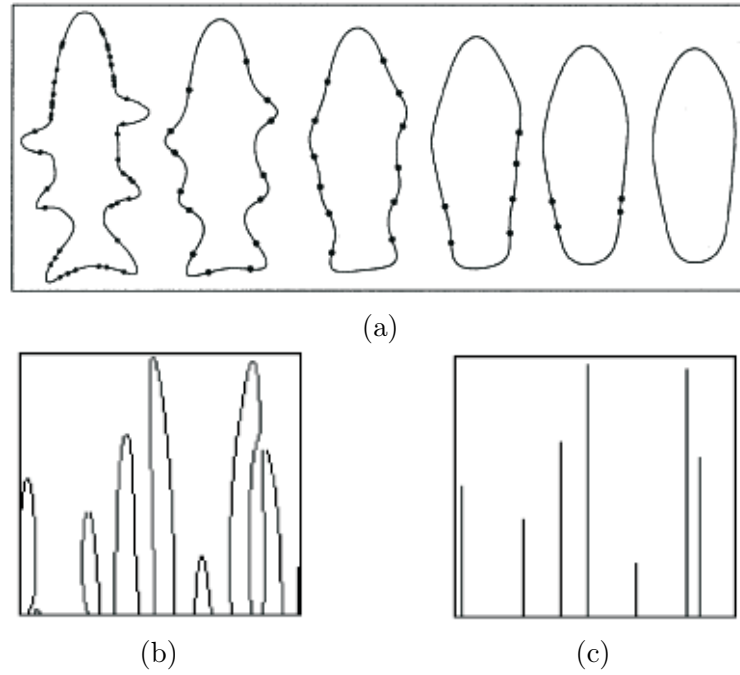


Figura 3.3: Obtenção do descritor CSS. (a) Larguras $\sigma = 1, 4, 7, 10, 12, 14$ do filtro gaussiano. (b) Árvore de intervalo. (c) Cumes da árvore de intervalo. (Retirado de [51])

Apesar de terem melhorado o método, o casamento ficou mais complexo e de custo mais alto. Posteriormente o método foi melhorado para trabalhar em Recuperação de Imagens [2]. Mokhtarian and Mackworth denominaram a assinatura *Scale Space* de *Curvature Scale Space* (CSS) tornando este nome bastante usado na literatura.

3.3.5 Transformadas Espectrais

As técnicas de processamento de imagens são baseadas no domínio do espaço ou no domínio da frequência, até então as técnicas abordadas são baseadas do domínio do espaço.

Descritores que trabalham com o espectro são melhores diante de problemas de sensibilidade ao ruído e variações na forma [51]. Dois dos principais descritores que podem ser citados são os descritores de Fourier (FD) e descritores wavelets

(WD).

O uso de wavelets para descrever formas é bastante recente e vem ganhando espaço na comunidade científica. Uma de suas vantagens é que trabalha no domínio do espaço e no domínio da frequência. Contudo, seu algoritmo de casamento torna seu uso proibitivo em Recuperação de Imagens devido à sua complexidade bastante elevada [51]. Por diversos fatores tais como a facilidade de ser calculado, existência de significado físico de cada descritor e facilidade de normalização o descritor de Fourier é mais usado.

Descritores de Fourier

Descritores de Fourier podem ser obtidos da seguinte forma: definido um contorno s constituído de N pontos e um ponto de origem no mesmo, cada ponto pertencente a esse contorno pode ser denotado por um par coordenado (x_i, y_i) , $i = 0, 1, \dots, N-1$. Assim o contorno s é gerado por todo o conjunto de pares coordenados existentes: $s(k) = [x(k), y(k)]$, onde $x(k) = x_k$ e $y(k) = y_k$. E por razões de simplificação, cada par ordenado é representado na forma de um número complexo, obtendo-se assim um contorno unidimensional. Tal número é da forma:

$$s(k) = x(k) + jy(k) \quad (3.7)$$

Assim, o eixo x é o eixo real e o eixo y é o eixo imaginário. A partir desses dados a transformada discreta de Fourier (*Discrete Fourier Transform* - DFT) é definida:

$$a(u) = \frac{1}{N} \sum_{k=0}^{N-1} s(k) \exp[-j2\pi uk/N] \quad (3.8)$$

para $u = 0, 1, 2, \dots, N-1$. Os coeficientes complexos $a(u)$ são chamados de descritores de Fourier. Para se reconstruir o contorno s utiliza-se a transformada inversa

de Fourier:

$$s(k) = \sum_{u=0}^{N-1} a(u) \exp[j2\pi uk/N] \quad (3.9)$$

para $k = 0, 1, 2, \dots, N - 1$. Muitas vezes, torna-se desnecessário utilizar os N coeficientes, já que o contorno pode possuir um grande número de pontos. Utilizam-se então apenas os M pontos (M geralmente é uma potência de 2), neste caso:

$$\hat{s}(k) = \sum_{u=0}^{M-1} a(u) \exp[j2\pi uk/N] \quad (3.10)$$

para $k = 0, 1, 2, \dots, N - 1$. Quanto menor o valor de M mais detalhes são perdidos no contorno.

Descritores de Fourier são sensíveis a translação, rotação e mudança de escala. Contudo, através de simples mudanças pode-se obter invariância quanto a estas transformações [22]. A tabela 3.1 mostra as modificações requeridas para cada tipo de transformação.

TRANSFORMAÇÃO	CONTORNO	DESCRITOR DE FOURIER
Identidade	$s(k)$	$a(u)$
Rotação	$s_r(k) = s(k)e^{j\theta}$	$a_r(u) = a(u)e^{j\theta}$
Translação	$s_t(k) = s(k) + \Delta_{xy}$	$a_t(u) = a(u) + \Delta_{xy}\delta(u)$
Mudança de escala	$s_s(k) = \alpha s(k)$	$a_s(u) = \alpha a(u)$
Pontos de partida	$s_p(k) = s(k - k_0)$	$a_p(u) = a(u)e^{-j2\pi k_0 u/N}$

Tabela 3.1: Mudanças nos descritores de Fourier

Descritores de Fourier são amplamente utilizados pois conseguem sem muito esforço, segundo Gonzalez et al. [22], capturar a essência de uma fronteira. Vários pesquisadores tem utilizado descritores de Fourier e contribuído para sua melhoria. Dentre as principais contribuições pode-se destacar:

- Lin et al. e Mitchell et al. [29; 32] estenderam o método para que pudesse

ser aplicado a formas parciais. Originalmente descritores de Fourier eram aplicados apenas a contornos fechados.

- Arbter introduziu descritores de Fourier invariante ao afinamento para ser aplicado a formas afinadas [3].
- Ruaber propôs o UNL FD, (nome baseado em Universidade Nova de Lisboa, Portugal) que é capaz de descrever contornos disjuntos ou articulados [38].
- Richard e Hemami introduziram uma medida de distância complexa que visa obter a verdadeira medida de distância entre dois conjuntos de FDs [39].
- Rui et al. propôs uma medida de distância para classificar similaridade de caracteres usando coeficientes de Fourier [40].

Descritores de Fourier são amplamente utilizados pois conseguem sem muito esforço, segundo Gonzalez et al. [22], capturar a essência de uma fronteira. Vários pesquisadores tem utilizado descritores de Fourier e trabalhos foram concebidos para a aplicação do método a formas parciais [51], dentre estes tem se o método de Lin et al. e Mitchell et al. [29; 32]. Outras contribuições para a melhoria do método podem ser encontradas, dentre elas as principais são:

- Arbter introduziu descritores de Fourier invariante ao afinamento para ser aplicado a formas afinadas [3].
- Ruaber propôs o UNL FD, (nome baseado em Universidade Nova de Lisboa, Portugal) que é capaz de descrever contornos disjuntos ou articulados [38].
- Richard e Hemami introduziram uma medida de distância complexa que visa obter a verdadeira medida de distância entre dois conjuntos de FDs [39].
- Rui et al. propôs uma medida de distância para classificar similaridade de caracteres usando coeficientes de Fourier [40].

3.4 Métodos Estruturais Baseados em Contorno

Uma outra classe de métodos baseados em contorno é a classe que envolve métodos estruturais. Estes tem como principal característica o fato de quebrar o contorno em primitivas de imagem para então gerar uma descrição da forma. Cada método estrutural determina um tipo diferente de primitiva bem como uma maneira de organizar tais primitivas. O resultado de um método estrutural é uma *string* na forma $S = s_1, s_2, \dots, s_n$. Os principais métodos estruturais são: código da cadeia, aproximação por polígonos e decomposição em curvas.

3.4.1 Código da Cadeia

Este método foi criado por Freeman [20], e consiste em representar uma fronteira por uma seqüência conectada de segmentos de linha reta com tamanho e direção determinados. Para a determinação da direção o conceito da conectividade-de-4 ou conectividade-de-8 é utilizado, tais conceitos são baseados em vizinhança-de-4 e vizinhança-de-8. A figura 3.4(a) mostra o código direcional-de-4, e a figura 3.4(b) mostra o código direcional-de-8.

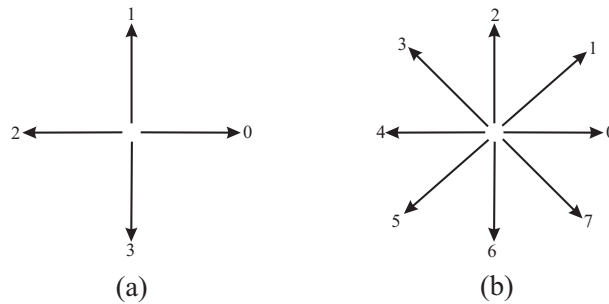


Figura 3.4: Código direcional. (a)Código direcional-de-4. (b) Código direcional-de-8.

Dessa forma, um valor é associado a cada direção de segmento, enquanto que o tamanho do mesmo é dado pela grade de reamostragem, que é usualmente utilizada em processamento de imagens digitais. A reamostragem da fronteira é uti-

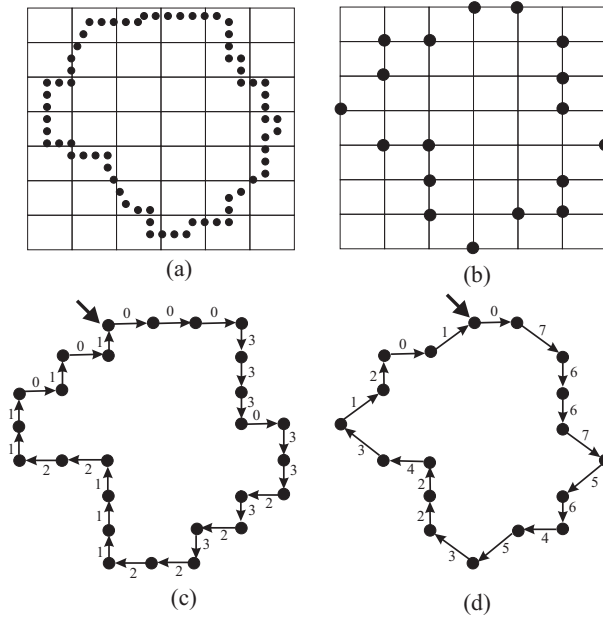


Figura 3.5: Código da cadeia. (a)Fronteira com grade de reamostragem. (b)Resultado da reamostragem. (c)Código da cadeia direcional de 4 (d)Código da cadeia direcional de 8

lizada para que não seja gerada uma cadeia de código muito grande e também para que pequenas perturbações ao longo da fronteira não afetem o resultado. O procedimento da reamostragem consiste em definir a imagem em uma grade com espaçamento maior do que a grade inicial e marcar os pontos da fronteira que estejam nas intersecções das linhas da grade. Os pontos que não estiverem em intersecção são aproximados para as que estiverem mais próximas. Dessa forma, pontos que estiverem muito próximos podem desaparecer ao serem aproximados para a mesma intersecção. A figura 3.5(a) mostra uma fronteira a ser reamostrada colocada sobre uma grade e a figura 3.5(b) mostra o resultado da reamostragem.

Definido o código direcional a ser usado, o código de cadeia é gerado seguindo a fronteira no sentido horário. A figura 3.5(c) mostra o resultado da reamostragem em código direcional de 4 e a figura 3.5(d) mostra o resultado da reamostragem em código direcional de 8.

Nos códigos gerados cada segmento é representado por um valor. Considerando que o ponto inicial é o que está indicado pela seta na parte superior, na cadeia direcional de 4 tem-se 00033303323232211122110101 e na cadeia direcional de 8 tem-se 076675645322431201, como ilustra as figuras 3.5(c)-(d). Para que não haja dependência em relação ao ponto inicial é feita a seguinte normalização: o código é tratado como uma seqüência circular de números sendo que o ponto inicial é dado de forma que o número inteiro formado seja de menor magnitude possível. Normalização de tamanho também é possível através da alteração do tamanho da grade de reamostragem. Normalização feitas em representações do código da cadeia só são exatas se as próprias fronteiras forem invariantes a rotação e à mudança de escala, o que dificilmente ocorre. Além disso a *string* gerada é bastante extensa e o método é muito sensível à ruídos, dificultando assim o seu uso em Recuperação de Imagens.

3.4.2 Aproximação por Polígonos

Em tal abordagem as fronteiras são aproximadas por polígonos. Busca-se captar a essência da fronteira com o menor número possível de segmentos poligonais, de acordo com a precisão que se deseje, sendo que a melhor precisão possível gera exatamente a fronteira. No entanto, não se tem uma representação muito concisa neste caso. Portanto, a precisão não deve ser máxima e nem mínima já que neste caso não representaria adequadamente a forma [22].

O polígono obtido deve possuir perímetro mínimo, ou seja, dada uma fronteira conexa S , alguma técnica deve ser utilizada para que seja gerado um polígono que possua forma semelhante a da fronteira e com a menor soma dos lados possível. Comumente se utiliza sobrepor um conjunto de células à fronteira, de maneira que o polígono possa ser traçado dentro das células (figura 3.6).

Diferentes formas de se trabalhar com aproximação por polígonos podem ser

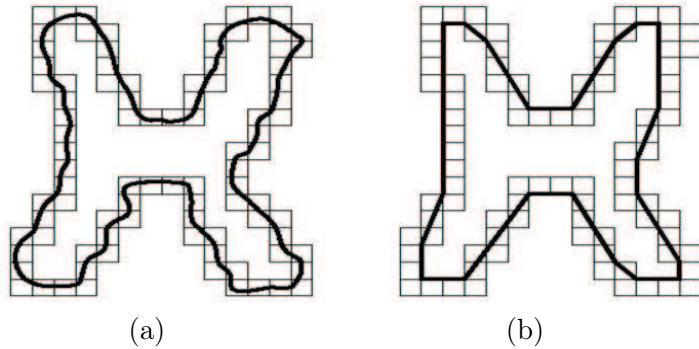


Figura 3.6: Aproximações por polígonos. (a) Fronteira de objeto coberto por células (b) Polígonos de perímetro mínimo que representa a fronteira

encontradas na literatura. Groskey et al. [23] usam os vértices dos polígonos como primitivas, sendo que para cada primitiva gera-se uma *string* contendo ângulo interno, distância ao próximo vértice e as coordenadas x e y . Apenas os cinco vértices mais agudos são selecionados para cada forma e para efetuar a busca um índice na forma de árvore é construído contendo tais características. Assim, a busca é realizada em dois passos: no primeiro, se uma característica modelo similar a consulta é encontrada, uma lista das formas associadas com a característica modelo é recuperado. No segundo passo da busca o casamento entre a imagem consulta e as imagens recuperadas é feito baseado em uma medida de distância entre as *strings* das primitivas.

Uma outra forma de se trabalhar com essa aproximação foi desenvolvida por Mehrotra e Gary [31], eles representam a forma como uma cadeia de vetores. Uma série de pontos da aproximação poligonal são detectados e um par destes pontos é escolhido para formar a base do vetor, e todos os outros pontos detectados são transformados neste sistema de coordenadas. Com isso a forma é representada pelo conjunto de coordenadas $(x_1, y_1), \dots, (x_n, y_n)$, que são também transformadas com os parâmetros S, T_x, T_y, θ correspondentes a base do vetor quanto à escala, translação e rotação. Uma árvore índice é construída e usada na busca, como

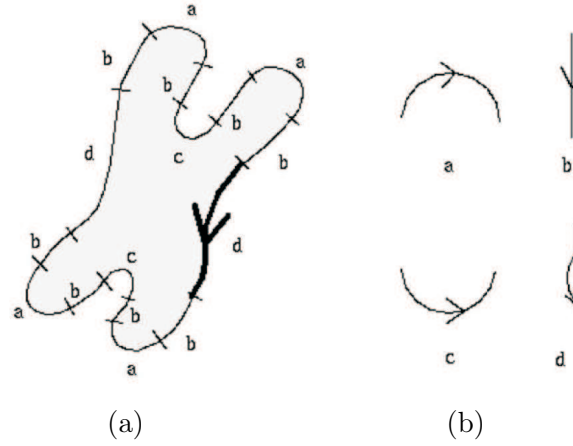


Figura 3.7: Descrição estrutural de uma forma. (a) Cromossomo dividido em primitivas. (b) Primitivas. Retirado de [51]

o casamento de uma ou mais características não garante casamento completo, a similaridade entre as formas é novamente verificada entre cada imagem recuperada e a imagem consulta. A distância Euclidiana é usada para medir a similaridade entre duas características.

O problema com esta aproximação é que ela se comporta bem somente quando as formas são simples ou seja, formas sintéticas, já para formas reais não se tem bons resultados [51].

3.4.3 Análise Sintática

Esta aproximação assume que imagens podem ser construídas assim como a linguagem [22]. Primitivas são predefinidas e assumidas como sendo os símbolos do alfabeto, cuja combinação permitirá a formação de palavras (formas mais complexas) que por sua vez se unirão formando frases (cenários). Por exemplo, dadas as primitivas a, b, c e d, (figura 3.7(b)), a forma da figura 3.7(a) pode ser representado pela *string*:

$$S = dbabcbabdbabcbab$$

Cada primitiva é um símbolo do alfabeto de uma gramática e uma linguagem é um conjunto de sentenças geradas por tal gramática. Assim sendo, pode-se definir gramática como sendo um formalismo gerador constituído de um alfabeto, símbolos iniciais e terminais, bem como regras de sintaxe para a geração das sentenças. Tal gramática pode ser definida a partir dos seguintes parâmetros:

$$G = (N, \alpha, P, S) \quad (3.11)$$

N = conjunto finito de variáveis chamados não terminais

α = conjunto finito de constantes chamadas terminais

P = conjunto de regras de reescrita chamadas produções

S = é o símbolo inicial, e pertence a N

Dada uma gramática e uma cadeia, sequências de derivações podem ser feitas até que a cadeia seja encontrada, sendo que derivações são feitas simplesmente reescrevendo os símbolos não terminais de acordo com as regras de produções. Um autômato é um formalismo reconhecedor que verifica se uma cadeia pertence a uma gramática.

Neste contexto, sentenças são *strings* de símbolos (padrões) e linguagens correspondem a classes de padrões.

A dificuldade do método é que os padrões gerados por uma linguagem precisam ser conhecidos, o que implica em necessidade de treinamento para construção das produções da gramática, para situações de domínio geral isso constitui um grande problema, já que a infinidade de classes de objetos existente geraria uma quantidade muito grande de regras de produção. Além disso, nem sempre se pode garantir que essas gramáticas geram apenas padrões válidos. O alfabeto também constitui outra dificuldade já que ele é dependente da aplicação.

3.5 Métodos Globais Baseados em Região

Quando o foco de atenção é a região e não o contorno, apenas propriedades internas ao contorno são observadas para descrever a forma. Métodos globais tratam a forma como um todo e produzem como resultado um vetor de característica. Assim distância métrica pode ser aplicada para obter o casamento entre formas.

Analogamente aos descritores de fronteiras há diferentes níveis de complexidade entre os descritores de regiões, bem como descritores globais e estruturais. Nas subseções a seguir a caracterização de alguns dos principais descritores de região é abordada.

3.5.1 Descritores de Região Simples

Estes descritores são bastante simples, e buscam extrair características da região e muitas vezes são utilizados em associação a outros descritores de região. Pode-se verificar que a região de uma maneira geral traz muito mais informação do que o contorno e que por tanto possui uma quantidade maior de descritores.

Perímetro é o tamanho da fronteira.

Excentricidade é definida pelo maior raio/ menor raio

Área é definida pelo numero de *pixels* contidos em uma região determinada por uma fronteira.

Compacidade é dada por $(\text{perímetro})^2/\text{área}$,

Média oferece uma medida estatística para descrever uma região através da média dos níveis cinza dos *pixels*. Uma variação do uso da média é contar quantos pixels estão com nível cinza acima ou abaixo dela.

Mediana análogo à média.

Valores de mínimo e máximo descrição através de *pixels* que possuem nível cinza mínimo ou máximo em comparação aos outros *pixels* contidos na região a que pertencem.

Descritores topológicos em imagens, topologia estuda as propriedades de uma figura que não sejam afetadas por deformações. Um tipo de descritor topológico é o número de buracos. Um buraco pode ser compreendido como a presença de características externas à região dentro da mesma. Neste caso a característica a ser definida é o nível cinza dos *pixels*. Ainda que a figura seja esticada ou rotacionada, a quantidade de buracos presentes em uma região da figura continua sendo a mesma. Outra característica que não é afetada por essas mudanças é a quantidade de componentes conexos, sendo que cada componente conexo é uma subregião de tamanho máximo tal que quaisquer dois pontos pertencentes a ela podem ser unidos por uma curva conexa que também pertença à subregião.

O número de Euler é uma característica topológica dada pela diferença entre o número de componentes conexos e o número de buracos.

3.5.2 Momentos Invariantes Geométricos

A estatística sempre contribuiu para o processamento de imagens. Dentre essas contribuições uma das mais importantes é a aplicação de momentos invariantes ao reconhecimento de padrões.

Normalmente, apenas sete momentos invariantes são calculados, sendo que os mesmos baseiam-se nos momentos centrais de segunda e terceira ordem. O fundamento que leva a calcular apenas estes sete momentos é que exatamente essa quantidade representa detalhes distinguíveis pelo ser humano [52]. Um descritor baseado em momento pode usar todos os sete ou apenas alguns deles. Outro fato

importante que se deve notar é que momentos caracterizam principalmente a regularidade do objeto pois refletem o espalhamento do mesmo em relação ao seu ponto médio [52]. O cálculo desses 7 momentos pode ser feitos da seguinte forma:

1. Cálculo dos momentos centrais de segunda e terceira ordem, μ_{00} , μ_{11} , μ_{20} , μ_{02} , μ_{21} , μ_{12} , μ_{30} , μ_{03}

$$\mu_{pq} = \sum_{n=1}^N ((x(n) - \bar{x})^p \cdot (y(n) - \bar{y})^q) \quad (3.12)$$

onde N é o número de pontos (x, y) do contorno fechado, e $\bar{x}$ e $\bar{y}$ são as médias das componentes x e y dos pontos de contorno, respectivamente.

2. Normalização dos momentos centrais μ_{pq} , gerando os momentos centrais normalizados m_{pq} , através das expressões:

$$m_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \gamma = \frac{1}{2}(p + q) + 1 \quad (3.13)$$

3. Cálculo dos momentos invariantes através das expressões:

$$\begin{aligned}
M1 &= m_{20} + m_{02} \\
M2 &= (m_{20} - m_{02})^2 + 4m_{11}^2 \\
M3 &= (m_{30} - 3m_{12})^2 + (m_{03} - 3m_{21})^2 \\
M4 &= (m_{30} + 3m_{12})^2 + (m_{03} + 3m_{21})^2 \\
M5 &= (m_{30} - 3m_{12})(m_{03} + 3m_{21})[(m_{30} + 3m_{12})^2 - 3(m_{03} + 3m_{21})^2] \\
&\quad + (m_{03} - 3m_{21})(m_{03} + 3m_{21})[(m_{03} + 3m_{21})^2 - 3(m_{30} + 3m_{12})^2] \\
M6 &= (m_{20} - m_{02})(m_{30} + m_{12})[(m_{03} + m_{21})^2] \\
&\quad + 4m_{11}(m_{30} - m_{12})(m_{03} + 3m_{21}) \\
M7 &= (3m_{12} - m_{30})(m_{30} + m_{12})[(m_{30} + m_{12})^2 - 3(m_{03} + m_{21})^2] \\
&\quad + (m_{03} - 3m_{21})(m_{03} + 3m_{21})[(m_{03} + m_{21})^2 - 3(m_{30} + m_{12})^2] \quad (3.14)
\end{aligned}$$

Um dos primeiros estudos sobre a utilização desta metodologia foi feito por Hu [26], que se baseou no trabalho de matemáticos como Boole, Cayley e Sylvester, e na teoria das formas algébricas:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad p, q = 0, 1, 2, \dots \quad (3.15)$$

Hu usou uma combinação não-linear de momentos de baixa ordem para derivar um conjunto de momentos invariantes, também chamados de momentos geométricos. Tais momentos possuem propriedades de invariância com relação a transformações geométricas.

O principal problema com momentos geométricos é que nem sempre fornecem parâmetros suficientes para descrever formas com precisão.

3.5.3 Momentos Invariantes Algébricos

Esse descritor empregado no QBIC [35] foi desenvolvido por Taubin e Cooper [43] e consiste de calcular os m momentos centrais que são dados como autovalores de matrizes predefinidas cujos elementos são fatores escalados dos momentos centrais [51]. Momentos invariantes algébricos podem ser construídos em ordem arbitrária, sendo esta uma de suas principais vantagens. Contudo, tais momentos tendem a trabalhar bem nos objetos onde a distribuição dos *pixels* é mais importante do que o perfil da forma.

3.5.4 Momentos Ortogonais

Momentos ortogonais surgiram como uma derivação dos momentos algébricos (equação 3.15), e foram introduzidos em descrição de imagens por Teague [44], sendo que a mudança realizada consistiu em substituir o kernel convencional $x^p y^q$, por um kernel mais geral $P_p(x)P_q(y)$. Podem ser usados os polinômios de Legendre e de Zernike no kernel geral. Tais momentos são denominados ortogonais porque são conjuntos completos de bases ortogonais.

Momentos de Legendre são dados por:

$$L_{nm} = \frac{(2m+1)(2n+1)}{4} \sum_x \sum_y P_m(x)P_n(y)f(x,y) \quad (3.16)$$

onde

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n \quad (3.17)$$

Momentos de Zernike baseiam-se em polinômios que os tornam independentes e facilmente calculados até uma ordem elevada [52]. Algo importante a se observar é que momentos de Zernike implicam a normalização da forma a um disco circular

de diâmetro unitário, já que os momentos de Zernique se encontram definidos dessa forma. Assim, a forma é completamente envolvida pelo disco e passa a ter área πR^2 , sendo que R é o raio do disco unitário. Feita esta normalização os momentos de Zernique são extraídos através da equação 3.18.

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y B_{nm}(\rho, \theta) f(\rho, \theta) \quad (3.18)$$

onde n é a ordem, m a repetição, (ρ, θ) são coordenadas polares, $B_{nm}(\rho, \theta)$ é a função base de Zernique de ordem (n, m) que é dada pela equação 3.19 e $f(\rho, \theta)$ é a função do suporte binário da forma do objeto em função de ρ e θ .

$$B_{nm}(\rho, \theta) = R_{nm}(\rho) e^{jm\theta}, p \leq 1 \quad (3.19)$$

$R_{nm}(\rho)$ é o polinômio radial de Zernike:

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left[\frac{n+|m|}{2} - s \right]! \left[\frac{n-|m|}{2} - s \right]!} \rho^{n-2s} \quad (3.20)$$

Assim, a forma pode ser descrita através de um conjunto de momentos, sendo que quanto maior a ordem, ou seja o número de momentos usados, maior será a quantidade de detalhes. Momentos de Zernike são invariantes a transformações geométricas. Momentos pseudo-Zernique são obtidos usando um polinômio radial com valor real.

Assim como outros descritores momentos de Zernique possuem desvantagens: necessidade de normalizar a forma em um disco unitário, grande dificuldade no cálculo do kernel, presença de características radiais e características circulares não consistentes, afinal, uma é do domínio espacial e outra do domínio da frequência, e também as características espectrais podem resultar em fatores menos significantes na descrição da forma [51].

3.6 Métodos Estruturais Baseados em Região

Assim como os métodos estruturais baseados em contorno, métodos estruturais baseados em região decompõem a região em subpartes para, então, descrevê-la.

3.6.1 Fecho Convexo

Esse descritor decompõe uma região em suas subregiões de forma que todas elas contenham somente partes convexas, ou seja, não deve existir nenhuma concavidade. Caso haja uma subregião não convexa a mesma é preenchida com uma ou mais subregiões mínimas de forma a tornar a mesma convexa.

Por definição uma região R é convexa se e somente se para quaisquer dois pontos $x_1, x_2 \in R$, todo o segmento traçado entre os pontos x_1 e x_2 está contido em R . Se houver alguma concavidade, parte do segmento ou todo ele passa por fora da região. O fecho convexo de R é na verdade a menor região convexa H que satisfaz a condição $R \subset H$, caso não haja nenhuma concavidade o fecho convexo da região é igual a ela própria. A diferença $H - R$ é chamada de deficiência convexa D da região R [51][22].

Para que não haja um excesso de pequenas deficiências convexas devido a ruído, falhas na digitalização e segmentação, deve ser feita uma suavização da região antes de realizar o particionamento.

Uma boa representação para forma consiste de uma árvore de concavidades construída através de um processo recursivo. Inicialmente o fecho convexo é obtido juntamente com suas deficiências convexas. Em seguida o fecho convexo e as deficiências convexas das mesmas também são obtidos. Esse processo é realizado sucessivamente até que não haja nenhuma deficiência convexa.

A figura 3.8(a) mostra o fecho convexo de uma região maior S . Suas deficiências são as partes mais claras nomeadas de S_1, S_2, S_3, S_4 e S_5 , sendo que S_1 e S_5

também possuem deficiências convexas. Na árvore de concavidades, figura 3.8(b), as deficiências de convexas de cada fecho são representadas como nós filhos.

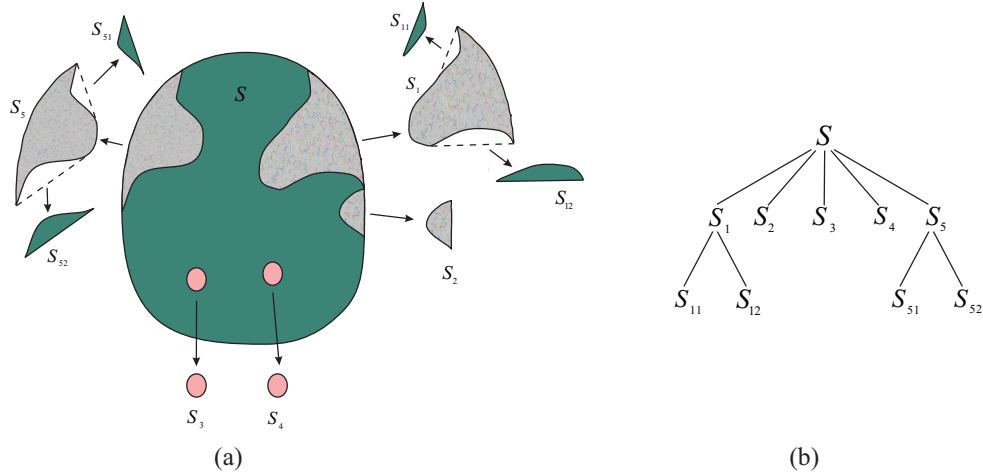


Figura 3.8: Fecho convexo (a)Fecho convexo com suas concavidades. (b) Árvore de concavidades

3.6.2 Eixo Medial

O eixo medial, que pode ser interpretado como sendo o esqueleto de uma forma, é obtido através da MAT (*Medial Axis Transform*), proposta por Blum [7]. Nessa abordagem a forma é preenchida com uma série de discos de raio máximo de maneira que os discos não ultrapassem suas fronteiras. A junção de todos os centros dos discos maximais resulta em uma linha que corresponde exatamente ao esqueleto da forma, figura 3.9. De maneira equivalente pode-se considerar que o esqueleto é constituído por todo ponto que possui distância mínima para pelo menos dois pontos da borda simultaneamente. A distância pode ser calculada diretamente da distância Euclidiana.

A partir do uso de discos com raios maximais, apenas um pequeno conjunto de distâncias precisa ser representado dentro dos círculos. Como pode ser observado na figura 3.10, o círculo possui 29 pontos, mas apenas 7 distâncias distintas

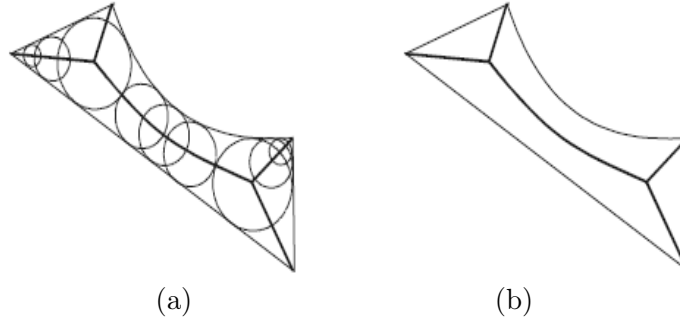


Figura 3.9: Eixo medial. (a)Circunferência com raios maximais. (b)União dos centros das circunferências

$\sqrt{32}$	5	$\sqrt{20}$	$\sqrt{17}$	4	$\sqrt{17}$	$\sqrt{20}$	5	$\sqrt{32}$
5	$\sqrt{18}$	$\sqrt{13}$	$\sqrt{10}$	3	$\sqrt{10}$	$\sqrt{13}$	$\sqrt{18}$	5
$\sqrt{20}$	$\sqrt{13}$	$\sqrt{8}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{8}$	$\sqrt{13}$	$\sqrt{20}$
$\sqrt{17}$	$\sqrt{10}$	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{10}$	$\sqrt{17}$
4	3	2	1	0	1	2	3	4
$\sqrt{17}$	$\sqrt{10}$	$\sqrt{5}$	$\sqrt{2}$	1	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{10}$	$\sqrt{17}$
$\sqrt{20}$	$\sqrt{13}$	$\sqrt{8}$	$\sqrt{5}$	2	$\sqrt{5}$	$\sqrt{8}$	$\sqrt{13}$	$\sqrt{20}$
5	$\sqrt{18}$	$\sqrt{13}$	$\sqrt{10}$	3	$\sqrt{10}$	$\sqrt{13}$	$\sqrt{18}$	5
$\sqrt{32}$	5	$\sqrt{20}$	$\sqrt{17}$	4	$\sqrt{17}$	$\sqrt{20}$	5	$\sqrt{32}$

Figura 3.10: Círculo de raio 3 representado dentro de uma grade 9x9. (Retirado de [30])

$(0, 1, \sqrt{2}, 2, \sqrt{5}, \sqrt{8}, 3)$. Isso ocorre em virtude do fato de se usar valores inteiros como coordenadas e também por causa da simetria que se tem no círculo. Por causa desse efeito o nome SAT (*Symmetry Axis Transform*) também é utilizado.

O conjunto de *pixels* dentro de um círculo muda de acordo com seu raio, assim como o número de distâncias. Contudo, o número de distâncias muda menos que o número de elementos dentro do círculo.

Uma maneira simples e eficiente para representar e armazenar a distância euclidiana é por meio da estrutura *Sorted Exact Distances Representation* (SEDR), proposta por Costa e Estrozi [11]. Essa estrutura armazena a distância euclidiana d , juntamente com sua posição relativa e pode ser representada através de

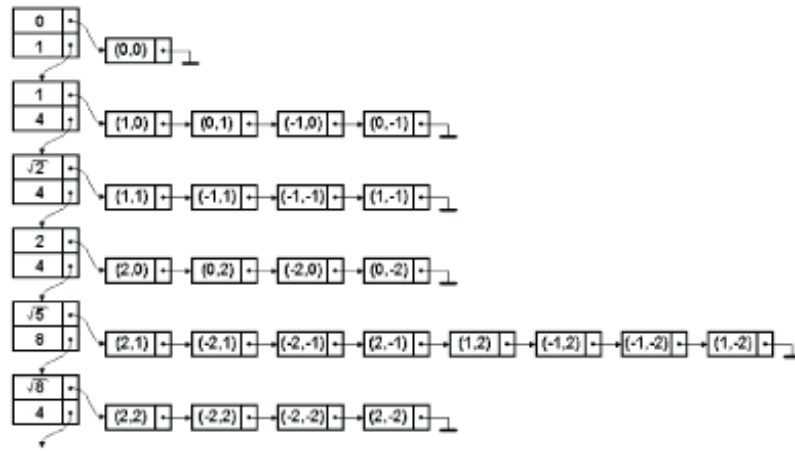


Figura 3.11: Estrutura de representação SEDR. (Retirado de [30])

uma lista ligada onde cada nodo (além da distância d) contém o número total de elementos equidistantes ed , um ponteiro para uma sublista que armazena as coordenadas relativas destes elementos com relação à origem e um ponteiro para o próximo elemento (figura 3.11). Essa é uma lista ordenada de forma ascendente pela distância.

Essa abordagem dispensa fazer referência direta à distância, pois como a lista é ordenada os índices fazem a correspondência com as mesmas. A vantagem é a possibilidade de se trabalhar com valores inteiros e não reais. Quanto às coordenadas dos pontos, utiliza-se representá-las em função da distância para o centro, pois a posição do mesmo é armazenada no nodo e tem-se a simetria que ajuda nessa forma de tratar os dados. Com isso, é suficiente armazenar os valores i e j (referentes à distância, por exemplo, distância $\sqrt{2}$ tem $i=1$ e $j=1$). As combinações possíveis são: (i,j) , $(i,-j)$, $(-i,-j)$ e $(-i,j)$. Contudo, algumas distâncias possuem um grande número de elementos equidistantes e essa quantidade é sempre múltipla de quatro. Para que tal problema seja resolvido é necessário repetir a distância na estrutura. Na tabela 3.2 é possível observar como as coordenadas são armazenadas

na estrutura SEDR.

DISTANCE	0	1	$\sqrt{2}$	2	$\sqrt{5}_1$	$\sqrt{5}_2$	$\sqrt{8}$	3	$\sqrt{10}_1$	$\sqrt{10}_2$	$\sqrt{13}_1$	$\sqrt{13}_2$
INDEX	0	1	2	3	4	5	6	7	8	9	10	11
(i, j)	(0,0)	(1,0)	(1,1)	(2,0)	(2,1)	(1,2)	(2,2)	(3,0)	(3,1)	(1,3)	(3,2)	(2,3)

Tabela 3.2: Coordenadas armazenadas na SEDR

Para cada distância euclidiana, todos os *pixels* do contorno são analisados frente a todas as posições relativas dentro daquela distância. Associa-se um rótulo ao *pixel* caso ele ainda não tenha sido associado a nenhum *pixel* do contorno. Esse procedimento é chamado de dilatação com propagação de rótulos.

No capítulo 4 outra maneira de se obter o esqueleto será discutida, afinal o esqueleto é utilizado na obtenção das saliências. É importante ressaltar que a estrutura SEDR não é utilizada no método proposto, uma vez que a transformada imagem floresta calcula todas as distâncias ao mesmo tempo que atribui rótulos, tornando desnecessários o armazenamento das distâncias e a inserção de discos maximais. Portanto, considera-se a abordagem baseada em transformada imagem floresta mais robusta na obtenção do esqueleto, mas não se pode tirar o mérito da abordagem através de SEDR, já que ela faz exatamente o que é proposto na definição de eixo medial, e esta por sua vez é uma definição clássica de esqueleto.

Depois da definição do esqueleto como sendo o eixo medial diversas outras surgiram, diferindo na maneira de encará-lo e por consequência de obtê-lo. Por exemplo, alguns autores consideram o esqueleto como sendo a linha central de uma forma, e como tal, seu procedimento para obter tal linha consiste de afinar sucessivamente a forma até que se chegue até ela.

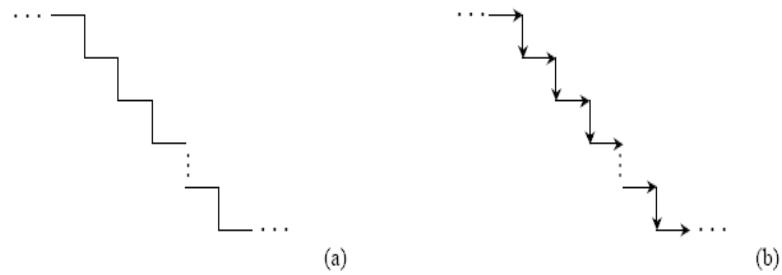


Figura 3.12: Estrutura de elementos primitivos em escada

3.7 Descritores Relacionais

Feita a descrição de fronteiras e de regiões, um outro passo a ser dado para obtenção de uma melhor descrição de uma imagem é descrever as relações estruturais entre os objetos (fronteiras e regiões). Assim como na abordagem estrutural, se utilizam regras de produções na forma de gramáticas para obtenção das relações a serem estabelecidas.

De maneira simplificada, uma estrutura em escada pode ser uma boa idéia para ilustrar como esse procedimento é possível: sendo a e b elementos primitivos, deseja-se descrever como os segmentos gerados da segmentação relacionam entre si de maneira a constituir um objeto - a escada. Assim a e b deverão representar, respectivamente, um segmento horizontal e um segmento vertical, com isso uma estrutura recursiva pode ser observada: após um a sempre haverá um b , em seguida outro a , e assim sucessivamente. A figura 3.12(a) mostra um elemento primitivo na forma de escada e a figura 3.12(b) mostra como esse elemento pode ser exibido sob a forma de primitivas de imagem.

Uma gramática pode ser estabelecida para que a estrutura codificada da figura acima possa ser gerada:

$$S \rightarrow aA$$

$$A \rightarrow bS$$

$$A \rightarrow b$$

A prova que a gramática realmente gera o que ela se propõe pode ser verificada da seguinte forma: a partir do símbolo inicial S , um elemento primitivo a é gerado juntamente com um símbolo não terminal A . Na segunda derivação a ser realizada, A gera um outro elemento primitivo b e retorna ao ponto inicial através de S , ou A simplesmente produz um b e pára em seguida. Assim qualquer estrutura da forma de uma escada pode ser gerada a partir da gramática.

Contudo, da forma tal qual foi colocada, essa maneira de se estabelecer a relação entre os elementos de uma fronteira não pode ser aplicada a formas bidimensionais. Assim, o uso de uma forma mais geral é recomendado, e uma forma geral que pode ser utilizada é a descrição de seções de imagens por segmentos de linha orientados, que não precisa ser unidos exclusivamente através de conexões cabeça-cauda. Com isso pode se definir operações entre as primitivas, sendo que essas operações constituem as conexões entre os segmentos. Soma e diferença, por exemplo, podem ser definidas como operações, ou seja, qualquer operação entre vetores pode ser realizada entre as primitivas abstratas, até mesmo a inversão de sua direção.

Algumas estruturas disjuntas, no entanto, não são bem descritas dessa forma. Elas necessitam de descrições através de árvores.

Para a representação da região da figura 3.13(a) na forma de árvore a relação ‘dentro de’ foi estabelecida. Tal árvore pode ser observada na figura 3.13(b). Cada nó da árvore deve conter, além de informações que o descrevam, informações que o relacionem com seus vizinhos.

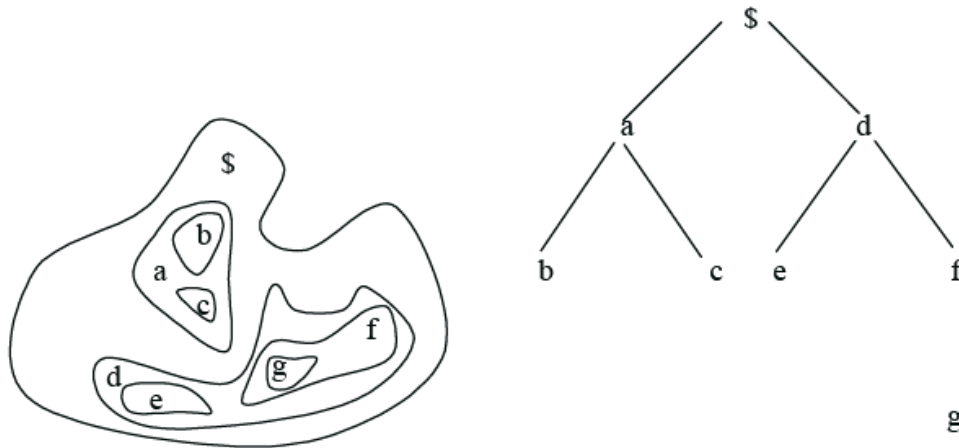


Figura 3.13: Descrição relacional. (a) Uma região composta simples. (b) Representação através de árvore

3.8 Medidas de similaridade

A partir do momento em que se tem uma forma descrita em termos de suas características, esta passa a ser considerada um padrão, e a recuperação baseada em forma é transformada em um problema de casamento de padrões. Daí em diante algoritmos de casamento de padrão e funções de distâncias métricas são usados para dizer o quão dois padrões são similares. Essa mudança de paradigma se justifica pelo fato de que criar uma maneira eficiente e compacta para representar a forma é mais fácil do que tentar comparar duas formas, pois além da simplicidade, diminui significativamente os requisitos de tempo e memória. Quando uma forma é transformada em um padrão, apenas as características mais importantes são armazenadas, daí o motivo do padrão ser chamado de vetor de características.

Da comparação entre dois vetores (ou conjuntos) de características deve resultar uma medida de similaridade, sendo que muitas vezes é utilizada a distância, que diz o quanto duas imagens estão distantes. Sendo assim, quanto menor a distância melhor o casamento e distância zero significa casamento perfeito. Como a distância

mede a dissimilaridade entre imagens, é comum usar o termo “dissimilaridade” para designar a distância.

3.8.1 Distância Euclidiana

Esta é a distância mais citada na literatura. Ela produz a menor distância real para chegar de um ponto a outro. Dados os pontos $p = (p_x, p_y)$ e $q = (q_x, q_y)$, a distância euclidiana, f_{euc} , é dada por:

$$f_{euc}(p, q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} \quad (3.21)$$

Embora seja de grande importância, não se pode negar que o custo computacional da mesma para grandes quantidades de dados é bastante elevado.

3.8.2 Distância de Minkowski

Dados dois pontos q e r , a distância de Minkowski, L_p , é dada por:

$$L_p(q, r) = \left(\sum_{i=0}^k |q_i - r_i|^p \right)^{\frac{1}{p}} \quad (3.22)$$

onde p é um parâmetro que modifica a distância de Minkowski resultando em outras distâncias conhecidas. Se $p = 2$ tem-se a distância euclidiana, se $p = 1$ tem-se distância de Manhattan, *city block* ou distância taxicab. Para p se aproximando do infinito, a distância se transforma na métrica do máximo:

$$\max_i (|q_i - r_i|) \quad (3.23)$$

Caso $p \geq 1$ tem-se uma distância métrica, e caso $p < 1$ a desigualdade triangular não é obedecida o que resulta na inexistência de distância métrica.

3.8.3 Distância Bottleneck

A distância de bottleneck é definida para conjuntos finitos de pontos e como tal, encontra a correspondência entre pontos de dois conjuntos através de uma função f . Sendo A e B dois conjuntos de pontos, ambos de tamanho n , e $d(a, b)$ a distância entre dois pontos, a distância de bottleneck $F(A, B)$ é o mínimo sobre todas as correspondências 1:1 entre A e B , onde cada correspondência é dada pela distância máxima $d(a, f(a))$.

Diversas alternativas podem ser utilizadas na distância $d(a, b)$ entre dois pontos, dentre elas a distância L_p . Contudo aproximações mais rápidas costumam ser utilizadas [48; 49].

Transformações geométricas também podem ser tratadas na distância de bottleneck. Embora o tempo requerido seja bastante grande, o problema de decisão para translação, por exemplo, possui diversas alternativas [48].

A distância de bottleneck possui variações. Dentre elas pode-se destacar: distância do peso mínimo, a distância mais uniforme e a distância do desvio mínimo.

3.8.4 Distância de Hausdorff

A distância de Hausdorff é um método clássico de casamento de formas baseado em correspondência que é utilizado para localizar objetos e medir a similaridade entre formas [51]. Ela trata cada ponto da forma como se fosse um ponto de característica. Dados 2 conjuntos de pontos referentes a duas formas, a distância de Hausdorff entre as formas A e B é dada por:

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (3.24)$$

onde

$$h(A, B) = \max_{a \in A} \min_{b \in B} d(a, b) \quad (3.25)$$

$d(a, b)$ é a distância entre os pontos de A e B , e normalmente se utiliza distância euclidiana. Portanto, $h(A, B)$ é o máximo obtido sobre todos os pontos em A das distâncias para um ponto de B . Devido a sensibilidade ao ruído, melhorias na distância de Hausdorff foram feitas e podem ser encontradas na literatura. Zhang e Lu [51] abordam algumas destas melhorias. Embora a distância de Hausdorff permita o casamento parcial entre formas, não possui invariância à mudança de escala, rotação ou translação, fato que restringe bastante o seu uso.

3.8.5 Similaridade da Função Assinatura

Levando em consideração a assinatura da forma, (seção 3.3.2) uma maneira de realizar a comparação de duas assinaturas deve ser definida. Usualmente, essa comparação é feita através de um emparelhamento dos elementos das assinaturas A e B , sendo que este emparelhamento não precisa ser um-para-um: o par do elemento i de A para o elemento j de B , pode ser seguido por um par de $(i, j + 1)$, $(i + 1, j)$, ou $(i + 1, j + 1)$. Um elemento pode ser um vértice ou um segmento. A partir do conjunto de pontos emparelhados a distância euclidiana pode ser calculada.

Holmes et al. [25] descreve uma maneira eficiente de encontrar a similaridade entre assinaturas. Com base no fato de que a distância euclidiana varia bastante diante de pequenas alterações nas imagens para as quais deseja-se comparar suas assinaturas, Holmes et al. propõe o uso do problema clássico do transporte para resolver esta questão. O problema do transporte visa minimizar o custo de entregar quantias integrais de bens produzidos por n depósitos para m mercados e ainda equilibra abastecimento e demanda. Como cada depósito está conectado a

um mercado isso gera um transporte de remessa entre nós sem intermediários, ocasionando um custo unitário $c_{i,j}$ de transportar de um depósito i para um mercado j .

Na aplicação do problema de transporte ao casamento de assinaturas, Holmes et al. considera uma assinatura abastecedora de bens e outra assinatura requeridora de bens: enquanto os elementos de uma assinatura são interpretados como depósitos os elementos da outra são interpretados como mercados. A partir disso uma solução para o problema clássico é aplicada [25].

3.8.6 Distância de Fréchet

Em alguns casos, pode ser de interesse uma correlação entre as fronteiras comparadas. Por exemplo, quando a distância de Hausdorff é utilizada em curvas a distância de um ponto na fronteira A para um ponto da fronteira B pode ser pequena. Caso fosse desejável que a relação de espaço R , entre o ponto inicial e o ponto para o qual se está medindo a distância, ambos na fronteira A , fosse a mesma para a fronteira B , a distância de Hausdorff não traria resultados satisfatórios.

A distância de Fréchet trabalha através de incrementos na posição a medida que calcula a distância entre as fronteiras A e B : a partir de um ponto inicial em cada fronteira caminha-se nas duas fronteiras simultaneamente calculando a distância para pontos correspondentes. Formalmente, tomando duas curvas parametrizadas, $A(\alpha(t))$ e $B(\beta(t))$, definidas a partir de A e B , sendo que suas respectivas parametrizações α e β são funções contínuas do mesmo parâmetro $t \in [0, 1]$, tal que $\alpha(0) = \beta(0) = 0$ e $\alpha(1) = \beta(1) = 1$. A distância de Fréchet é a máxima distância $d(A(\alpha(t)), B(\beta(t)))$, tal que $\alpha(t)$ e $\beta(t)$ são dados através do mínimo sobre todos os incrementos monotônicos de parametrização. Para o domínio discreto das imagens 2D $t \in \{0, 1\}$.

3.8.7 Distância City-Block

A métrica *city block*, também conhecida como distância de Manhattan, é baseada em vizinhança-de-4 e, como consequência, só permite caminhar em quatro direções para ir de um ponto a outro. Com isso, a distância de Manhattan nem sempre fornece a menor distância em linha reta entre dois pontos assim como a distância euclidiana, mas é bastante utilizada por seu cálculo ser menos custoso do que o da distância euclidiana. Dados dois pontos $p = (p_x, p_y)$ e $q = (q_x, q_y)$ a distância *city block* é definida da seguinte forma:

$$dist(p, q) = |p_x - q_x| + |p_y - q_y| \quad (3.26)$$

Portanto, se p e q são vizinhos-de-4 a distância entre eles é 1, caso ambos sejam vizinhos diagonais a distância entre eles é 2. Por esta razão a métrica *city block* também é referenciada como métrica “1-2”. Dado um ponto p , seus *pixels* adjacentes na vizinhança-de-4 têm distância 1 e seus *pixels* adjacentes na vizinhança diagonal têm distância 2 [36].

3.8.8 Distância Chessboard

A métrica *chessboard* possui esse nome porque saindo de um *pixel* p é possível mover em qualquer direção, assim como os movimentos de um rei no tabuleiro de xadrez. Portanto, esta métrica é baseada em uma vizinhança-de-8 [36].

Assim como na distância *city block*, a maior motivação para o uso da métrica *chessboard* é o fato desta poder substituir a distância euclidiana com maior eficácia, seu cálculo é mais rápido pois gasta um número menor de operações de cpu. Dados dois pontos $p = (p_x, p_y)$ e $q = (q_x, q_y)$ a distância *chessboard* é definida como:

$$dist(p, q) = \max(|p_x - q_x|, |p_y - q_y|) \quad (3.27)$$

3.8.9 Casamento Elástico

O método do casamento elástico utiliza um modelo deformado gerado a partir da soma de um modelo original $\sigma(s)$ e uma deformação $\theta(s)$ [6]. O modelo original é uma aproximação de uma curva real gerada a partir de funções de interpolação:

$$\phi(s) = \sigma(s) + \theta(s) \quad (3.28)$$

onde $\sigma = (\sigma_x, \sigma_y)$ e $\theta = (\theta_x, \theta_y)$. A partir do modelo deformado obtido e de um objeto na imagem, a similaridade é medida pela minimização de uma função composta:

$$\begin{aligned} F &= S + B + M \\ &= \alpha \int_0^1 \left[\left(\frac{d\theta_x}{ds} \right)^2 + \left(\frac{d\theta_y}{ds} \right)^2 \right] ds \\ &\quad + \beta \int_0^1 \left[\left(\frac{d^2\theta_x}{ds^2} \right)^2 + \left(\frac{d^2\theta_y}{ds^2} \right)^2 \right] ds \\ &\quad + \int_0^1 I_E(\phi(s)) ds \end{aligned} \quad (3.29)$$

sendo que I_E é o objeto de imagem, S e B são chamados de energia de tensão e energia de curva respectivamente, e M é o grau de sobreposição entre o modelo deformado e o objeto na imagem. Contudo, a medida obtida da composição das três funções ainda não é suficiente pois não consegue diferenciar corretamente uma imagem da outra. São necessários ainda os parâmetros:

N - é a complexidade da forma e é medida como a quantidade de zeros da função de curvatura associada com os modelos dos contornos.

C - é a correlação entre as funções de curvatura do modelo original e do modelo

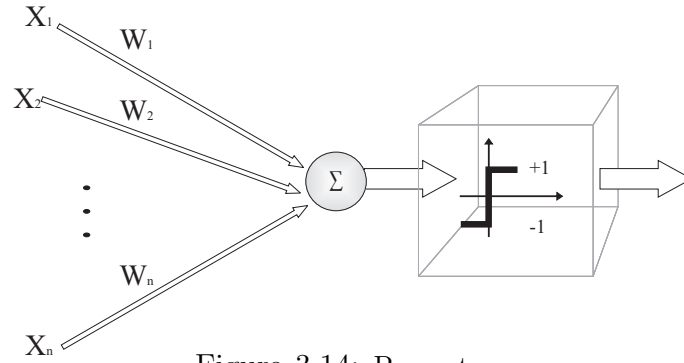


Figura 3.14: Perceptron

deformado.

O último passo a ser dado para comparar duas imagens é tomar os cinco parâmetros (S, B, M, N, C) e usá-los em uma rede neural de retro-propagação para que seja feita a classificação [6].

Em redes neurais busca-se obter resultados a partir de uma simulação da forma tal qual se acredita que estejam organizados, no cérebro, os neurônios e suas inúmeras ligações. Nessa simulação os neurônios são representados por elementos de computação não linear, os perceptrons, e as ligações entre eles são dadas por uma organização em rede desses elementos.

O perceptron é capaz de aprender uma função de decisão linear que distingue dois conjuntos de treinamento linearmente separáveis. Para isso, utiliza uma soma ponderada das entradas. Considerando $X = (x_1, x_2, \dots, x_n)$ como um vetor de características (padrão), e $W = (w_1, w_2, \dots, w_n, w_{n+1})$ um conjunto de pesos a serem aplicados ao vetor. A soma ponderada é dada na equação 3.30 e o perceptron é ilustrado na figura 3.14.

$$d(X) = \left(\sum_{i=1}^n w_i x_i \right) + w_{n+1} \quad (3.30)$$

Os pesos tentam traduzir o que ocorre no cérebro humano chamado sinapse ¹,

¹Sinapse é o meio pelo qual os neurônios estão conectados uns aos outros. As sinapses transmi-

já que modificam as entradas, conforme o referido processo.

No modelo ilustrado na figura 3.14, há também uma função que mapeia a saída da soma na saída final do dispositivo, e é chamada de função de ativação. Tal função se comporta como um elemento de limiarização definindo a saída do perceptron como $+1$ se $d(X) > 0$, o que significa que o padrão pertence à classe w_1 , e como -1 no caso de $d(X) < 0$. Neste último caso o padrão X pertence a w_2 . Se a função de decisão resultar em 0, tem se especificamente o caso em que o padrão está sobre a superfície de decisão que separa as duas classes. Assim não se pode determinar exatamente a qual classe pertence o referido padrão.

Assim o problema se restringe a encontrar W usando um conjunto de treinamento formado por padrões de cada uma das duas classes. Um algoritmo de treinamento é utilizado para ajustar os pesos, até que X possa ser atribuído corretamente à sua classe.

Contudo, na prática a maioria dos problemas envolve mais de uma classe, por isso foi desenvolvido o modelo de redes neurais multicamadas *feedforward*. A arquitetura básica dessa rede consiste em camadas de elementos computacionais, como perceptrons, estruturalmente idênticos, que são arranjados de maneira que a saída de cada neurônio em uma camada alimenta a entrada de cada neurônio da camada seguinte. Nesse tipo de rede normalmente se utiliza treinamento por retro-propagação onde o valor de saída, assim que conhecido, ajusta as entradas das camadas anteriores. O erro é propagado para trás na rede começando com o erro na camada de saída, o objetivo é minimizar tal erro.

Como se pode constatar, o método implica em grande custo computacional o que torna inviável o seu uso em recuperação de imagens *online*. Em uma comparação com os sistemas de recuperação de imagens QBIC [35] e QVE [24], os autores

tem estímulos através de diferentes concentrações de $Na+$ (Sódio) e $K+$ (Potássio), e o resultado disto pode ser estendido por todo o corpo humano.

demonstraram que o número de operações de CPU é menor do que em ambos, contudo, o número de passos no processo de deformação torna o casamento de alto custo. Outra desvantagem é que o descritor não é invariante à rotação e o tipo de deformação é arbitrária ou dependente da aplicação [51].

3.9 Considerações Finais

A recuperação baseada em forma depende principalmente de bons métodos para a representação e descrição de formas, pois, comparar uma imagem consulta inteira com cada imagem de uma base de dados é algo que pode ser bastante demorado. As características das imagens da base de dados são extraídas através de descritores e só então a busca é realizada. Portanto, a busca se reduz a um problema de casamento de padrão: as características da imagem consulta são comparadas com as características de cada imagem na base de dados. Neste ponto entram as medidas de similaridade, pois através destas se pode medir o quão dois padrões, representantes de duas imagens, são parecidos. Distância euclidiana é uma das medidas mais utilizadas. Contudo, se requisitos de tempo são importantes, utiliza-se uma medida mais fácil de calcular, tal como distância *city block* ou *chessboard*.

Várias categorias de métodos para representar e descrever formas podem ser encontradas, e dentro de cada uma a quantidade de métodos é bastante considerável, o que totaliza uma quantidade gigantesca de maneiras de transformar uma forma em um pequeno conjunto de características. Neste capítulo alguns destes métodos foram discutidos dando uma visão geral sobre a representação e descrição de formas, sendo que bons descritores devem possuir diversas características, tais como: ser conciso, descrever eficientemente a forma e, principalmente, ser invariante a transformações geométricas. Assim normalizações também é um assunto importante dentro de descritores de forma, pois através delas um descritor pode

se tornar invariante, sendo assim, estas também foram discutidas neste capítulo.

Dentre os principais descritores que se pode citar tem-se os descritores de Fourier que possuem uma ampla gama de aplicação, já que são capazes de descrever forma e textura de imagens, além disso são amplamente usados e difundidos o que os torna alvo de constantes adaptações que tem trazido melhorias importantes em contextos bem definidos. Descritores simples embora não possam ser usados separadamente também merecem destaque pois constituem uma maneira simples e rápida de evitar falsos acertos. O descritor assinatura da forma também merece destaque, pois o fato de transformar uma fronteira em função unidimensional o torna bastante útil, alguns descritores o utilizam como modo de representar a forma, para só então descrevê-la.

Um descritor novo que vem se destacando devido a invariância à transformações geométricas é saliências do contorno, por isso, este terá um capítulo inteiro dedicado a ele.

Capítulo 4

Saliências do Contorno

4.1 Introdução

Um descritor proposto recentemente merece destaque. É o descritor saliências do contorno, que pode ser obtido através do esqueleto [47; 46; 45].

Saliências capturam a própria essência da forma uma vez que qualquer parte sobressalente ou mesmo reentrância contidas na forma são consideradas saliências e estas são justamente as características que diferenciam uma forma da outra. Aplicações utilizando esta maneira de tratar a forma é de grande utilidade em descrição de imagens. Saliências do contorno configura-se um descritor aplicado exclusivamente a formas. Uma outra denominação para saliências do contorno é saliências da forma, já que uma forma pode ser expressa através de seu contorno. Trata-se de um descritor robusto a transformações geométricas como rotação, translação e mudança de escala.

Uma distinção que pode ser feita é quanto ao referencial da saliência. No caso de saliência do contorno tem-se uma abordagem relativa ao *pixel*. Torres et al. [45] apresenta também saliências do segmento (SS - *Segment Saliences*), que é um enfo-

que baseado em segmento. Nas próximas seções são abordadas maneiras distintas de obtenção da saliência: uma com base somente na área de influência e outra com base na associação ao esqueleto. Em ambas utiliza-se a IFT (Image Foresting Transform), que será discutida na próxima seção. O capítulo é finalizado com uma seção dedicada ao algoritmo de casamento utilizado no descritor saliências.

4.2 Transformada Imagem-Floresta

Uma maneira muito eficiente para se encontrar as saliências é apresentada em [47; 46; 45]. Trata-se da utilização de grafos para a representação da imagem através de seus relacionamentos de conexidade entre nodos e arcos, sendo os nodos *pixels* e os arcos relações de adjacência entre os *pixels*. Tal técnica utilizada é denominada Transformada Imagem-Floresta (IFT), sendo que Falcão aborda amplamente o assunto [17; 19; 18; 16]. A IFT é uma transformada projetada para operadores de processamento de imagem que gera uma representação multiescala.

Para determinar se um *pixel* é de saliência é preciso saber a sua área de influência, ou seja, é preciso definir quais os *pixels* que estão sob a influência de um ponto do contorno. Dessa forma conhecer o relacionamento de conexão estabelecido entre os *pixels* é fundamental. A principal vantagem dessa abordagem é que ela transforma o problema de particionar a imagem em um conjunto de regiões discretas ao cálculo de um caminho de custo mínimo na floresta utilizando uma função custo. Assim, a IFT calcula simultaneamente a Transformada da Distância Euclidiana (EDT- *Euclidean Distance Transform*) e as regiões discretas de Voronoi. Cada ponto do contorno é tomado como uma raiz e a partir desta um caminho de custo mínimo é determinado. Cada caminho constitui uma árvore e a união delas constitui uma floresta que cobre a imagem inteira. A IFT gera duas diferentes representações da forma: contornos multiescala por dilatação exata e esqueletos

multiescala por propagação de rótulos.

O algoritmo da IFT utiliza um conjunto de sementes S previamente rotuladas com o objetivo de propagar seus respectivos rótulos por toda a imagem, sendo que todo pixel p pertencente ao contorno está no conjunto S .

O algoritmo 1, proposto por Falcão et al. [16], obtém a floresta de caminho ótimo. Para isso percorre a imagem empilhando/desempilhando *pixels* enquanto:

- calcula o custo do caminho armazenando-o no mapa de custo C
- constrói o mapa raiz R

Sejam $p = (x_p, y_p)$ e $q = (x_q, y_q)$ *pixels* quaisquer. Uma relação de adjacência euclidiana A é definida entre os dois:

$$q \in A(p) \Leftrightarrow (x_q - x_p)^2 + (y_q - y_p)^2 < \rho^2 \quad (4.1)$$

onde $\rho = \sqrt{2}$. A partir desse relacionamento definido entre os *pixels*, o custo do caminho é calculado também baseado em distância euclidiana. Dado o caminho, $\pi = \{(x_{p_1}, y_{p_1}), (x_{p_2}, y_{p_2}), \dots, (x_{p_n}, y_{p_n})\}$, seu custo é definido como:

$$f_{euc}(\pi) = \begin{cases} \sqrt{(x_{p_n} - x_{p_1})^2 + (y_{p_n} - y_{p_1})^2}, & \text{se } p_1 \in S, \\ +\infty, & \text{se } p_1 \notin S \end{cases} \quad (4.2)$$

Toda raiz possui um rótulo e cada *pixel* na imagem possui uma única raiz e por consequência um rótulo. Para saber qual a raiz de um certo *pixel* é preciso conhecer qual o custo mínimo entre este *pixel* e o contorno. O custo é calculado de acordo com a equação 4.2 e caso seja mínimo entre um *pixel* raiz (pertencente ao contorno) e um *pixel* qualquer, o rótulo da raiz é atribuído a este *pixel*.

Algoritmo 1 IFT

Entrada: (1)Imagem I , (2)Conjunto de sementes S devidamente rotuladas,
(3)Relação de Adjacência Euclidiana A .

Saída: (1)Uma floresta de caminho ótimo P , (2)Mapa de custos C ,
(3)Mapa raiz R .

Estrutura Auxiliar: Fila de prioridade Q

```

1: para todo pixel  $p$  da imagem  $I$  faça
2:    $C(p) \leftarrow +\infty$ 
3: fim para
4: para todo pixel  $p \in S$  faça
5:    $P(p) \leftarrow nil$ 
6:    $R(p) \leftarrow p$ 
7:    $C(p) \leftarrow 0$ 
8:   Insira  $p$  em  $Q$ 
9: fim para
10: enquanto  $Q$  não está vazia faça
11:   Remova um pixel  $p = (x_p, y_p)$  de  $Q$  tal que  $C(p)$  é mínimo
12:   para cada pixel  $q = (x_q, y_q)$  tal que  $q \in A(p)$  e  $C(q) > C(p)$  faça
13:      $C' \leftarrow (x_q - x_{R(p)})^2 + (y_q - y_{R(p)})^2$ , onde  $(x_{R(p)}, y_{R(p)})$  é raiz de  $p$ 
14:     se  $C' < C(q)$  então
15:       se  $C(q) \neq +\infty$  então
16:         Remova  $q$  de  $Q$ 
17:       fim se
18:        $P(q) \leftarrow p$ 
19:        $R(q) \leftarrow R(p)$ 
20:        $C(q) \leftarrow C'$ 
21:       Insira  $q$  em  $Q$ 
22:     fim se
23:   fim para
24: fim enquanto

```

Assim como pode ser observado no algoritmo para se obter a IFT, inicialmente todas as sementes possuem custo zero, pois não é preciso caminhar para encontrar uma semente quando se está em uma. Enquanto isso, o custo de um *pixel* qualquer deve ser o maior possível (infinito), pois não se sabe a princípio o seu custo real. Deve ser definido que a raiz de uma semente é sempre ela mesma e que a mesma não possui um predecessor. Os demais *pixels*, em algum momento, receberão um

predecessor e também uma raiz da qual receberá um rótulo. Todos os *pixels* somente deverão ser inseridos em uma fila para que se saiba de onde se deve partir na construção dos grafos orientados. Assim, cada semente é a raiz de uma árvore e o conjunto de árvores define uma floresta cobrindo a imagem inteira.

No próximo passo os *pixels* p de custo mínimo são removidos da fila, um de cada vez, e sua vizinhança é analisada. Quando o custo de um *pixel* q na vizinhança de p é maior do que o custo deste último significa que q pode estar na árvore errada. Para resolver o impasse, um novo custo C' é calculado considerando que ele está na árvore de seu vizinho p . Se este custo for menor que o seu custo atual, então q deve ser re-arranjado. Seu predecessor, sua raiz e seu custo devem ser modificados: o predecessor passa a ser p , a raiz passa a ser a raiz do seu predecessor e o custo passa a ser C' . Em seguida, q deve ser inserido na fila para que se possa prosseguir na árvore gerada. Caso o antigo custo q fosse infinito, isso implicaria que q nunca teve seu custo calculado e que portanto nunca esteve na fila. Caso contrário, q está na fila e um caminho passando por ele está sendo traçado. Portanto essa condição deve ser verificada para impedir que q seja inserido duas vezes na fila, pois isso implicaria em dois caminhos passando por ele e tornaria o resultado incoerente.

Um mapa de rótulos L pode ser construído com base no mapa raiz, ou mesmo durante o próprio algoritmo da IFT: cada vez que um *pixel* recebe uma raiz, muda-se o seu rótulo. Portanto, ao final da IFT, o mapa de rótulos L pode ser exibido como uma imagem na qual não se pode distinguir o contorno, mas que fornece importantes informação sobre as regiões em que cada *pixel* exerce influência, ou seja regiões discretas de Voronoi 4.7(c).

4.3 Rotulação dos Pixels do Contorno

Para que a IFT seja aplicada é necessário que os *pixels* estejam rotulados ou pelo menos que o conjunto de sementes S tenha sido determinado. Assim, um passo anterior a IFT deve ser dado nesse sentido, e tal passo vem a ser exatamente o algoritmo de rotulação que encontra as sementes e em seguida atribui rótulos incrementais as mesmas começando com o valor 1. Assim, os rótulos variam no intervalo $1, 2, \dots, N$ onde N é o número de sementes.

O algoritmo de rotulação passa por algumas dificuldades quando encontra formas muito intrincadas ou muito finas pois certas ambigüidades podem ocorrer. Na figura 4.1 é possível ver alguns destes casos ambíguos. Em 4.1(a) o objeto é muito fino e todos os seus *pixels* são considerados como sendo de borda já que não é possível definir quais pertencem a borda e quais pertencem ao interior da forma. Em 4.1(b) partes espúrias decorrentes de falhas de segmentação dificultam a rotulação dos *pixels*. E nos casos 4.1(c)-(e), a ambigüidade é facilmente verificada: não se pode saber ao certo se a borda é de um único objeto ou se são dois objetos que se tocam.

A fim de minimizar os problemas citados, algumas restrições na definição de borda devem ser feitas. Esta passa a ser definida como um contorno fechado 8-conexo e orientado. Com isso, os *pixels* em um caminho qualquer são 8-adjacentes e a borda deverá separar nitidamente o interior do exterior do objeto: o interior estará sempre a esquerda do contorno seguindo o sentido anti-horário, e a direita se for seguido o sentido horário [15].

Para que os *pixels* possam ser rotulados corretamente algumas normas devem ser obedecidas:

1. Um único contorno por borda de objeto deve ser identificado;
2. Linhas e ramos devem ser eliminados (figura 4.1(a) e (b));

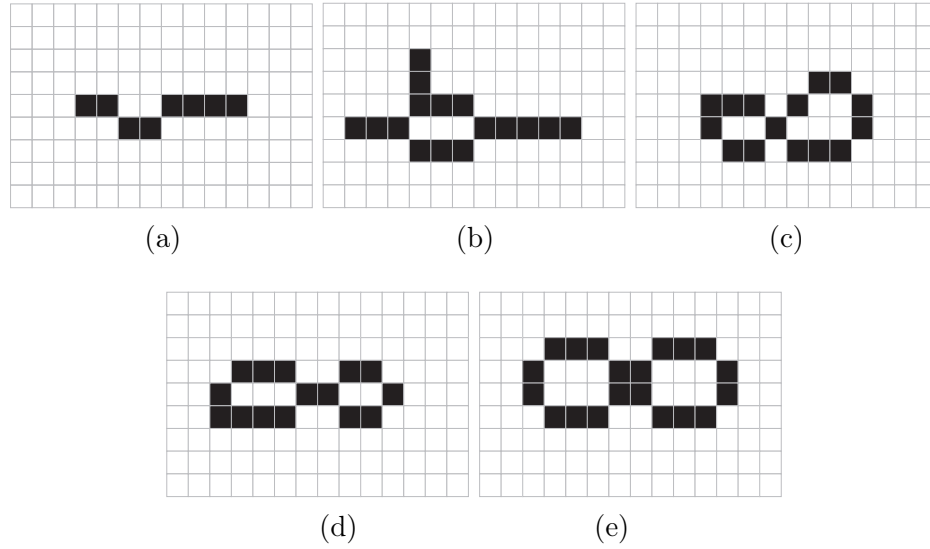


Figura 4.1: Casos ambíguos. (a) Linha. (b) Contorno com ramos. (c) Contorno com pixel de estrangulamento. (d) Dois contornos que se tocam. (e) Contorno que contém dois lados que se tocam.

3. Bordas com *pixel* de estrangulamento (figura 4.1(c)) devem ser tratadas como contornos com ramos iniciando nesses *pixels*;
4. Casos de dois contornos que se tocam (figura 4.1(d)) devem ser tratados como tal;
5. Casos de contornos que tocam seu lado oposto (figura 4.1(e)) devem ser tratados como tal.

Quanto ao caso da figura 4.1(a), em que os objetos possuem um *pixel* de largura, a resolução mais simples e viável é admitir que o objeto tenha passado por dilatação antes da aplicação do algoritmo, caso contrário as bordas não podem ser extraídas. Nos casos 4.1(b), 4.1(d) e 4.1(e) deve se evitar arcos entre *pixels* 8-adjacentes que passem por dentro e por fora do objeto, para cada arco $(p, q) \in A_8$ (figura 4.2(a)), onde p e q são *pixels*, existe o conceito de *pixel* à esquerda $e(p, q)$ e *pixel* à direita $d(p, q)$ do arco (figura 4.2(c)-(d)). Um arco passa por dentro ou por fora do objeto

quando $I(e(p, q)) = I(d(p, q))$ e portanto não é considerado um arco válido. Já o caso da figura 4.1(c) exige adicionalmente que o *pixel* de estrangulamento seja localizado, pois a corretude na rotulação depende de que a mesma seja iniciada em um *pixel* imediatamente depois do *pixel* de estrangulamento, evitando que se passe duas vezes pelo mesmo: se isso ocorresse tal *pixel* receberia um rótulo para em seguida receber um outro distinto, provocando erro na rotulação. Um *pixel* de estrangulamento é caracterizado por possuir 4 arcos partindo dele, sendo que todos obedecem a norma de que $I(e(p, q)) \neq I(d(p, q))$. Assim, uma rotina de localização deste *pixel* deve ser usada, seguida do algoritmo de rotulação.

A rotulação correta nos casos ambíguos pode ser vista na figura 4.3.

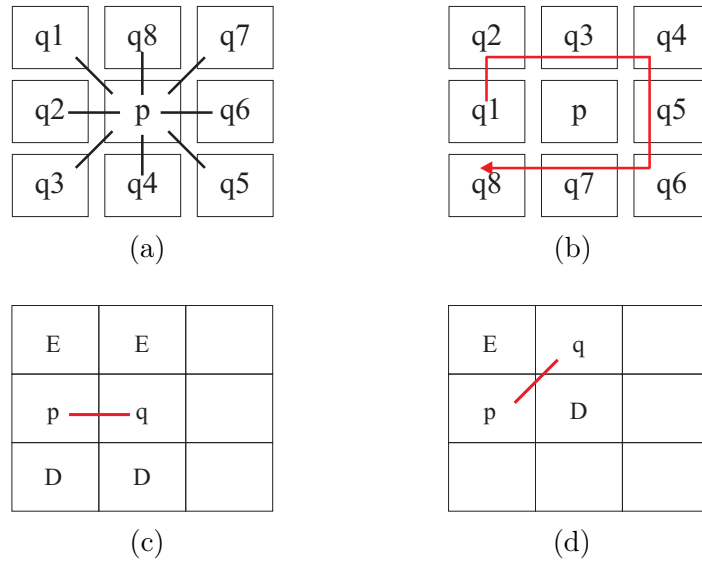


Figura 4.2: Relações na vizinhança do pixel. (a) Arcos na vizinhança-de-8. (b) Pixels no sentido horário. (c)-(d) Pixel à direita D e à esquerda E do arco entre p e q

A estratégia do algoritmo de rotulação é visitar cada *pixel* do contorno uma única vez caminhando no sentido anti-horário (ou horário) [15]. Assim que a busca for feita e todos os *pixels* do contorno tiverem sido encontrados o algoritmo atribui um rótulo a cada um. Uma ordem para os *pixels* na vizinhança-8 deve ser determi-

nada. Na figura 4.2(b) essa ordem pode ser observada: tomando um *pixel* central p , os *pixels* são percorridos em sentido horário partindo do *pixel* correspondente ao vizinho-de-4 esquerdo q_1 .

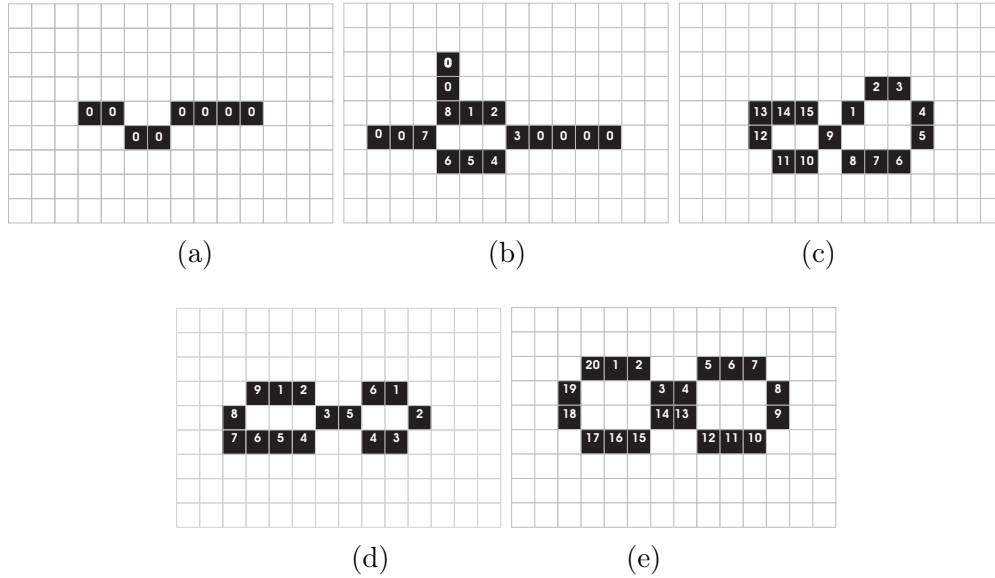


Figura 4.3: Rotulação de casos ambíguos. (a) Linha. (b) Contorno com ramos. (c) Contorno com pixel de estrangulamento. (d) Contornos que se tocam. (e) Contorno que toca o lado oposto.

Dados os parâmetros de entrada e saída assim como estruturas auxiliares, o algoritmo de rotulação (algoritmo 2) é dado a seguir. A imagem de cores C , que é usada como estrutura auxiliar, diz respeito a utilização da pilha: a cor $C(p)$ de um *pixel* qualquer pertencente à imagem é **branco** se nunca foi inserido na pilha, **cinza** se está na pilha e **preto** se já foi removido da pilha.

Algoritmo 2 Rotulação

Entrada: Imagem binária I a ser rotulada**Saída:** Imagem rotulada L , na qual cada pixel do contorno possui um rótulo**Estrutura Auxiliar:** (1) Pilha Q , (2) Imagem de cores C , (3) Imagem P de predecessores, (4) Imagem B de bordas.

```

1: para todo pixel  $p$  pertencente a  $I$  faça
2:    $L(p) \leftarrow 0$ 
3:    $B(p) \leftarrow 0$ 
4:    $C(p) \leftarrow \text{branco}$ 
5:   se  $I(p) = 1$  e existe  $q \in A_4(p)$  tal que  $I(q) = 0$  então
6:      $B(p) \leftarrow 1$ 
7:   fim se
8: fim para
9: para todo pixel  $s$ ,  $s \in I$ , tal que  $B(s) = 1$  e  $C(s) = \text{branco}$  faça
10:  se existe  $q$ ,  $q \in A_8(s)$ , tal que  $B(s) = 1$  e  $I(e(s, q)) \neq I(d(s, q))$  então
11:     $P(s) \leftarrow \text{nil}$ 
12:     $C(s) \leftarrow \text{cinza}$ 
13:    Insira  $s$  em  $Q$ 
14:    enquanto  $Q$  não estiver vazia faça
15:      Remova  $p$  de  $Q$ 
16:       $C(p) \leftarrow \text{preto}$ 
17:      para todo  $q$ , tal que  $q \in A_8(p)$  faça
18:        se  $q = s$  e  $P(p) \neq s$  então
19:          vá para a linha 32
20:        fim se
21:        se  $B(q) = 1$  e  $C(q) \neq \text{preto}$  e  $I(e(s, q)) \neq I(d(s, q))$  então
22:           $P(q) \leftarrow p$ 
23:          se  $C(p) = \text{branco}$  então
24:            Insira  $q$  em  $Q$ 
25:             $C(q) \leftarrow \text{cinza}$ 
26:          fim se
27:        fim se
28:      fim para
29:    fim enquanto
30:     $l \leftarrow 1$ 
31:    Esvazie  $Q$ 
32:    enquanto  $p \neq \text{nil}$  faça
33:       $L(q) = l$ 
34:       $p \leftarrow P(p)$ 
35:       $l \leftarrow l + 1$ 
36:    fim enquanto
37:  fim se
38: fim para

```

Na figura 4.7(b) pode ser visto o resultado da rotulação da figura 4.7(a). A rotulação se inicia do valor 1 gerando tons muito próximos do preto, a medida que os rótulos são incrementados os tons se aproximam do branco.

4.4 Saliências Através da Área de Influência

Saliências podem ser determinadas a partir dos *pixels* e para avaliar se um *pixel* é de saliência é necessário conhecer a área sob a qual este exerce influência, sendo que quanto maior o valor atribuído a tal área maior a probabilidade deste *pixel* ser de saliência. A figura 4.4 mostra as áreas de influência interna e externa dos pontos A e B , sendo que A possui área externa grande enquanto que B possui área interna grande. Portanto, ambos são pontos de saliência.

Um algoritmo de dilatação exata com propagação de rótulos (EDLP - *Exact Dilation with Label Propagation*) [47] ajuda a determinar a área de influência e portanto as saliências. Neste contexto o algoritmo EDLP utilizado é a transformada imagem floresta (IFT). Diz-se que o algoritmo é de dilatação exata com propagação de rótulos por que a partir de um contorno rotulado, ele expande esses rótulos para dentro e para fora do contorno (dilação) propagando-os de forma coerente à medida que realiza essa tarefa. A rotulação é coerente pois um *pixel* no interior do contorno possui o mesmo rótulo de um *pixel* do contorno somente quando está mais próximo deste do que de qualquer um outro deste contorno.

Quando a imagem está completamente rotulada, e conta-se com um mapa de rótulos L , é dito que ela está dividida em regiões discretas de Voronoi, que podem ser também obtidas a partir do diagrama de Voronoi [9]. Tal diagrama foi criado com o propósito de resolver um problema simples: como dividir uma cidade em áreas irregulares de forma que a área coberta por um carteiro vinculado a uma determinada agência de correio seja otimizada: nenhum carteiro deverá ficar so-

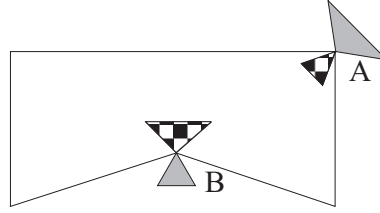


Figura 4.4: Partes hachuradas são a área de influência interna e as partes cinza são a área de influência externa do ponto. A é uma saliência convexa e B é uma saliência côncava

brecarregado enquanto outro está com pouco trabalho. Neste caso particular a divisão da imagem em regiões discretas de Voronoi, feita pelo algoritmo EDLP, nada mais é do que associar cada ponto da imagem com algum ponto pertencente ao contorno que esteja mais próximo que qualquer um outro do mesmo. As regiões discretas de Voronoi obtidas do algoritmo EDLP fornecem a área de influência dos pontos através de simples contagem. Com isso é possível determinar quais deles possuem alta curvatura já que são caracterizados por terem área de influência maior do que a de outros pontos do contorno. Através desta área tem-se uma estimativa do ângulo de abertura:

$$\theta = \frac{2A}{r^2} \quad (4.3)$$

onde r é o raio de dilatação e A a maior área de influência entre as áreas interna e externa relativas ao *pixel* em questão. O raio r é uma orientação para limitar até onde os *pixels* serão contados. Se não houvesse essa limitação o posicionamento da forma dentro da imagem afetaria os valores de saliência. Por exemplo, uma saliência pouco significativa bem distante das bordas da imagem, seria mais importante que uma saliência significativa próxima da borda, já que inúmeros *pixels* estão sob influência do *pixel* distante da borda. Com um raio pequeno esse problema não ocorre: Torres et al. [47] sugere um raio de 10 *pixels* para que o lado oposto da forma não seja ultrapassado [47; 46; 45].

A partir do ângulo θ obtido para cada ponto, é possível estipular um ângulo padrão. Assim, qualquer ponto que possua curvatura maior que esse ângulo padrão é na verdade uma saliência. Por exemplo, pode-se determinar o ângulo padrão igual a 70° . Dessa forma, qualquer ponto que possua $\theta > 70^\circ$ é uma saliência. Uma saliência pode ser caracterizada como convexa se a área influência externa é maior que a interna ou côncava caso contrário. A figura 4.4 mostra um ponto de saliência côncavo (B) e um convexo (A) e suas respectivas áreas de influência.

Na figura 4.7(k) os pontos claros mostram as saliências utilizando esse método.

4.5 Esqueletos

Uma outra maneira de se obter as saliências é usando os esqueletos da forma, visto que toda extremidade de esqueleto interno corresponde a uma saliência convexa, e extremidades do esqueleto externo correspondem a saliências côncavas.

Inúmeras são as aplicações de esqueletos devido à sua capacidade de representar formas de maneira compacta. De maneira intuitiva pode se dizer que o esqueleto nada mais é que a linha central de uma forma ou região. Contudo, diferentes definições podem surgir segundo a maneira pela qual o esqueleto foi obtido. Linha central é baseada no entendimento de que o processo de obter o esqueleto é o afinamento sucessivo da região de interesse até que se chegue à tal linha. Outra possibilidade é definir o esqueleto em termos da Transformação do Eixo Medial (*Medial Axis Transform* - MAT) 3.6.2 [7].

Outra maneira de se obter o esqueleto é através da IFT, já que esta resulta em um mapa de rótulos L que possibilita separar os *pixels* em regiões discretas de Voronoi. O esqueleto consiste das divisões destas regiões. Sendo que tais divisões podem ser obtidas facilmente através do cálculo da imagem diferença Dif , utilizando a seguinte equação:

$$Dif(p) = \max_{\forall q \in A_4(p)} \{\min(\delta(p, q), N - \delta(p, q))\}, \quad (4.4)$$

onde L é uma imagem rotulada, $A_4(p)$ são os *pixels* na vizinhança-de-4 relativa a p , N é o número de sementes e $\delta(p, q) = L(q) - L(p)$. A diferença dentro da vizinhança-de-4 é calculada para cada *pixel* $p \in L$. Dessa equação pode-se observar que valores altos são gerados quando os valores dos *pixels* em L são muito diferentes. Caso sejam próximos valores baixos são obtidos. No primeiro caso tem-se mudanças bruscas nos níveis de cinza que deverão resultar em linhas fortes no esqueleto e no segundo caso mudanças suaves que devem produzir linhas pouco significativas no esqueleto. Ao representar a imagem diferença (figura 4.7(d)), pode-se perceber linhas mais claras próximas ao centro da forma em que se tem uma forte diferença entre os *pixels*, ou seja, trata-se de linhas significativas do esqueleto, tais linhas escurecem a medida que se aproximam das extremidades pois nesses locais a diferença entre os *pixels* não é tão acentuada como no centro da forma. Portanto, quanto maior a diferença de um *pixel* para seus vizinhos, maior o valor atribuído a esse *pixel* e maiores são as chances de que o mesmo pertença ao esqueleto.

Como pode ser observado, a imagem diferença possui inúmeros detalhes sendo que a maioria não pode ser vista através do sistema visual humano. Estes são os detalhes menos relevantes que geram linhas muito escuras e imperceptíveis e muitas vezes não devem estar presentes no esqueleto, na figura 4.7(e) tais detalhes extraídos da figura 4.7(d) são exibidos. Assim, é necessário filtrar a imagem diferença para obter um esqueleto que represente a forma convenientemente, para tal, aplica-se um *threshold*, ou seja, um limiar que separa os pontos com tons mais claros (os mais relevantes) dos mais escuros (os menos relevantes).

Nas figuras 4.7(e)-(i) vários diferentes limiares foram aplicados. Assim, o es-

queleto é gerado em diferentes escalas.

Segundo Plotze et al. ainda não se tem uma boa definição do esqueleto ideal [37], mas mesmo assim é possível verificar que os esqueletos gerados com mais detalhes não preservam somente as principais características da imagem, por isso, é preciso determinar um valor adequado para o *threshold*. Uma estimativa para tal valor, é encontrada por Torres et al. [47], que sugere 5% de N , sendo que N é o número de sementes e por consequência o número de regiões discretas de Voronoi. Esta estratégia se fundamenta no fato de que a máxima diferença que pode ser obtida é $N/2$ e que, no entanto, não se pode obter diferenças tão grandes já que uma vizinhança de quatro está sendo analisada e os *pixels* vizinhos tendem a ter valores próximos devido a rotulação incremental. Assim qualquer ponto que possua uma diferença superior a 5% de N deve ser preservado, pois é uma linha significativa.

Um fato importante que se observa é que a medida que este valor de *threshold* é modificado, ainda que minimamente, as taxas de precisão e revocação são profundamente afetadas. Desta forma um balanceamento entre o nível de detalhes e a preservação das propriedades de invariância com relação a mudanças geométricas é conveniente. Valores menores produzem mais detalhes e por consequência mais saliências e acabam por resultar em baixa precisão pois com o excesso de saliências casamentos incorretos entre as mesmas ocorrem com maior facilidade elevando imagens não relevantes no *ranking*. Valores grandes produzem um número menor de saliências, e isto torna os casamentos mais restritos, pois apenas as características essenciais da forma estão sendo julgadas, resultando em precisão mais elevada. Por outro lado, o valor do *threshold* não pode ser grande a ponto de fazer desaparecer por completo o esqueleto. Experimentalmente é possível constatar que um valor satisfatório para o *threshold* é de fato 5% de N . A figura 4.7(i) está nesse percentual. Observa-se que o esqueleto gerado preserva apenas ramos relativos à

saliências significativas.

A figura 4.7(j) mostra como as saliências da imagem original se relacionam com o esqueleto. Na figura 4.7(l) os pontos claros mostram as saliências obtidas dessa forma. Observa-se, portanto que associar as extremidades do esqueleto aos pontos do contorno é uma maneira bem mais eficiente para se obter as saliências, já que o uso puro e simples da área de influência como referencial (figura 4.7(k)), mostrou que saliências pouco significativas, decorrentes simplesmente da limitação da imagem digital, recebem mais destaques do que saliências de fato importantes.

4.6 Geração do conjunto de características

A figura 4.5 mostra como saliências podem ser obtidas através da relação entre o esqueleto e o contorno. Os valores contidos na imagem diferença obtida pela equação 4.4 podem ser interpretados como sendo o comprimento máximo do segmento mais curto entre duas raízes eqüidistantes [47], segmento $\overline{dab}$. Tal fato ocorre em virtude dos rótulos incrementais. Com isso é necessário apenas fazer a relação entre a extremidade do esqueleto e o meio deste segmento, já que necessariamente esta é saliência. Desta maneira, partindo da extremidade de esqueleto c deve-se encontrar a raiz deste *pixel* através do mapa raiz R para em seguida percorrer a distância $\overline{dab}/2$ no sentido horário (ou anti-horário). Feito isto, a posição da saliência torna-se conhecida.

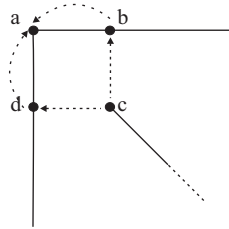


Figura 4.5: Relação entre o esqueleto e a saliência

Obtida a posição da saliência, é necessário atribuir um valor para a mesma. Aqui novamente a IFT é de grande importância. Através do mapa de rótulos L gerado pela IFT é possível obter o tamanho da área de influência de cada ponto do contorno através da simples construção de um histograma. Afinal, a área de influência é apenas a contagem dos *pixels* que estão em certa região discreta limitada por um raio. O valor da saliência é tomado como sendo a área de influência do ponto.

Com isso, a construção do conjunto de características está quase completa: já se sabe a posição bem como o valor da saliência. Falta apenas determinar se a saliência é côncava ou convexa. No primeiro caso o valor da saliência torna-se negativo e no segundo permanece positivo. Determinar o tipo da saliência é trivial: se a extremidade é de esqueleto interno, a saliência é convexa, caso contrário é côncava.

Para tornar o conjunto de características mais eficiente a posição a ser utilizada deve ser a posição relativa da saliência com relação ao contorno: escolhe-se um ponto arbitrário como sendo o ponto inicial, com isso cada ponto tem uma posição, em seguida divide-se o valor da posição pelo total de pontos no contorno, valor da posição/ N . Essa última medida visa garantir que haja normalização quando mudanças de escala são realizadas.

O conjunto de características portanto não vem a ser um vetor mas sim dois vetores associados, onde o primeiro contém o valor das saliências e o segundo suas respectivas posições relativas. Na figura 4.6 uma imagem teve seu conjunto de características gerados e exibidos em um gráfico saliências x posição.

4.6.1 O Algoritmo de Casamento

Cada forma é representada por um conjunto de características constituído pelo valor da saliência juntamente com sua posição relativa: o valor da saliência é dado

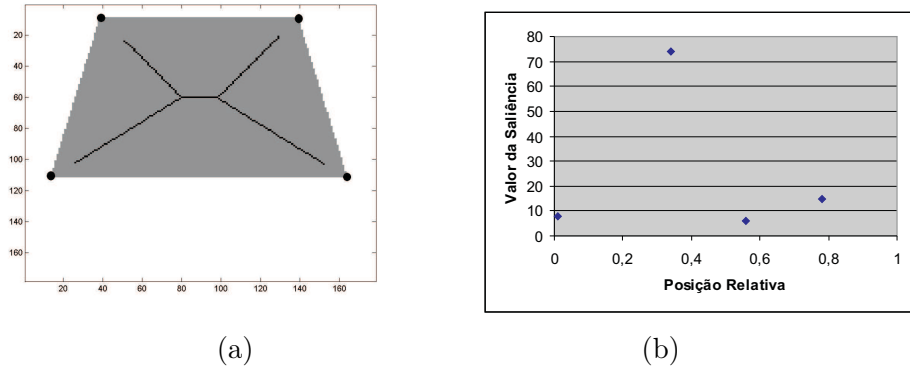


Figura 4.6: Conjunto de características. (a) Imagem com saliências associadas ao esqueleto. (b) Gráfico mostrando saliências x posição.

através da área de influência e a posição relativa pela posição do *pixel* em relação ao contorno. Uma forma em posições distintas dentro da imagem possui sempre o mesmo conjunto de características, exceto pela posição relativa que não será necessariamente a mesma, já que ela depende do ponto inicial e este pode ser qualquer um. Ainda assim trata-se da mesma forma. O algoritmo de casamento deve prever essa disparidade relacionada ao ponto inicial fazendo correções à medida que realiza as comparações. O algoritmo do CSS (*Curvature Scale Space*) trabalha da maneira desejada no descritor saliências, leva em consideração a posição relativa e realiza um deslocamento toda vez que encontra saliências similares.

Considerando o casamento entre uma forma representada por F_A e uma outra representada por F_B , o primeiro passo do algoritmo é ordenar ambas em ordem decrescente de valores de saliências. Isso é necessário para que as comparações sejam facilitadas. Em seguida as saliências similares são obtidas e armazenadas em uma lista de pares, cada par contém uma saliência de F_A e uma de F_B . Duas saliências não precisam ter o mesmo valor para serem consideradas similares, basta que o módulo da diferença entre elas seja menor ou igual a 20% da maior saliência de F_A ou F_B conforme a primeira saliência usada na diferença: se a primeira

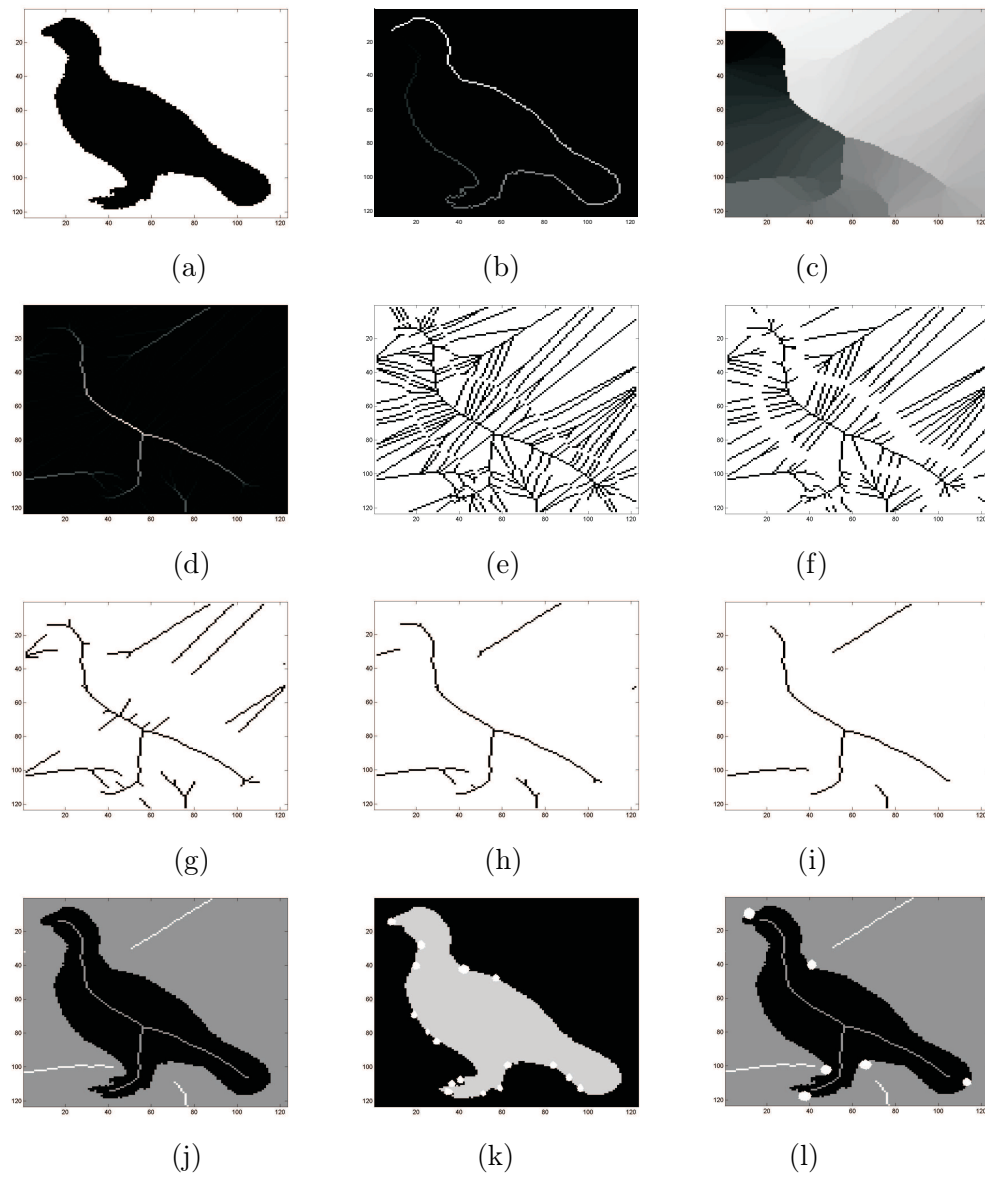


Figura 4.7: Saliências através de IFT. (a) Forma original. (b) Contorno rotulado. (c) Resultado da IFT. (d) Imagem diferença, (e)-(i) Diferentes limiares aplicados à imagem diferença. (j) Relação entre o esqueleto e a imagem original. (k) Pontos claros são saliências obtidas através da área de influência (l) Pontos claros são saliências obtidas da relação entre a extremidade do esqueleto e o contorno.

pertencer a F_A , a maior saliência deve ser relativa a F_A . Feito isto, para cada par candidato a casamento, o parâmetro α é calculado. Este parâmetro é obtido calculando-se a distância, em termos de posição relativa, entre as duas saliências que supostamente casam. Com isso, todas as saliências de F_A devem ser deslocadas por α e calcula-se a distância entre F_B e o novo F_A . Esse procedimento é feito para todos os pares gerando várias distâncias diferentes. É escolhido o menor valor obtido já que este é o casamento mais próximo.

No caso das saliências não estarem muito distantes, a distância é obtida através de métrica euclidiana envolvendo as posições relativas e o valor das saliências. Caso contrário, é feita a soma entre os valores de saliência. Resumindo: se as saliências estão próximas, o valor dado ao casamento é aproximadamente a diferença entre elas e no caso de serem distantes tem-se exatamente a soma entre elas. O fato de ser aproximadamente a diferença é devido ao uso das posições relativas, que não chegam a afetar o resultado significativamente já que estão no intervalo $[0,1]$.

O tamanho dos vetores de características também é considerado durante a comparação. A última saliência do menor vetor é o ponto de parada no cálculo da distância. No final a diferença entre o tamanho dos dois vetores é adicionada ao valor resultante.

Dados dois contornos distintos representados pelos conjunto de saliências $F_A = \{(u_{A1}, s_{A1}), \dots, (u_{An}, s_{An})\}$ e $F_B = \{(u_{B1}, s_{B1}), \dots, (u_{Bm}, s_{Bm})\}$, sendo que s_{Ai} é o valor da i -ésima saliência, e u_{Ai} a posição relativa da mesma. F_A e F_B podem possuir tamanhos diferentes, sendo que este não vem a ser um problema para o algoritmo de casamento, que é dado abaixo:

1. Crie $F'_A = \{(u'_{A1}, s'_{A1}), \dots, (u'_{An}, s'_{An})\}$ e $F'_B = \{(u'_{B1}, s'_{B1}), \dots, (u'_{Bm}, s'_{Bm})\}$ ordenando F_A e F_B por ordem decrescente dos valores das saliências
2. Crie a lista L contendo pares de pontos candidatos a casamento de F'_A e F'_B .

Um par $((u'_{Ai}, s'_{Ai}), (u'_{Bj}, s'_{Bj}))$ está em L se $|s'_{Ai} - s'_{Bj}| \leq 0.2s'_{A1}$. Um par $((u'_{Bj}, s'_{Bj}), (u'_{Ai}, s'_{Ai}))$ está em L se $|s'_{Bj} - s'_{Ai}| \leq 0.2s'_{B1}$.

3. Para cada par de candidato a casamento na forma $P_{ij} = \{(u'_{Ai}, s'_{Ai}), (u'_{Bj}, s'_{Bj})\}$ em L , encontre o parâmetro de deslocamento α tal que $\alpha = u'_{Ai} - u'_{Bj}$. Desloque os pontos de saliência de F_A por α , obtendo $F''_A = \{(u''_{A1}, s''_{A1}), \dots, (u''_{An}, s''_{An})\}$.
4. A distância d entre F''_A e F_B é dada por

$$d = \sum_{k=1}^{\min\{n,m\}} d_k$$

Onde

$$d_k = \begin{cases} \sqrt{(u''_{Ak} - u_{Bk})^2 + (s''_{Ak} - s_{Bk})^2}, & \text{se } |u''_{Ak} - u_{Bk}| \leq 0.2, \\ s''_{Ak} + s_{Bk}, & \text{caso contrário} \end{cases}$$

Finalmente, se $n \neq m$, a quantidade s de pontos não casados é adicionada a d .

5. Repita os passos 3 e 4 considerando os outros pares candidatos em L .
6. Selecione a menor distância d como sendo a distância entre F_A e F_B

4.7 Considerações finais

Mais de uma maneira pode ser usada para obter o descritor saliência do contorno, sendo que para a maioria das abordagens a IFT é utilizada. Isso acontece porque para determinar se um *pixel* é de saliência ou não é preciso conhecer sua área de influência e a IFT é um método bastante robusto que é capaz de fornecer as regiões discretas de Voronoi. Por ser uma transformada baseada em grafo é

possível calcular o custo do caminho e determinar a qual árvore um *pixel* pertence. Portanto, todas as árvores que formam a floresta cobrindo a imagem são obtidas.

Um fator importante para o funcionamento da IFT é a rotulação. Os *pixels* devem ser rotulados para que a IFT possa distinguir entre as diversas raízes. Cada *pixel* do contorno é uma raiz com um rótulo distinto que dará origem a uma árvore. Assim, um algoritmo de rotulação é aplicado para que a IFT propague os rótulos para dentro e para fora do contorno obtendo o mapa de rótulos.

O mapa de rótulos é fundamental na obtenção dos descritores de saliência, seja para o caso das saliências obtidas pela área de influência, seja para o caso das saliências obtidas da relação entre o esqueleto e o contorno. No primeiro caso um histograma é construído e a área de influência determinada, no segundo caso, o esqueleto é extraído através da imagem diferença originada do mapa de rótulos, e da aplicação de um limiar em tal imagem que tem como finalidade extrair apenas as linhas significativas do esqueleto. A partir daí a relação entre o esqueleto e o contorno é estabelecida. Pode-se constatar que a obtenção das saliências através do esqueleto é bem mais eficiente, pois nesse caso, não se tem a influência de deformidades presentes na imagem, ao contrário do que se verificou em saliências através da área de influência.

Por último o que se pode destacar é o algoritmo de casamento usado para comparar saliências. Torres. et al. utilizam o algoritmo do CSS, baseado em pontos de inflexões que devem ser comparados e que, portanto, não difere muito de saliências. Assim, é conveniente testar a construção do descritor saliência e principalmente como ele se comporta em um ambiente de recuperação de imagens, já que essa é uma de suas finalidade. Os testes realizados são abordados no capítulo a seguir.

Capítulo 5

A Implementação

5.1 CBIR

Foi implementado um CBIR baseado em forma utilizando as saliências, onde módulos possuem funções específicas, como pode ser observado na figura 5.1. A parte delimitada pela linha tracejada é feita estaticamente, já que extrair as características de toda uma base de dados demanda tempo e não precisa ser feita mais do que uma vez. Antes de começar a realizar a recuperação de imagens é preciso processar toda a base de dados armazenando os vetores de características de forma a gerar uma nova base de dados, na qual o índice para a consulta é o próprio conjunto de características.

A parte delimitada por uma linha sólida na figura 5.1 é feita dinamicamente: submetida a consulta, suas características são extraídas, e um módulo de busca varre toda a base de dados de características aplicando o algoritmo de casamento que retorna um valor de dissimilaridade. Em seguida esse mesmo módulo constrói um *ranking* com base no valor d de dissimilaridade obtido.

A seguir os principais módulos são detalhados.

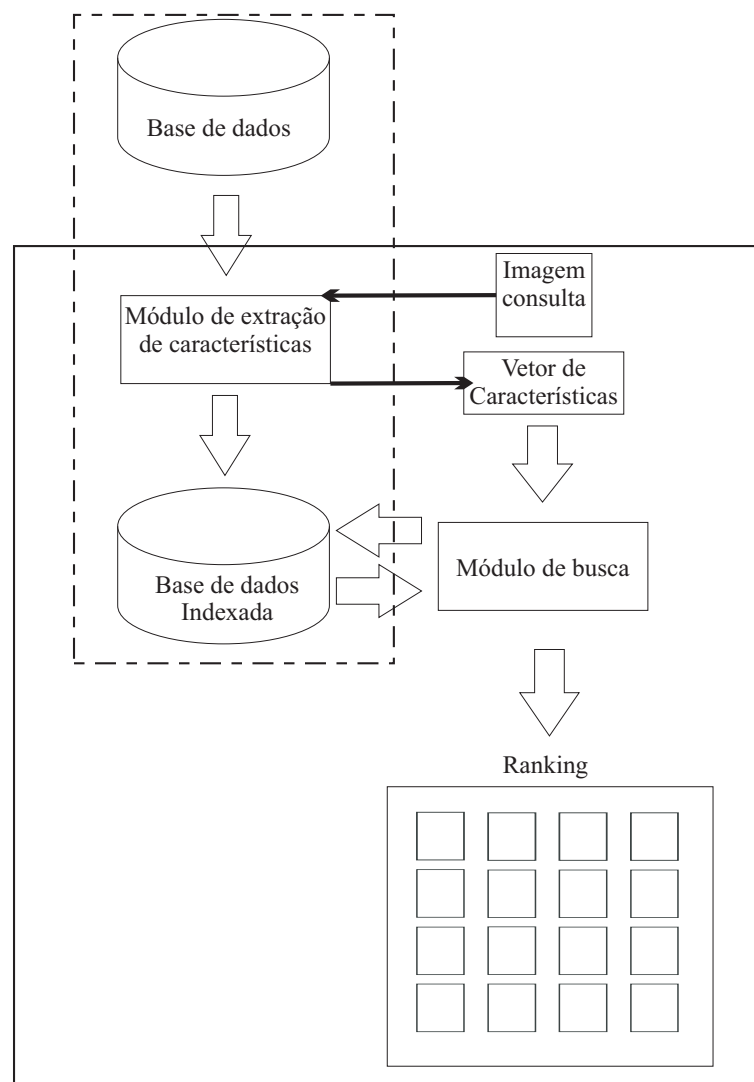


Figura 5.1: Esquema seguido na recuperação

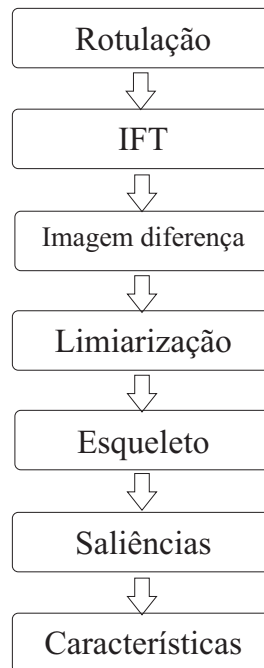


Figura 5.2: Extração de características

5.1.1 Módulo de Extração de Características

O módulo de extração de características trabalha efetuando cada uma das etapas mostradas na figura 5.2: inicialmente é feita a rotulação dos *pixels* semente através do algoritmo 2 encontrado na seção 4.3. Em seguida a transformada imagem-floresta gera um mapa de rótulos L a partir das sementes rotuladas. O próximo passo é aplicar a equação 4.4 no mapa de rótulos de forma a obter a imagem diferença. Um limiar adequado deve ser usado na imagem diferença para extrair o esqueleto, e este deve ser usado em um relacionamento com o contorno para que finalmente as saliências possam ser definidas. Por último, gera-se o vetor de características contendo a posição relativa e o valor da saliência.

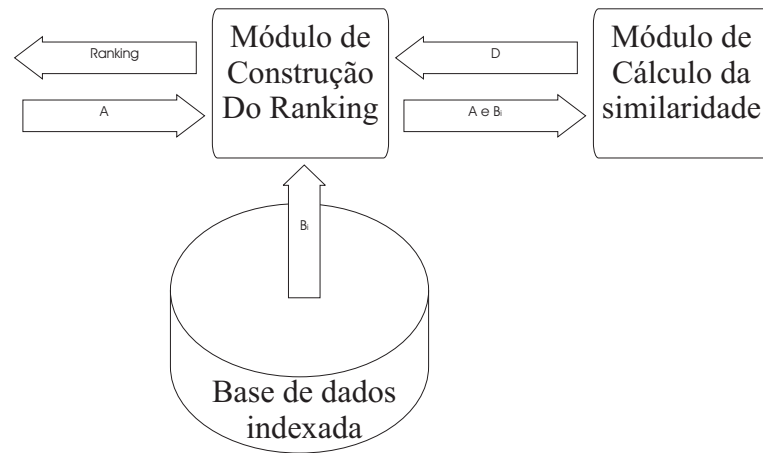


Figura 5.3: Módulo de busca

5.1.2 Módulo de Busca

A figura 5.3 ilustra o funcionamento do módulo de busca: o vetor de características A , extraído da imagem consulta, é passado ao módulo de construção do *ranking*. Este, por sua vez, obtém uma imagem processada na base de dados de características B_i , passando-a juntamente com o vetor A para o módulo que calcula a medida de similaridade, o valor d resultante do cálculo feito pelo algoritmo de casamento é retornado ao módulo de construção do *ranking* que o armazena em uma lista ordenada, ao final essa lista é exatamente o *ranking* a ser retornado ao módulo de recuperação. Sendo $i = 1, \dots, N$ e N a quantidade de imagens na base de dados.

5.2 Ferramentas e linguagens

A principal linguagem de programação utilizada na implementação foi a linguagem C, sendo que o Visual C++ 6.0 foi a plataforma escolhida. Todas as rotinas do módulo de extração de características e do módulo de busca foram implementadas nesta linguagem.

Devido a sua grande facilidade na manipulação de matrizes e imagens, o Matlab foi usado na *interface* do sistema. Um programa escrito para esta plataforma ficou responsável por toda a troca de informações com o usuário. Assim, as consultas são submetidas através do Matlab: o usuário busca a imagem consulta que é manipuladas e gravadas em arquivo pelo Matlab, em seguida o comando é passado ao programa em **C** que realiza todas as principais tarefas do CBIR. Ao término da execução da consulta o programa em **C** grava o *ranking* em um arquivo retornando o comando ao Matlab que exibe os resultados.

5.3 Interface

Como foi dito anteriormente a interface foi feita em Matlab, um ambiente bem simples e fácil de se usar foi desenvolvido. A entrada de dados consiste apenas da submissão da imagem consulta, sendo que esta pode ser qualquer imagem em dois tons de cinza do usuário, nos formatos .jpeg, .jpg e .bmp. É importante frisar que a imagem consulta deve ser a forma de um objeto, caso o usuário queira fazer a recuperação de uma imagem que não contenha uma forma isolada deverá segmentar a imagem. A figura 5.4 mostra como a imagem consulta é adquirida.

Escolhida a imagem para a qual se fará a consulta, a mesma é exibida no espaço à esquerda e o usuário deverá clicar no botão “consultar” abaixo da imagem para efetuar a busca no banco de dados. Quando a busca terminar, o usuário pode visualizar todas as imagens retornadas em grupos de 36, no qual as imagens estão ordenadas por relevância. A figura 5.5 mostra a exibição do resultado da consulta.

Um gráfico é disposto para que as características das imagens possam ser comparadas pelo usuário. Feita a consulta um gráfico saliência x posição é exibido mostrando as características da imagem consulta, figura 5.6. Em um *combo box* uma das 36 imagens exibidas pode ser escolhida para que suas características sejam

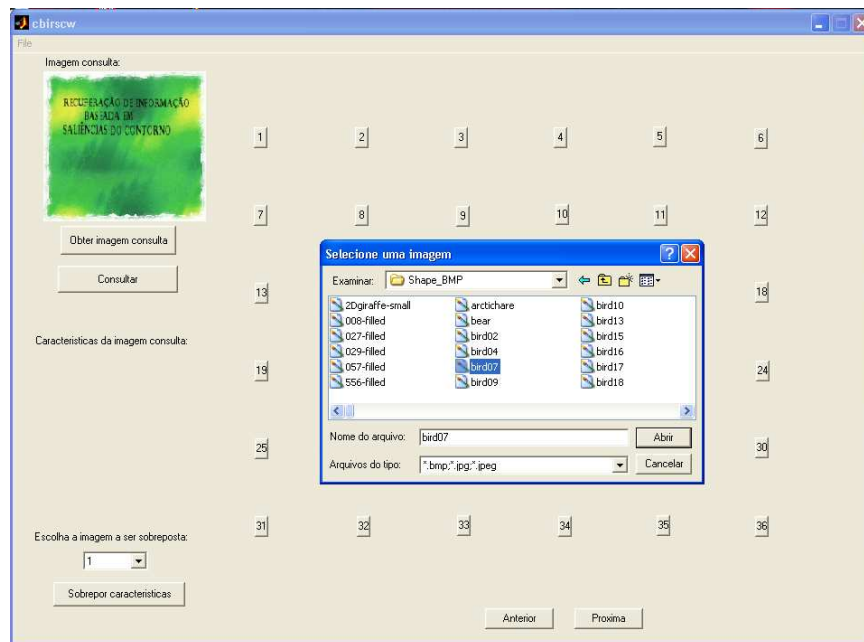


Figura 5.4: Interface do programa mostrando como adquirir imagem consulta

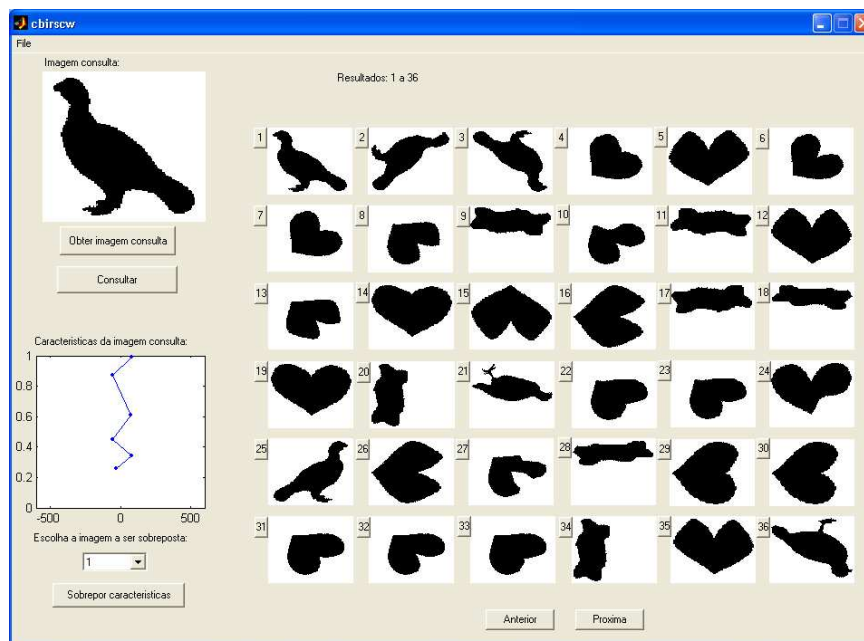


Figura 5.5: Interface do programa mostrando resultado da consulta

sobrepostas às da imagem consulta, figura 5.7. Os asteriscos no gráfico da figura 5.8 são as características da primeira imagem do *ranking* sendo sobrepostas.

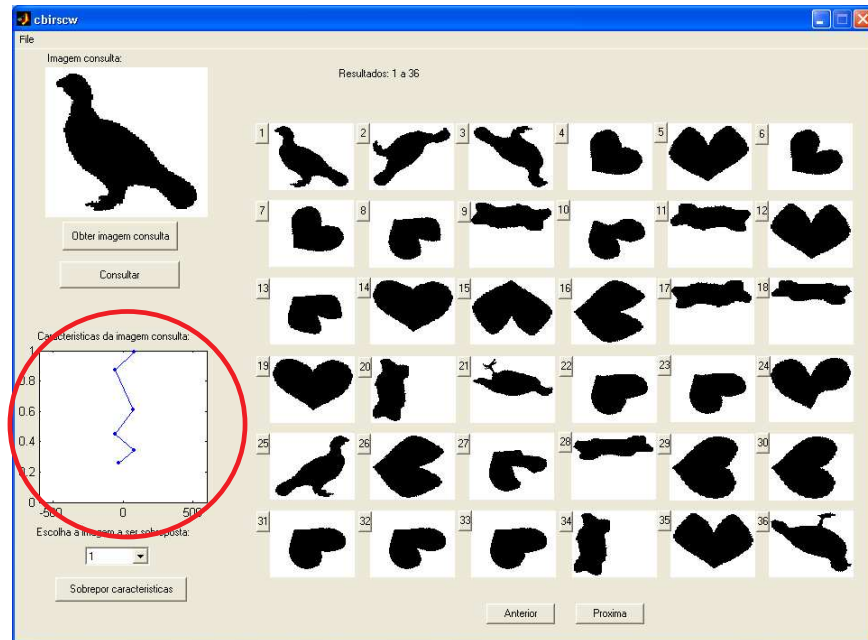


Figura 5.6: Gráfico de características da imagem consulta

Outra ferramenta que o usuário pode usar na análise dos dados é a visualização das saliências de cada imagem no *ranking*, clicando no número à esquerda da imagem exibida, uma janela se abre e os pontos de saliência aparecem marcados na imagem. Na figura 5.9 os pontos de saliência são exibidos, eles foram ressaltados para que possam ser melhor visualizados.

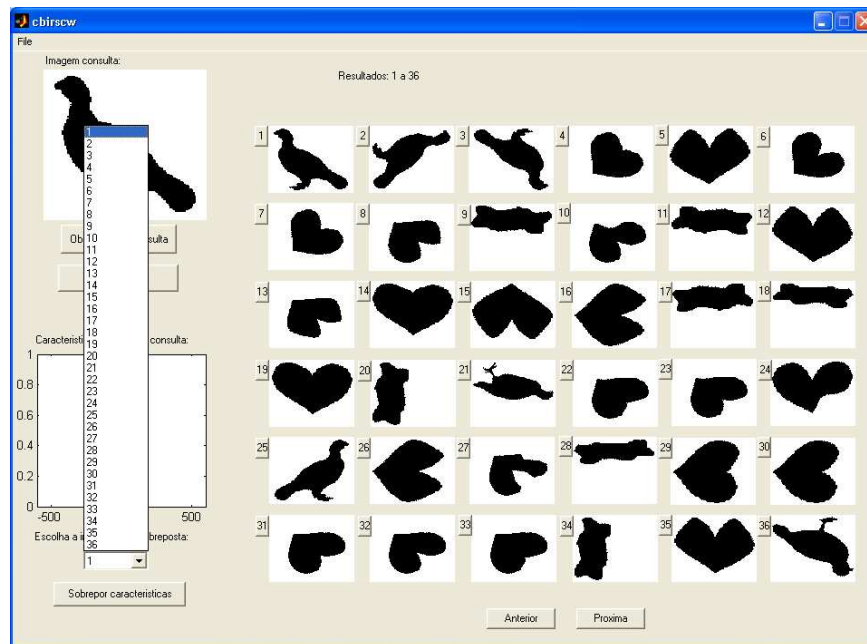


Figura 5.7: Escolha da imagem para sobrepor características.

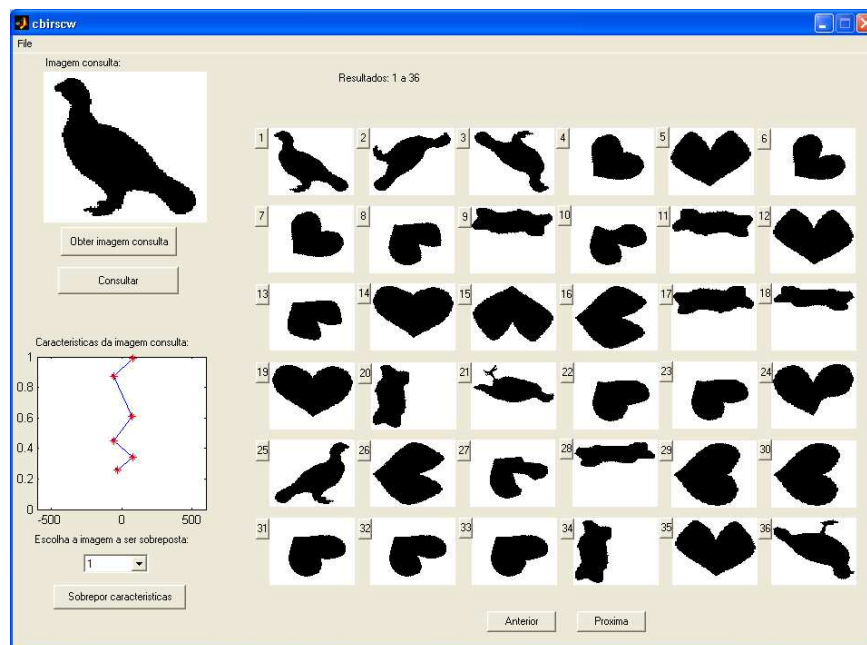


Figura 5.8: Características sobrepostas: imagem consulta e imagem no ranking

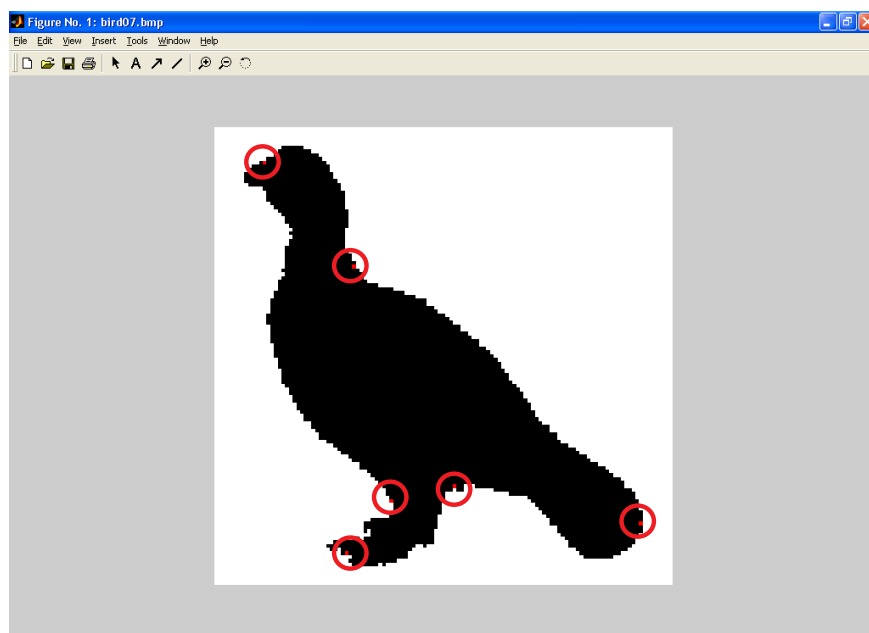


Figura 5.9: Saliências de imagem no ranking

Capítulo 6

Resultados

6.1 Introdução

Foi realizado um experimento sobre uma base de dados contendo formas armazenadas sob o formato .bmp. 83 formas distintas podem ser observadas e cada uma possui 9 variações quanto à escala e rotação. Algumas, inclusive, combinam essas duas transformações, totalizando 747 imagens binárias de tamanho variável (imagens de 113x46 a 429x429). A imagem da forma juntamente com suas variações é considerada uma classe e o conjunto de imagens relevantes à uma determinada consulta é toda a classe à qual essa imagem consulta pertence, assim como imagens de classes que sejam similares (caso existam classes nessas condições).

Neste capítulo os resultados dos experimentos são discutidos. Gráficos Precisão x Revocação são analisados na seção 6.3 e na seção 6.4 considerações a respeito dos resultados são feitas, uma conclusão a respeito dos resultados abordando as principais questões finaliza o capítulo.

6.2 As Consultas

Nas consultas uma imagem exemplo é dada (assim como foi mostrado no capítulo 5). Suas características são extraídas dinamicamente, em seguida a busca é efetuada e um *ranking* é construído. Neste *ranking* um valor de dissimilaridade é atribuído a cada imagem da base de dados frente a imagem consulta utilizando o algoritmo de casamento da seção 4.6.1.

Várias consultas foram realizadas utilizando imagens exemplo da própria base de dados a fim de verificar o desempenho do método. Os *rankings* resultantes tiveram grande variabilidade, alguns exibiram grande quantidade de imagens relevantes nas posições iniciais, enquanto outros nem tanto.

Na próxima seção é apresentada a análise dos gráficos precisão x revocação relativos às consultas realizadas.

6.3 Análise dos Gráficos Precisão x Revocação

Imagem consulta: óculos

Na figura 6.1 observa-se que a precisão se mantém bastante elevada para baixos níveis de revocação, ao atingir aproximadamente 13% de revocação a precisão sofre um forte declínio chegando a níveis bem baixos a partir desse ponto. Observou-se nesse caso uma forte incidência de casamentos parciais: a extremidade nos lados da forma do óculos casou com pés e/ou bicos de pássaros. Vários tijolos também ficaram bem posicionados no *ranking*, evidenciando que também ocorreu casamentos parciais, como ilustra a figura 6.11.

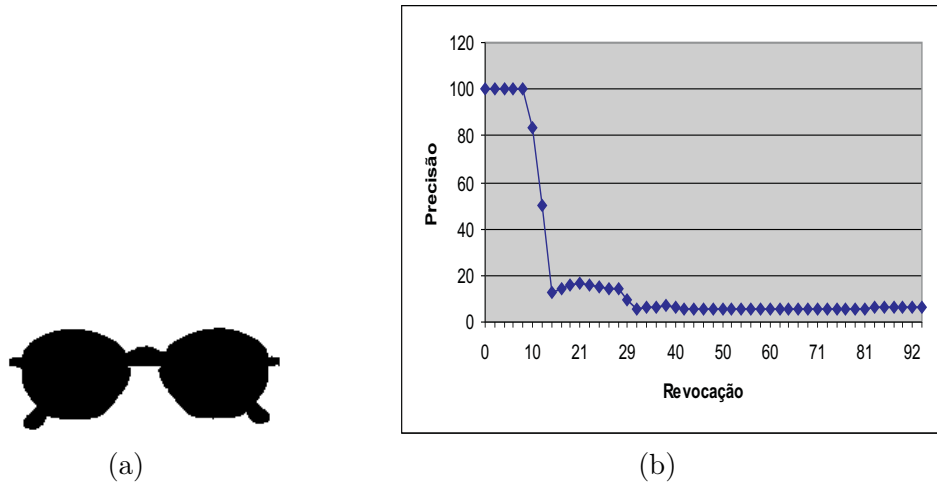


Figura 6.1: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

Imagem consulta: taça

Na figura 6.2 a precisão é máxima até aproximadamente 19% de revocação, caindo para patamares entre 40% à 59%. Ocorre uma nova queda na qual a precisão se mantém em cerca de 20% até aproximadamente 80% de revocação, para então sofrer mais uma queda. O que se percebeu desta consulta foi que os valores repetidos de dissimilaridade ocorridos como resultado do algoritmo de casamento influenciaram o gráfico P x R. Os ciclos ocorridos onde se observa uma precisão por um período para logo depois haver uma queda significa que várias imagens não relevantes receberam o mesmo valor de dissimilaridade que um conjunto de imagens relevantes. Em geral quando isso acontece, a ordem das imagens no *ranking* é alterada. As imagens relevantes tentam a aparecer aleatoriamente ao longo da listagem de imagens o que não contribui para altos valores de precisão. Além disso casamentos parciais ocorreram com formas de ossos, já que estes apareceram em grande quantidade entre os documentos relevantes, como ilustra a figura 6.12.

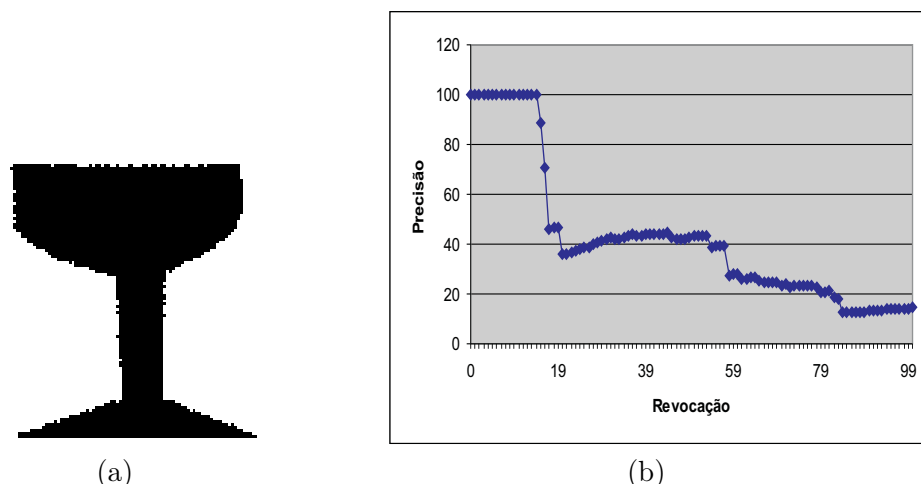


Figura 6.2: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

Imagem consulta: coração

Na figura 6.3 a precisão foi bastante satisfatória ao longo da tarefa de recuperação, se mantendo aproximadamente ótima até revocação de 40% e com 80% de revocação a precisão se encontrava em 80%. Somente após este percentual houve queda significativa. Portanto, nesta consulta o sistema se comportou muito bem, atingindo altos níveis de revocação com precisão proximamente ótima. Neste caso nenhum tipo de casamento parcial significativo pode ser notado (figura 6.13).

Imagem consulta: pássaro07

Na figura 6.4 o desempenho do sistema foi bastante ruim, o nível de precisão caiu drasticamente após revocação de 5% e se manteve baixo, em torno de 20% até atingir revocação de 100%. Vários casamentos parciais ocorreram, o peito do pássaro com o lado do coração, pés/bicos do pássaro com a extremidade do óculos, houve também casamentos parciais com tijolos (figura 6.14).

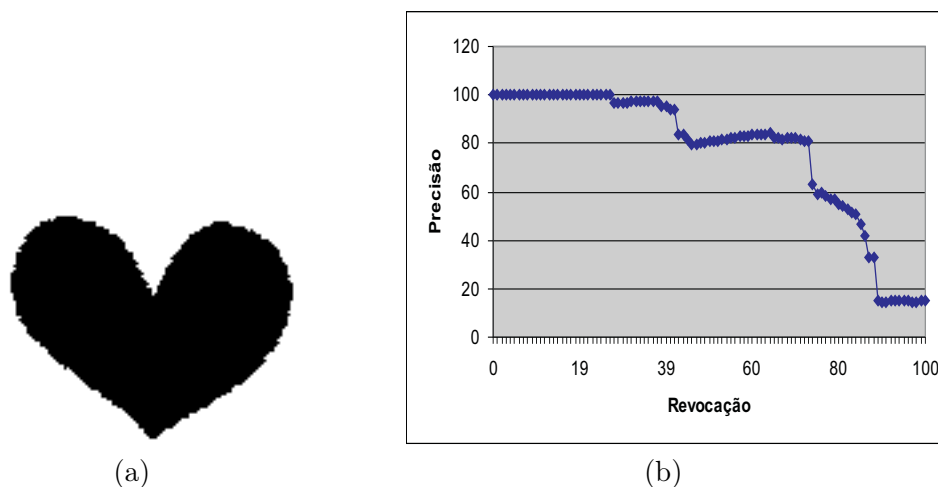


Figura 6.3: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

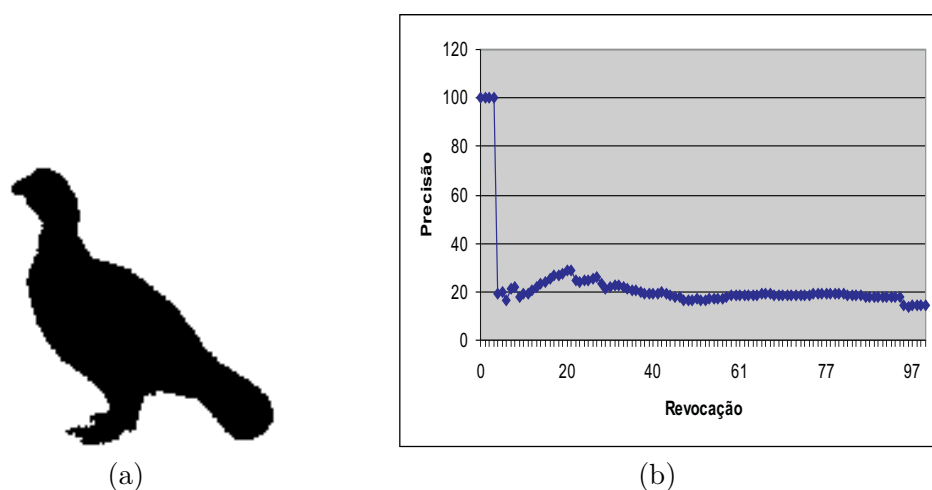


Figura 6.4: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

Imagem consulta: pássaro02

Na figura 6.5 o sistema também não se comportou adequadamente, a precisão começou a sofrer uma queda gradual a partir de 5% de revocação chegando em 20% em aproximadamente 35% de revocação e se mantendo constante até atingir 100% de revocação. Observou-se também as mesmas ocorrências de casamentos

parciais do caso anterior (figura 6.15).

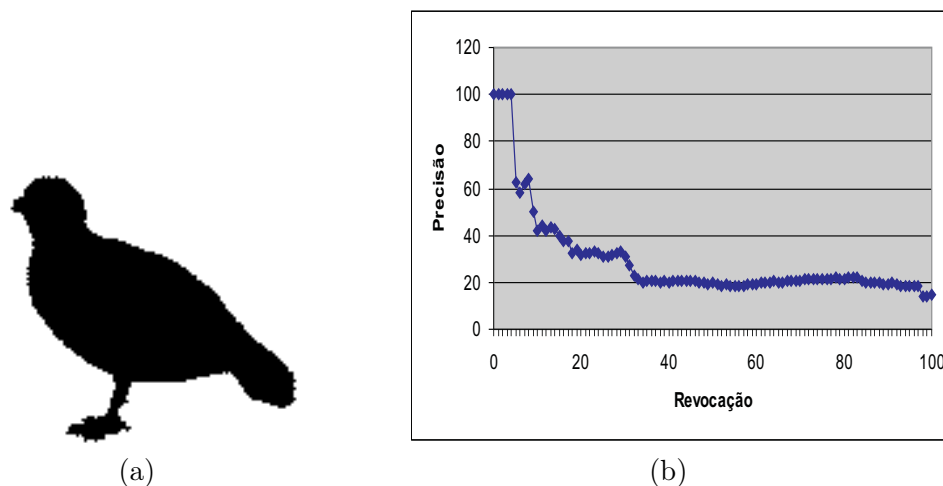


Figura 6.5: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

Imagem consulta: coelho

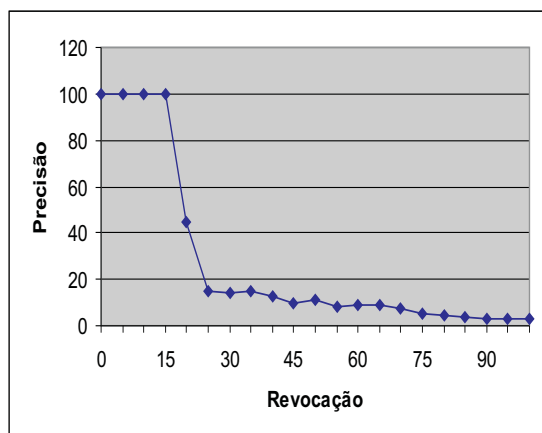
Na figura 6.6 a precisão se manteve ótima até cerca de 20% de revocação quando houve uma queda levando a precisão para baixo de 20% até a revocação de 100%. O problema maior nesta consulta foi também devido aos valores repetidos da dissimilaridade. Quase todas as imagens relevantes receberam um mesmo valor. Como outras não relevantes também receberam esse valor, e a ordem no *ranking* é aleatória, não se pôde constatar a efetividade do sistema. Casamentos parciais ocorreram principalmente entre as orelhas do coelho e ossos e também entre os pés do coelho e do pássaro, figura 6.16.

Imagem consulta: osso

Na figura 6.7 a precisão se mantém ótima apenas para revocação até 10%, após esta taxa a precisão sofre um declínio onde permanece entre 40 e 20% até atingir



(a)



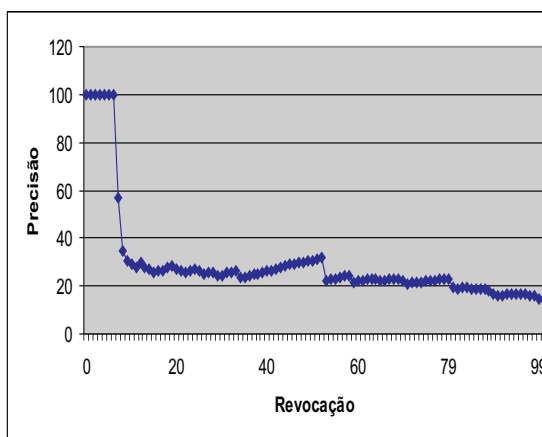
(b)

Figura 6.6: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

revocação 100%. Casamentos parciais ocorreram principalmente com tijolos, já que a extremidade dos mesmos se assemelha à extremidade dos ossos (figura 6.17).



(a)



(b)

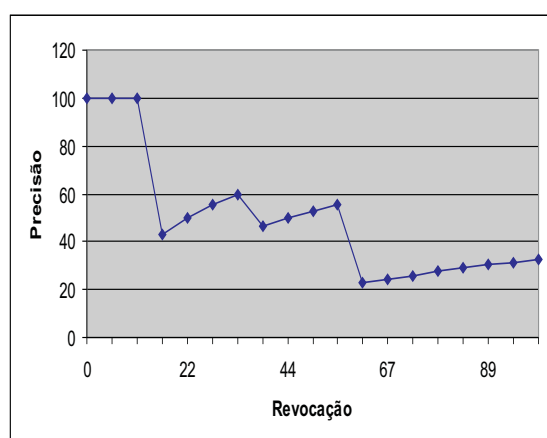
Figura 6.7: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

Imagem consulta: peixe

Na figura 6.8 a curva do gráfico Precisão x Revocação é bastante curiosa. Até 11% apresenta precisão 100%, caindo para 40% mas subindo à 60% quando a revocação está em 33%. Isso ocorre porque uma grande quantidade de documentos relevantes foram recuperados neste momento possibilitando aumento na precisão. Novamente os valores repetidos de dissimilaridade são os responsáveis, já que os documentos não relevantes recuperados que fizeram a precisão cair possuem o mesmo valor de dissimilaridade que os documentos relevantes recuperados. Esse efeito é verificado mais uma vez neste gráfico, quando a revocação está em aproximadamente 38%. Casamentos parciais ocorreram principalmente com ossos contento extremidades pontiagudas, já que estas se assemelham as nadadeiras do peixe (figura 6.18).



(a)



(b)

Figura 6.8: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

Imagem consulta: máscara

Na figura 6.9 a precisão não ultrapassa níveis médios durante a tarefa de recuperação: atinge picos de 50% e se mantém entre 20% e 40% até atingir a revocação

100%. O excesso de deformidades dessa forma permitiu vários casamentos parciais com todo tipo de imagem (figura 6.19).

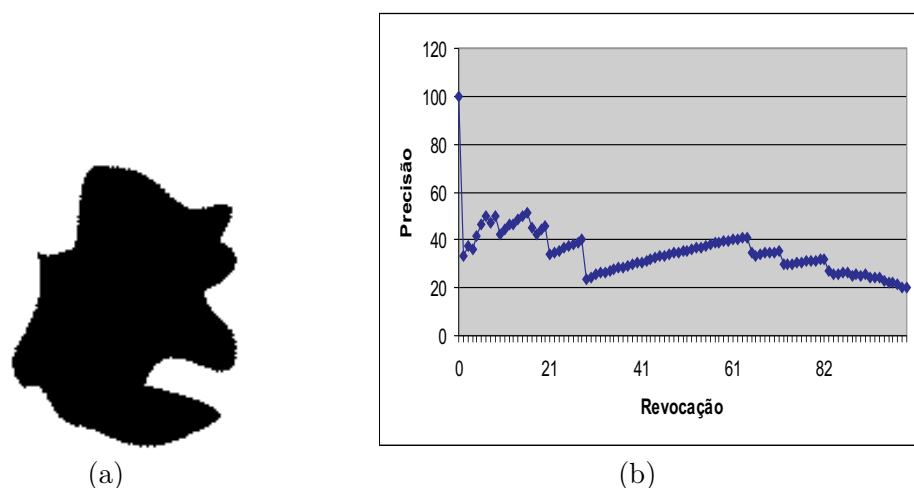


Figura 6.9: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

Imagem consulta: tijolo

Na figura 6.10 a precisão se mantém ótima até aproximadamente 10%, em seguida cai para cerca de 20% onde se mantém constante até atingir revocação 100%. Neste caso o valor de dissimilaridade repetido também prejudica o resultado. Casamentos parciais ocorreram principalmente com ossos e taças (figura 6.20).

6.4 Discussão

6.4.1 Normalizações no Algoritmo de Casamento

Normalizações foram inseridas no algoritmo a fim de melhorar os resultados das consultas. Essas normalizações foram inseridas exatamente no cálculo da distância em que se verificou uma certa discrepância entre os valores gerados e os valores

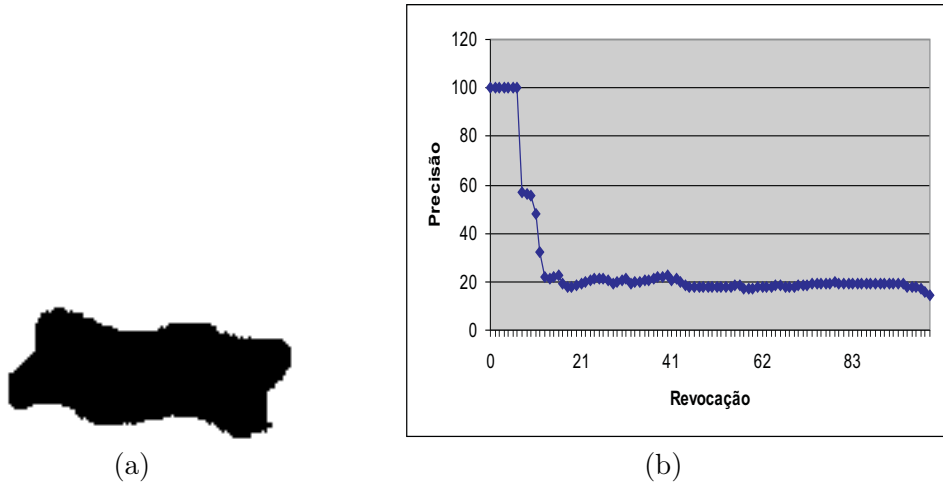


Figura 6.10: Avaliação de consulta (a)Imagem Consulta (b)Gráfico P x R

esperados. Por exemplo, em algumas situações em que os vetores de característica eram bastante similares, valores muito altos de dissimilaridade foram gerados quando na verdade se esperava que fossem menores.

A distância é interpretada como sendo a medida de dissimilaridade, portanto, quanto menor o valor obtido, menor é a diferença entre duas formas que estão sendo comparadas. Considerando o vetor de características, o cálculo da distância é feito posição a posição e se duas saliências estão próximas, a distância é aproximadamente a subtração entre elas e caso contrário a distância é exatamente a soma. Ao final toma-se o somatório das distâncias calculadas em cada posição do vetor como sendo o valor total da distância. A seção 4.6.1 apresenta uma explicação detalhada deste cálculo.

A seguir as normalizações são expostas na sequência em que são realizadas:

1. Conta-se o número de saliências que estejam próximas nos dois vetores sendo comparados obtendo $d1$ - posteriormente este valor será utilizado;
2. Eleva-se o número de saliências que estejam próximas, $d1$, a décima potência

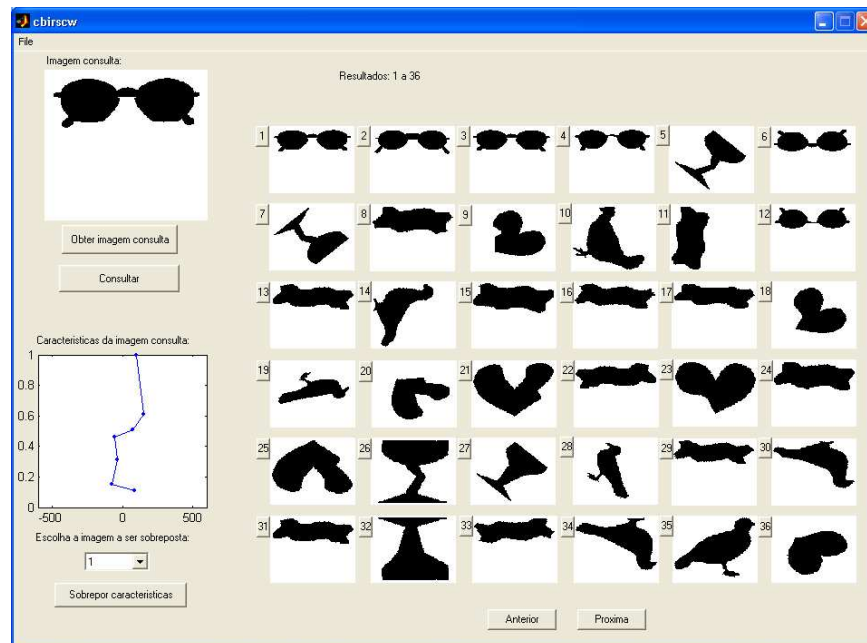


Figura 6.11: Ranking para imagem consulta óculos

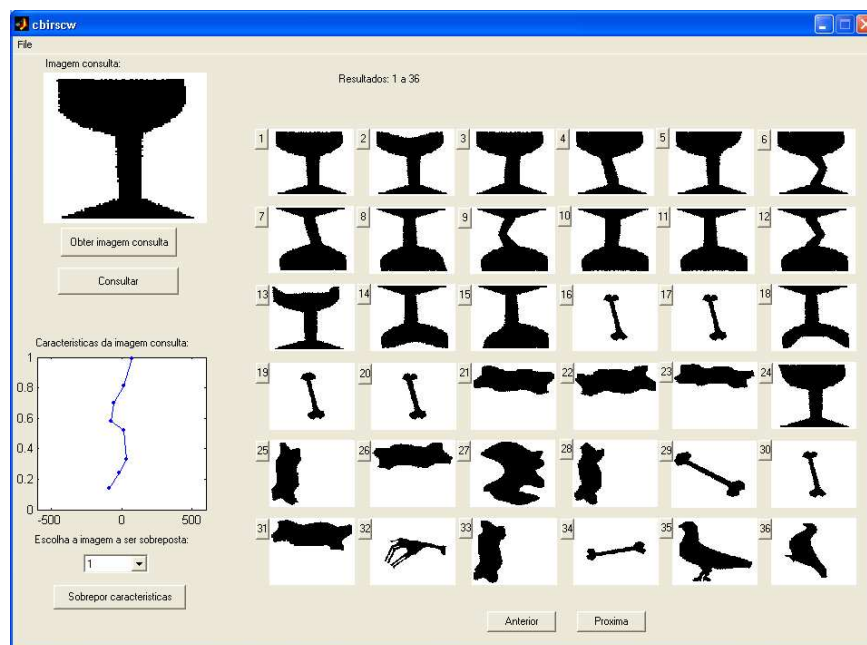


Figura 6.12: Ranking para imagem consulta taça

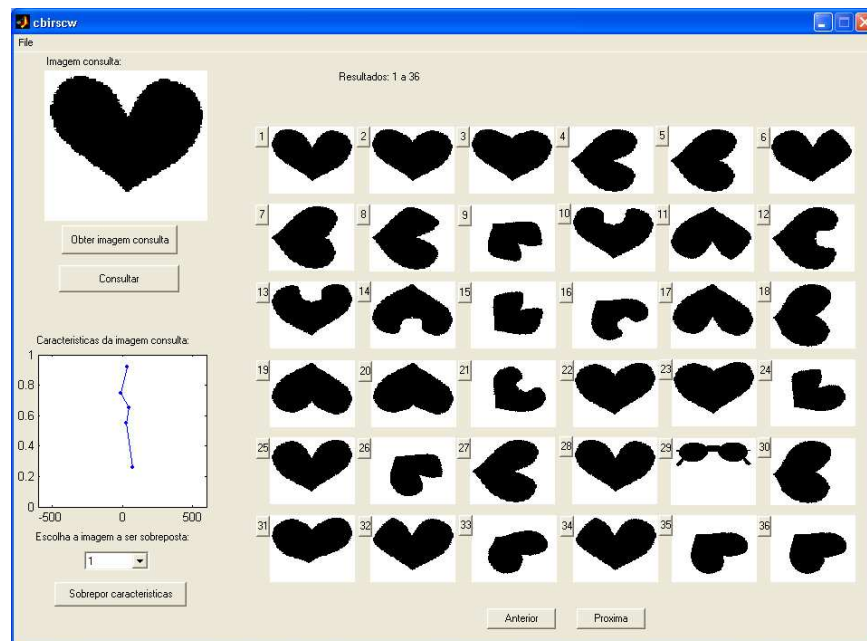


Figura 6.13: Ranking para imagem consulta coração

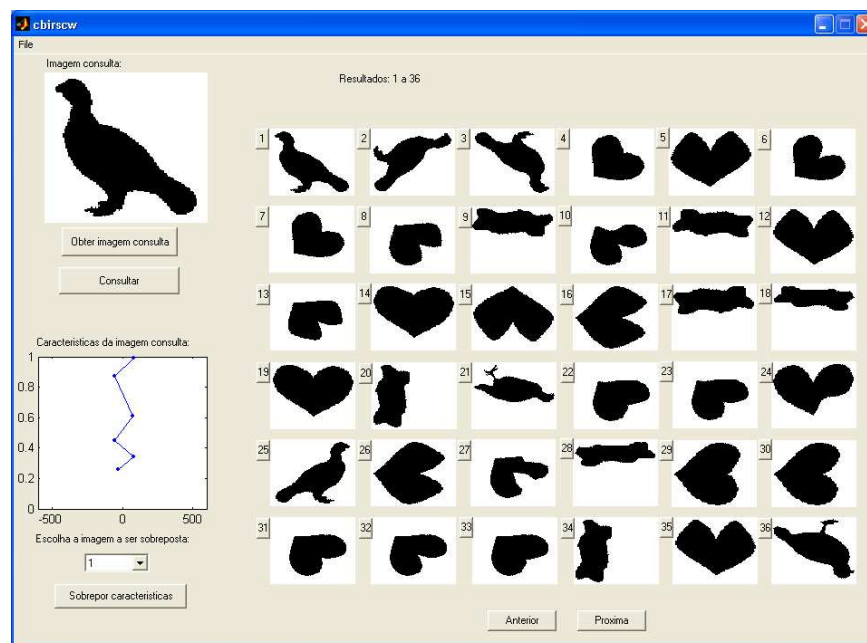


Figura 6.14: Ranking para imagem consulta pássaro07

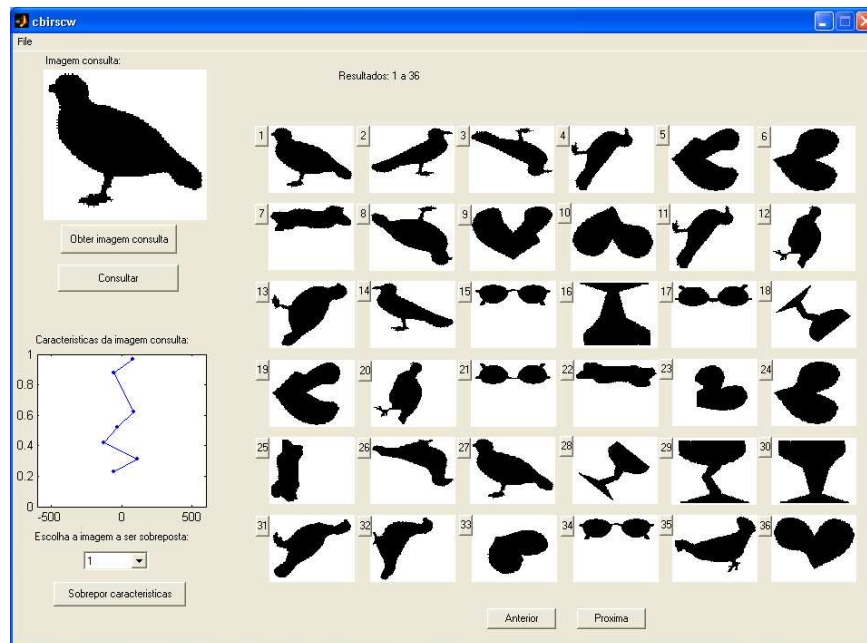


Figura 6.15: Ranking para imagem consulta pássaro02

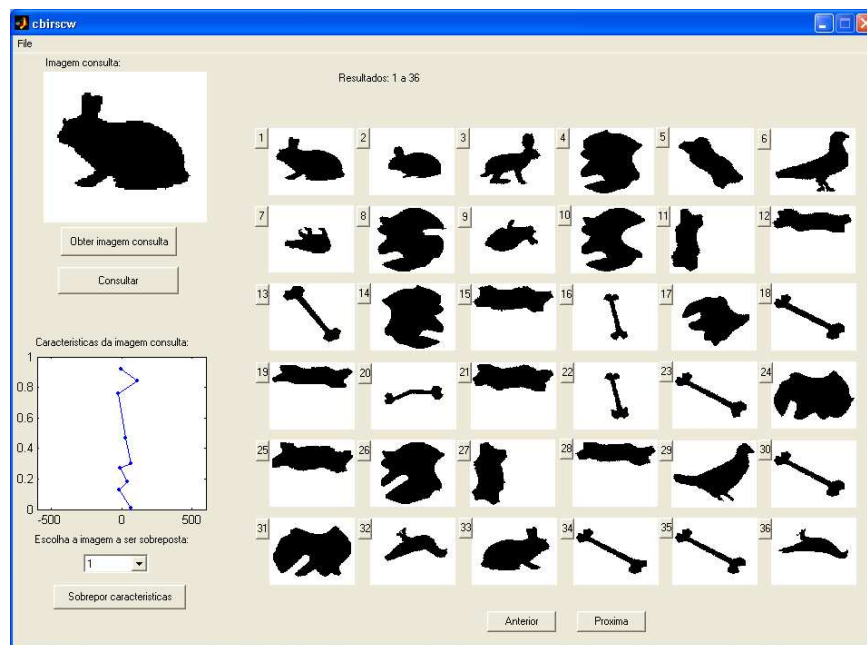


Figura 6.16: Ranking para imagem consulta coelho

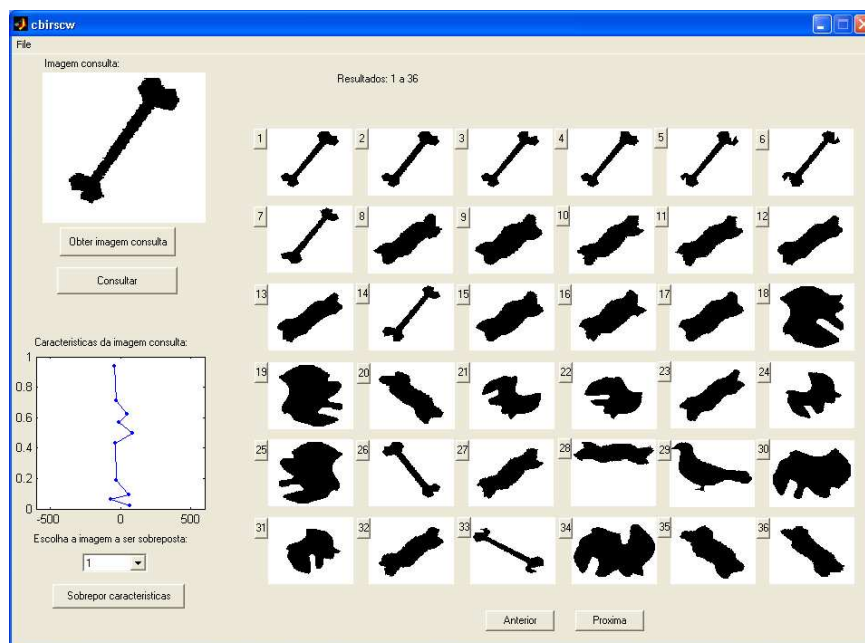


Figura 6.17: Ranking para imagem consulta osso

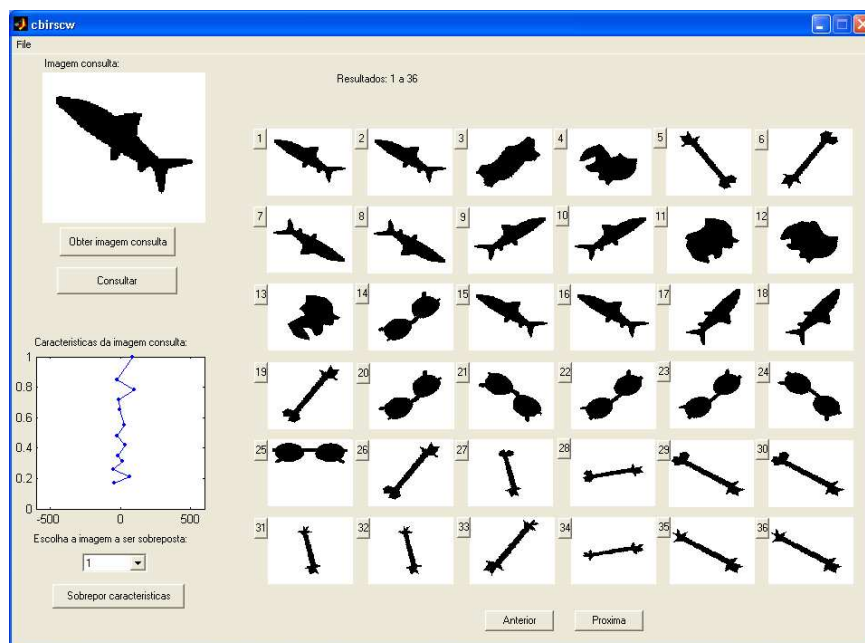


Figura 6.18: Ranking para imagem consulta peixe

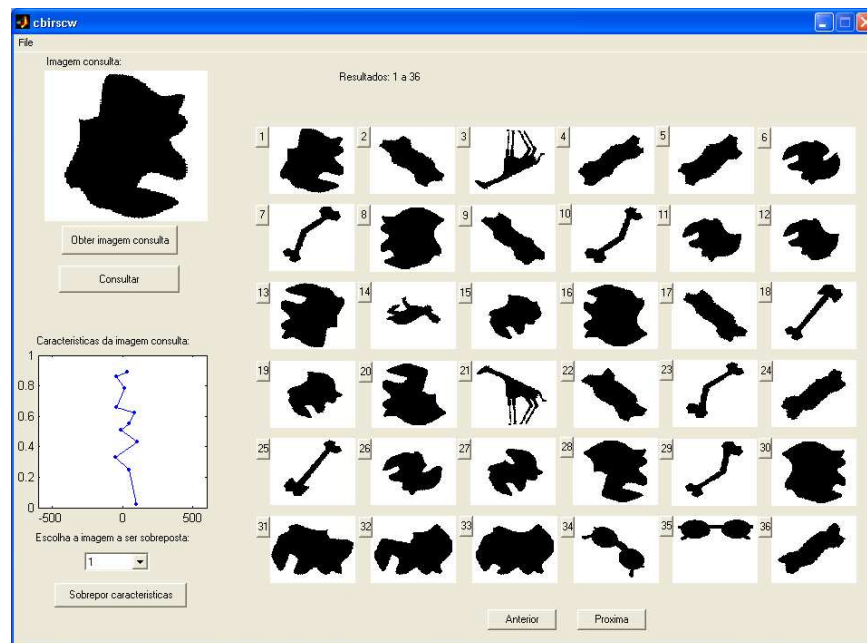


Figura 6.19: Ranking para imagem consulta máscara

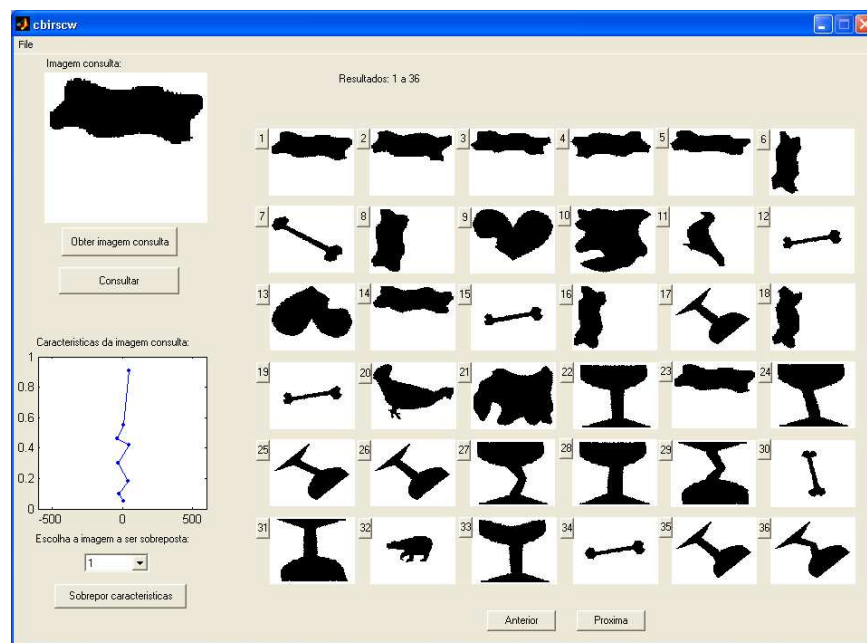


Figura 6.20: Ranking para imagem consulta tijolo

obtendo $d2$ - esse valor também será envolvido em cálculos;

3. Caso nenhuma saliência de um vetor esteja próxima a alguma saliência do outro vetor, o valor da distância é aumentado;
4. Divide-se o valor da distância por $d2$ - se nenhuma saliência casou, a distância d é bem grande e $d2$ é pequeno resultando em um valor de distância bem grande, o que traduz o desejo de que imagens muito diferentes tenham valor de dissimilaridade grande. Do contrário, se várias saliências casaram, a distância d é pequena e o valor $d2$ é grande resultando em pequena dissimilaridade.

6.4.2 Casamentos Parciais

Um problema verificado é que freqüentemente ocorrem casamentos parciais durante a recuperação. Uma forma se casa parcialmente com outra fazendo com que o valor de dissimilaridade seja reduzido e que portanto sua posição no *ranking* seja elevada. Um casamento parcial nesse contexto é quando uma parte das saliências de uma forma casa com parte das saliências de outra forma. Contudo, na maioria dos casos a similaridade que ocasiona o casamento é imperceptível já que ela ocorre em termos de um subconjunto de saliências.

Com a ocorrência desse tipo de casamento, torna-se difícil a determinação do conjunto de relevantes para uma imagem, o que por sua vez distorce os resultados obtidos para Precisão x Revocação. Um exemplo ocorrido é que freqüentemente formas de ossos se casam com formas de taças, isso porque as saliências do pé da taça se parecem com as saliências da extremidade do osso, ocorrendo um casamento parcial. Contudo, não se pode dizer, a princípio, que a imagem de um osso seja relevante quando a imagem consulta é uma taça. Uma outra situação que ilustra muito bem o fato é a seguinte: em uma consulta na qual a imagem consulta se

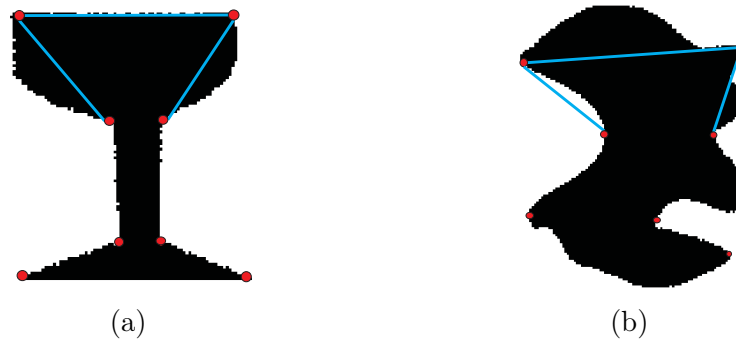


Figura 6.21: Casamentos Parciais entre imagens bastante distintas (a)Taça com saliências marcadas (b)Máscara com saliências marcadas

tratava de uma taça uma imagem aparentemente não-relevante foi retornada na posição 27 do *ranking*. Ao analisar este caso é possível perceber que ocorre um casamento parcial. Tal semelhança é mostrada na figura 6.21. Observa-se pontos marcando todas as saliências e linhas ligando saliências que casaram. Como os valores e a relação de localização entre elas são semelhantes, o casamento acontece elevando a figura 6.21(b) no *ranking*.

A principal constatação que pode ser feita a respeito é que o algoritmo pode falhar dependendo das características da base de dados. A aplicação do método em bases bem controladas em que uma imagem caracterize bem a forma ao mesmo tempo que não possibilite esse tipo de casamento parece ser melhor do que a aplicação do mesmo em uma base de domínio geral como a que se está utilizando, já que a origem extremamente distinta das imagens contribuem para tais casamentos. Assim como qualquer descritor destinado a reconhecimento/recuperação de imagens, este também encontra dificuldades quando aplicado a problemas de domínio geral [22]. Em uma base com um único tipo de imagem, as semelhanças são maiores e os casamentos estão mais para totais do que parciais, qualquer pequena diferença nas saliências de uma determinada forma com relação à imagem consulta

a farão descer no *ranking*.

6.4.3 Casamentos Exatos de Saliências

Outra tendência que se observa é a forte relevância dada pelo algoritmo de casamento à casamentos exatos de saliências. Quando duas saliências possuem exatamente o mesmo valor uma enorme diferença no valor total da dissimilaridade ocorre. Por exemplo, quando formas representadas por 5 saliências estão sendo comparadas, um caso em que 2 saliências casam exatamente e 3 não se casam pode ser mais importante do que quando 3 saliências casam não exatamente e 2 não se casam. Isso prejudica bastante o desempenho do sistema, pois quanto mais saliências casam mais semelhantes são as formas. No caso deste critério prevalecer, o impacto dos casamentos parciais poderia ser diminuído.

Portanto, aqui também se percebe a influência negativa de casamentos parciais. Já que o problema é exatamente a ocorrência deles e do peso dado aos mesmos, sendo que deveria ser privilegiadas no *ranking* formas com um maior número de casamentos entre as saliências. Percebe-se que pesos e uma análise melhor das saliências que casam contribuiriam bastante para um melhor desempenho do algoritmo de casamento.

6.4.4 Valores Repetidos de Dissimilaridade

Muitas vezes valores repetidos de dissimilaridade ocorrem como resultado do algoritmo de casamento. Com isso várias imagens no *ranking* possuem teoricamente a mesma classificação. No entanto sua posição, na prática, é dada pela ordem em que se encontram na base de dados.

O problema afeta algumas buscas, mas não chega a ser algo tão significativo na maioria dos casos. Ainda assim o gráfico Precisão x Revocação é alterado. Com

isso a inclusão de critérios de desempate na geração do *ranking* seriam bem vindos. Por exemplo, o número de saliências que casam poderia ser um bom critério.

6.4.5 Outras Dificuldades

Quando se tem formas muito intrincadas e mesmo muito finas existe grande dificuldade de se obter as saliências, afinal, para isso é preciso determinar o interior e a borda da imagem, e em alguns casos isso pode não ser possível.

6.5 Considerações Finais

Verifica-se que o sistema apresenta excelente precisão para baixos níveis de revocação. Seria desejável que a precisão se mantivesse alta à medida que a revocação cresce, mas ainda assim, o sistema cumpre seus propósitos de Recuperação de Informação já que as primeiras posições no *ranking* são as que realmente interessam [50].

Os resultados de Precisão x Revocação para muitas consultas não foram satisfatórios. Diversos fatores contribuem para este problema dentre eles, pode-se citar a produção de valores repetidos no algoritmo de casamento, casamentos exatos entre saliências e ainda a ocorrência de casamentos parciais em que as saliências de uma forma casam com algumas saliências da outra forma, enquanto outras são completamente diferentes, não sendo este um casamento adequado segundo os olhos de um especialista humano. Este problema está intrínseco na própria natureza da base de dados, já que esta permite a ocorrência desses casamentos parciais. Supõe-se que em uma base de dados de domínio específico o problema não ocorra. Como foi dito anteriormente, o domínio do problema ainda vem a ser uma das maiores dificuldades na recuperação de imagens baseada em conteúdo [22].

Outro fator que deve ser notado são as normalizações, uma vez que o algo-

ritmo não trabalha bem sem as mesmas. Tais normalizações são exatamente o que propiciam a alta precisão em baixos níveis de revocação.

Conclusão e Trabalhos Futuros

A recuperação de imagens embora tenha recebido bastante atenção dos pesquisadores ultimamente, ainda tem muito o que melhorar, pois trata-se de algo recente. As melhorias devem ser feitas principalmente no tocante a forma. Ambigüidades, perda de informação, e ruído dentre outros, tornam a recuperação baseada em forma pouco precisa.

Assim, descritores devem ser implementados de maneira eficiente não permitindo que esses problemas sejam transportados para a descrição da forma e ainda devem garantir invariância. Um grande leque de possibilidades foi aberto mostrando diversos descritores: descritores de fourier, momentos, *curvature scale space*, assinaturas, etc. O descritor escolhido para se trabalhar foi saliências do contorno, por ser novo e pouco explorado. Contudo, foi evidenciado que esse descritor não trabalha bem em domínios gerais, assim como ocorre em diversos contextos de recuperação/reconhecimento de imagens com outros descritores.

Além da qualidade do descritor, a métrica de comparação de imagens deverá capturar corretamente sua essência e ainda ser rápida para possibilitar seu uso em recuperação de informação.

Como trabalhos futuros, pretende-se:

- Desenvolver uma métrica de casamento própria para o descritor saliências do contorno. O algoritmo de casamento usado é o mesmo do descritor CSS, e pelo menos nos testes feito o desempenho não foi satisfatório.
- Fazer experimentos sobre diferentes maneiras de se determinar o valor da saliência. Quanto melhor for a representação de uma forma, melhor vai ser a recuperação.
- Estudar a criação de um novo descritor a partir do esqueleto.

Referências Bibliográficas

- [1] ABBASI, S., MOKHTARIAN, AND KITTLER, J. Curvature Scale Space Image in Shape Similarity Retrieval. *Multimedia Systems*, 7(6):467–476, 1999.
- [2] ABBASI, S., MOKHTARIAN, AND KITTLER, J. Enhancing CSS-based Shape Retrieval for Objectics with Shallow Concavities. *Vision Computacional*, 18:199–211, 2000.
- [3] ARBTER, K. Affine-Invariant Fourier Descriptors. In *J. C. Simon*. Elsevier Science Publishers, Amsterdam, North Holland, 1989.
- [4] ASADA, H., AND BRADY, M. The Curvature Primal Sketch. *IEEE Trans. Pattern Anal. Mach. Intell*, 1:2–14, 1986.
- [5] BATISTA, M. A. Retoque Digital. Dissertação de mestrado, Universidade Federal de Uberlândia, Uberlândia, MG, Brasil, Outubro 2003.
- [6] BIMBO, A. D., AND PALA, P. Visual Image Retrieval by Elastic Matching of User Sketches. *IEEE Trans. Pattern Anal. Mach.*, 19(2):121–132, 1997.
- [7] BLUM, H. A Transformation for Extracting New Descriptors of Shape. *Models for the Perception of Speech and Visual Form*, pages 260–268, 1967.
- [8] CASTELLI, C., AND BERGMAN, L. D. *Image Databases: Search and Retrieval of Digital Imagery*. Wiley, New York, USA, 2002.

- [9] CHOU, J. J. Voronoi Diagrams for Planar Shapes. *Computer Graphics and Applications*, 3:52–59, 1995.
- [10] COSTA, L. F., CAMPUS, A. G., AND MANOEL, E. T. M. Integrated Approach to Shape Analysis: Results e Perspectives. *International Conference on Quality Control by Artificial Vision*, 8:23–34, 2001.
- [11] COSTA, L. F., AND ESTROZI, L. F. Multiresolution shape representation without border shifting. *IEEE Electronics Letters*, 35(21):1829–1830, 1999.
- [12] DUBOIS, S. R., AND GLANZ, F. H. An Autoregressive Model Approach to two-Dimensional Shape Classification. *IEEE Trans. Pattern Anal. Mach.*, 8:627–637, 1986.
- [13] ERPEN, L. R. C. Recuperação de Informação Visual. Relatório técnico, Universidade Federal do Rio Grande do Sul, 2003.
- [14] FACON, J. Processamento e Análise de Imagens. Relatório técnico, Pontifícia Universidade Católica - PUC-Santa Catarina, 2002.
- [15] FALCÃO, A. X. Notas de aula. www.ic.unicamp.br/~afalcao/, 2003/2004/2005.
- [16] FALCÃO, A. X., COSTA, L. F., AND CUNHA, B. S. Multiescala Skeletons by Image Foresting Transform and its applications to neuromorphometry. *Pattern Recognition*, 35:1571–1582, 2002.
- [17] FALCÃO, A. X., AND CUNHA, B. S. Multiscale Shape Representation by Image Foresting Transform. In *SPIE on Medical Imaging*, volume 4322, pages 1091–1100, San Diego, CA, 2001.

- [18] FALCÃO, A. X., STOLFI, J., AND LOTUFO, R. A. Design of Connected Operators Using the Image Foresting Transform. In *SPIE on Medical Imaging*, volume 4322, pages 468–479, San Diego, CA, 2001.
- [19] FALCÃO, A. X., STOLFI, J., AND LOTUFO, R. A. The Image Foresting Transform: theory, algorithms and applications. *IEEE Trans. Pattern Anal. Mach. Intell*, 26(1):19–26, 2004.
- [20] FREEMAN, H. On the Encoding of Arbitrary Geometric Configurations. *IRE Trans. Electron. Comput.*, pages 260–268, 1961.
- [21] GOMES, J., AND VELHO, L. *Computação Gráfica: Imagem*. IMPA/SBM, Rio de Janeiro, 1994.
- [22] GONZALEZ, R. C., AND WOODS, R. E. *Digital Image Processing*. Prentice Hall, 2^a edition, 2002.
- [23] GROSKEY, W. I., NEO, P., AND MEHROTRA, R. A Pictorial Index Mechanism for Model-Based Matching. *Data Knowledge Eng.*, 8:309–327, 1992.
- [24] HIRATA, K., AND KATO, T. Query by Visual Example-Content Based Image Retrieval. In *Thrid International Conference on Extending Database Technology (EDBT'92)*, pages 56–71, Viena, Austria, 1992.
- [25] HOLMES, A. S., ROSE, C. J., AND TAYLOR, C. Measuring Similarity Between Pixel Signatures. *Image and Vision Computing*, 20:239–248, 2002.
- [26] HU, M. K. Visual Pattern Recognition by Moment Invariants. *IRE Trans. Inf. Theory*, 8:179–187, 1962.
- [27] JACOBS, C. E., FINKELSTEIN, A., AND SALESIN, D. H. Fast Multiresolution Image Query. In *ACM SIBGRAPI*, 1995.

- [28] KASHYAP, K. L., AND CHELLAPA, R. Stochastic Models for Closed Boundary Analysis: Representation and Reconstruction. *IEEE Trans. Inform. Theory*, 27:627–637, 1981.
- [29] LIN, C. C., AND CHELLAPA., R. Classification of Partial 2D Shapes Using Fourier Descriptors. *IEEE Pattern Anal. Mach. Intell*, 9(5):686–690, 1987.
- [30] LUPPE, M., COSTA, L. F., AND RODA, V. O. Parallel Implementation of Exact Dilations and Multi-Scale Skeletonization. *Real Time Imaging*, 9:163–169, 2003.
- [31] MEHROTRA, R., AND GARY, J. E. Similar-Shape Retrieval in Shape Data Management. *IEEE Comput.*, 9:57–62, 1995.
- [32] MITCHELL, O. R., AND GROGAN, T. A. Global and Partial Shape Discrimination for Computer Vision. *Opt. Eng.*, 23(5):484–491, 1984.
- [33] MOKHTARIAN, F., AND MACKWORTH, A. Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes. *IEEE Pattern Anal. Mach. Intell*, 8(1):34–43, 1986.
- [34] MOURA, D. M., SILVA, L. V., NERY, M. S., AND RODRIGUES, P. N. Recuperação de Imagens Baseado em Conteúdo. Relatório técnico, Universidade Federal de Minas Gerais - UFMG, Belo Horizonte, MG, Brasil, 2003.
- [35] NIBLACK, W., BARBER, R., EQUITZ, W., FLICKNER, M., GLASMAN, E., PETKOVIC, D., YANKER, P., FALOUTSOS, C., AND TAUBIN, G. The QBIC Project: Querying Image by Content Using Color, Texture and Shape. In *SPIE Storage and Retrieval for Image and Video Databases*, volume 1908, pages 173–187, San Jose, CA, USA, 1993.

- [36] PEIXOTO, A., AND VELHO, L. Transformadas de Distância. Relatório técnico, Pontifícia Universidade Católica - PUC-Rio, setembro 2000.
- [37] PLOTZE, R. O., AND BRUNO, O. M. Estudo e comparação de algoritmos de esqueletonização para imagens binárias. *IV Congresso Brasileiro de Computação - CBComp*, 2004.
- [38] RAUBER, T. W. Two-Dimensional Shape Description. Technical report: Gruninova-rt-10-94, Universidade Nova de Lisboa, Portugal, 1994.
- [39] RICHARD, C. W., AND HEMAMI, H. Identification of Three-Dimensional Objects Using Fourier Descriptors of the Boundary Curve. *IEEE Trans. System Man Cybernet*, 4:371–378, 1974.
- [40] RUI, Y., SHE, A. C., AND HUANG, T. S. A modified Fourier Descriptor for Shape Matching in MARS. In *Image Databases and Multimedia Search*, pages 165–177. Scientific Publishing Co., Singapore, 1997.
- [41] SHNEIDERMAN, B. *Designing the User Interface*. Addison Wesley, 2001.
- [42] SMITH, J. R., AND CHANG, S. F. Visual SEEK: A Fully Automatec Content-Based Image Query System. In *ACM Multimedia 96*, pages 87–98, Boston, MA, 1996.
- [43] TAUBIN, G., AND COOPER, D. B. Recognition and Positioning of Grid Objects Using Algebraic Moment Invariants. *SPIE Conference on Geometric Methods in Computer Vision*, 8:179–187, 1962.
- [44] TEAGUE, M. R. Image Analysis Via the General Theory of Moments. *J. Optical Soc. Am.*, pages 920–930, 1980.

- [45] TORRES, R. S., AND FALCÃO, A. X. Contour Saliency Descriptors for Effective Image Retrieval and Analysis. *Image and Vision Computing journal*, pages 1–11, 2006.
- [46] TORRES, R. S., FALCÃO, A. X., AND COSTA, L. F. Effective Image Retrieval by Shape Saliencies. In *XVI Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, pages 1530–1534, 2003.
- [47] TORRES, R. S., FALCÃO, A. X., AND COSTA, L. F. A Graph-based Approach for Multiscale Shape Analysis. *Pattern Recognition journal*, 37:1163–1174, 2004.
- [48] VELTKAMP, R. C., AND HAGERDOORN, M. Shape-of-the-Art in Shape Matching. Relatório técnico, Utrecht University, Utrecht, Netherlands, 2000.
- [49] VELTKAMP, R. C., AND HAGERDOORN, M. Shape Similarity Measures, Properties, and Constructions. Relatório técnico, Utrecht University, Utrecht, Netherlands, 2001.
- [50] YATES, B., AND NETO, R. *Modern Information Retrieval*. Addison Wesley, 1999.
- [51] ZHANG, D., AND LU, G. Review of Shape Representation and Description Techniques. *Pattern Recognition*, 37:1–19, 2004.
- [52] ZIBREIRA, C. Descrição e Procura de Vídeo Baseadas na Forma. Dissertação de mestrado, Universidade Técnica de Lisboa, Lisboa, Portugal, 2000.

© Cristiane de Fátima dos Santos & Célia Aparecida Zorzo Barcelos

Rua João Catanduva, 315 - Santa Mônica

38.408-240 Uberlândia (MG) - Brasil

Dissertação de Mestrado em Ciência da Computação - versão π

documento escrito em L^AT_EX 2 _{ϵ}

UFU - Universidade Federal de Uberlândia - FACOM