

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



**RECUPERAÇÃO DE IMAGENS: SIMILARIDADE PARCIAL
BASEADA EM ESPECTRO DE GRAFO E COR**

DALÍ FREIRE DIAS DOS SANTOS

Uberlândia - Minas Gerais

2012

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



DALÍ FREIRE DIAS DOS SANTOS

RECUPERAÇÃO DE IMAGENS: SIMILARIDADE PARCIAL BASEADA EM ESPECTRO DE GRAFO E COR

Dissertação de Mestrado apresentada à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como parte dos requisitos exigidos para obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Banco de Dados.

Orientadora:

Prof^a. Dr^a. Denise Guliato

Uberlândia, Minas Gerais
2012

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU

S237r Santos, Dalí Freire Dias dos, 1985-
Recuperação de imagens: similaridade parcial baseada em espectro
de grafo e cor / Dalí Freire Dias dos Santos. - 2012.
88 f. : il.

Orientadora: Denise Guliato.

Dissertação (mestrado) i Universidade Federal de Uberlândia, Pro-
grama de Pós-Graduação em Ciência da Computação.
Inclui bibliografia.

1. Computação - Teses. 2. Processamento de imagens - Teses.
3. Sistemas de recuperação da informação - Teses. I. Guliato, Denise.
II. Universidade Federal de Uberlândia. Programa de Pós-Graduação
em Ciência da Computação. III. Título.

CDU: 681.3

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da dissertação intitulada “**Recuperação de Imagens: Similaridade Parcial Baseada em Espectro de Grafo e Cor**” por **Dalí Freire Dias dos Santos** como parte dos requisitos exigidos para a obtenção do título de **Mestre em Ciência da Computação**.

Uberlândia, 17 de Agosto de 2012

Orientadora:

Prof^a. Dr^a. Denise Guliato
Universidade Federal de Uberlândia

Banca Examinadora:

Prof. Dr. Bruno Augusto Nassif Travençolo
Universidade Federal de Uberlândia

Prof. Dr. Ilmério Reis da Silva
Universidade Federal de Uberlândia

Prof. Dr. William Robson Schwartz
Universidade Federal de Minas Gerais

Dedicatória

Aos meus pais, irmão, esposa e amigos.

Agradecimentos

Obrigado meu Deus pela vida e por guiar meu caminho, obrigado minha mãe pela orientação e pela luta, obrigado meu pai e irmão pela força, obrigado Geórgia e família pelo carinho com que me receberam, obrigado aos colegas de trabalho e amigos por contribuírem na minha formação.

Gostaria de agradecer também aos amigos, professores e funcionários da Faculdade de Computação pela amizade e apoio no que fosse preciso. Em especial gostaria de agradecer ao professor Ilmério Reis da Silva, ao colega de pesquisa Robson de Carvalho Soares, ao secretário da Pós-graduação Erisvaldo Araújo Fialho, ao professor Manuel João da Fonseca da Universidade Técnica de Lisboa e a minha orientadora Denise Guliato.

A todos que de alguma forma contribuíram, muito obrigado!

“O que você sabe não tem valor algum. O valor está no que você faz com o que sabe.”
(Provérbio Chinês)

Resumo

Tradicionalmente, descritores de forma, ou descritores baseados em cor e textura, são utilizados para descrever o conteúdo visual das imagens. Embora essas abordagens apresentem bons resultados, elas não são capazes de diferenciar adequadamente imagens que contêm objetos com as mesmas cores, mas com organização espacial diferente ou não suportam a pesquisa parcial de imagens. Neste trabalho focamos em um caso particular da pesquisa parcial de imagens, que é encontrar objetos em imagens que contenham vários objetos, não deixando de lado a pesquisa total (encontrar imagens similares à original).

Dado que a cor é uma das propriedades visuais mais discriminativas, propomos um novo descritor baseado somente em cor capaz de encontrar imagens de objetos que estão contidos em outras imagens. Embora tenha apresentado melhores resultados quando comparado a trabalhos correlatos, esse novo descritor de cor não é capaz de discriminar objetos topologicamente diferentes mas que possuam as mesmas cores. Com o intuito de resolver esse problema, também propomos uma nova abordagem para a recuperação parcial de imagens que combina características topológicas e de cor em um único descritor. Esse novo descritor primeiramente realiza um processo de simplificação da imagem original, onde são identificadas as regiões de cor que compõem a imagem. Após esse processo de simplificação, a organização espacial das regiões de cor previamente identificadas é representada por meio de um grafo topológico, onde os vértices representam as regiões de cor e as arestas representam as conexões entre essas regiões. O descritor topológico é então calculado a partir do grafo de topologia utilizando a teoria espectral de grafos, evitando a necessidade de se realizar uma comparação direta entre grafos. Para suportar a pesquisa parcial de imagens, é realizada uma decomposição do grafo principal em diversos subgrafos.

Palavras chave: recuperação de imagens baseada em conteúdo, recuperação parcial de imagens, extração de características, descritor topológico, descritor de cor, espectro de grafos.

Abstract

Traditionally, local shape descriptors or color and texture based descriptors are used to describe the content of images. Although, these solutions achieving good results, they are not able to distinguish scenes that contain objects with the same colors, but with a different spatial organization or do not supports partial matching. In this work we focus on a particular case of the partial matching that is to find individual objects in images that contain various objects.

Since the color is one of the most visually distinguishable properties, we propose a new descriptor based only on color able to find pictures of objects that are contained in other images. Although our descriptor has shown better results when compared to related works, this new color descriptor is not able to discriminate objects topologically different but having the same colors. To overcome this problem, we also propose a new approach to the partial matching of images that combine color and topological features on a single descriptor. This new descriptor, first performs a simplification process of the original image, which identifies the color regions that make up the image. Then, we represent the spatial information among the color regions using a topological graph, where vertices represent the color regions and the edges represent the spatial connections between them. To calculate the descriptor from this graph representation, we use the spectral theory of graphs, avoiding the need to make a direct comparison between graphs. To support the partial matching, we propose a decomposition of the main graph into several subgraphs, and also calculate descriptors for these subgraphs.

Keywords: content based image retrieval, cbir, partial matching, feature extraction, topological descriptor, color descriptor, spectrum of graphs.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xiv
Lista de Abreviaturas e Siglas	xv
1 Introdução	16
1.1 Contribuições	17
1.2 Organização da Dissertação	19
2 Fundamentação Teórica	20
2.1 Imagens Digitais	20
2.1.1 Modelos de cor	21
2.1.2 Quantização de cores	23
2.1.3 Segmentação de imagens	24
2.2 Recuperação de Imagens por Conteúdo	26
2.2.1 Extração de características	28
2.2.2 Medidas de similaridade	32
2.3 Medidas de avaliação	34
2.3.1 <i>Precision and Recall</i>	35
2.3.2 <i>Average Precision</i>	36
2.3.3 <i>Mean Average Precision</i> (MAP)	37
2.4 Grafos	37
2.4.1 Conceitos básicos	38
2.4.2 Representação matricial	39
2.4.3 Teoria Espectral de Grafos	42
2.4.4 Representação de Imagens por meio de Grafos	44
2.5 Trabalhos Relacionados	45
2.6 Considerações Finais	48
3 <i>Color Blobs Descriptor</i>: Um novo descritor de cor para pesquisa parcial	50
3.1 Simplificação da Imagem	50

3.2	Medida de Similaridade	51
3.3	Considerações Finais	53
4	<i>Color Spectrum</i>: Combinando cor e topologia em um único descritor para pesquisa parcial	54
4.1	Simplificação da Imagem	54
4.2	Grafo Topológico	55
4.3	Medida de Similaridade	57
4.4	Descritor <i>Color Spectrum</i>	59
4.5	Decomposição em Subgrafos	60
4.6	Considerações Finais	62
5	Experimentos	63
5.1	Bases de Imagens	63
5.1.1	<i>Partial-COIL</i>	63
5.1.2	<i>Partial-COIL-T</i>	65
5.1.3	<i>Partial-Mock</i>	66
5.2	Influência do nível de quantização	68
5.3	Experimentos para o descritor CBD	69
5.3.1	Avaliação dos Resultados	70
5.4	Experimentos para o descritor <i>Color Spectrum</i>	74
5.4.1	Avaliação dos Resultados	75
5.5	Considerações Finais	79
6	Conclusão e Trabalhos Futuros	80
6.1	Conclusão	80
6.2	Trabalhos Futuros	81
	Referências Bibliográficas	82

Lista de Figuras

1.1	Exemplo de uma base de imagens para a consulta parcial.	18
1.2	Resultado desejado para uma consulta sobre o conjunto de imagens ilustrado na Figura 1.1.	18
2.1	Representação matricial de uma região de interesse de 10×10 <i>pixels</i> (b) de uma imagem digital (a).	20
2.2	Modelo de cor RGB.	22
2.3	Modelo de cor HSV.	22
2.4	Imagem original e suas representações quantizadas em 216, 64, 16 e 2 cores.	23
2.5	Exemplo do processo realizado pelo algoritmo <i>Mean Shift</i> até a convergência para um centro de massa.	25
2.6	Segmentação de uma fotografia utilizando a abordagem <i>Mean shift</i> [1].	26
2.7	Arquitetura típica de um sistema de recuperação de imagens por conteúdo [2].	27
2.8	Uma imagem colorida (a) e seus histogramas globais para cada canal RGB.	29
2.9	Exemplo de uma imagem em níveis de cinza e seus histogramas locais.	29
2.10	Exemplo de similaridade de formas baseada em região e contorno [3].	30
2.11	Exemplos de texturas.	31
2.12	Exemplo de uma função de distância métrica para comparar a similaridade entre fotos de animais.	33
2.13	Abrangência geométrica das funções da Família Minkowski.	33
2.14	Cálculo da distância cosseno entre dois vetores $\bar{a}$ e $\bar{b}$ [4]. $d(\bar{a}, \bar{b}) = \cos\theta$	34
2.15	Relacionamento entre os possíveis resultados para uma operação de busca.	35
2.16	Exemplo de uma curva Precisão x Revocação.	36
2.17	Exemplos de tipos de grafo.	38
2.18	Exemplo de um grafo (a) e um dos seus subgrafos (b).	39
2.19	Exemplo de grafos isomorfos $\{(a, 2), (b, 1), (c, 3), (d, 4), (e, 6), (f, 5)\}$	39
2.20	Um grafo qualquer (a) e sua respectiva matriz de adjacência (b).	40
2.21	Um grafo não ponderado (a) e sua respectiva matriz de graus (b).	41
2.22	Um grafo qualquer (a) e suas matrizes Laplaciana (b) e Laplaciana normalizada (c).	42
2.23	Um grafo qualquer (a) e sua respectiva matriz de adjacência (b).	43

2.24	Um <i>pixel</i> p qualquer as suas relações de vizinhança: (a) Vizinhança-4; (b) Vizinhança-8.	45
2.25	Representação de uma imagem em forma de RAG [5]: (a) Imagem original; (b) Imagem segmentada; (c) RAG.	45
2.26	Duas imagens distintas que apresentam o mesmo histograma de cores. . . .	46
2.27	Relacionamentos topológicos definidos por Egenhofer [6].	47
2.28	Relacionamentos topológicos originalmente definidos por Egenhofer (a) e a versão simplificada proposta por Fonseca et al. (b).	48
3.1	Processo de simplificação da imagem: (a) Imagem original; (b) Imagem borrada; (c) Imagem quantizada (216 cores); (d) Imagem simplificada. . . .	51
3.2	Cálculo do descritor CBD para uma imagem de consulta: (a) Imagem original; (b) Imagem simplificada; (c) CBD.	51
3.3	Cálculo do descritor CBD para uma imagem a ser comparada: (a) Imagem original; (b) Imagem simplificada; (c) CBD.	52
4.1	Processo de simplificação da imagem: (a) Imagem original; (b) Imagem borrada; (c) Imagem quantizada (16 cores); (d) Imagem simplificada. . . .	55
4.2	Relações topológicas entre regiões da imagem: (a) Disjuntos; (b) Inclusão; (c) Adjacência.	56
4.3	Imagem simplificada (a) e seu respectivo grafo topológico (b).	56
4.4	Extração do grafo topológico da Figura 4.1: (a) Imagem original; (b) Imagem simplificada; (c) Grafo topológico.	57
4.5	Cálculo do espectro de adjacência para a imagem da Figura 4.3. (a) Grafo topológico; (b) Matriz de adjacência; (c) Autovalores; (d) Espectro de adjacência.	58
4.6	Cálculo do espectro de adjacência para uma imagem de consulta. (a) Imagem simplificada; (b) Grafo topológico; (c) Matriz de adjacência; (d) Autovalores; (e) Espectro de adjacência.	58
4.7	Diagrama de blocos para a obtenção do descritor <i>Color Spectrum</i>	59
4.8	Cálculo do <i>Color Spectrum</i> para a imagem da Figura 4.3. (a) Grafo topológico; (b) Matriz de similaridade de cor $\mathbf{M}$; (c) Autovalores; (d) <i>Color Spectrum</i>	60
4.9	Subgrafos extraídos e suas partes correspondentes na imagem.	61
5.1	Objetos utilizados para criar a base de imagens COIL-100 [7].	64
5.2	Os 30 objetos aleatoriamente selecionados da base COIL-100.	64
5.3	Exemplos de imagens da base <i>Partial-COIL</i>	65
5.4	Exemplos de imagens da base <i>Partial-COIL-T</i>	66
5.5	Os 30 objetos criados para construir a base <i>Partial-Mock</i>	67

5.6	Exemplos de imagens da base <i>Partial-Mock</i>	68
5.7	Resultado do processo de simplificação aplicado à uma imagem original (a) utilizando diferentes níveis de quantização: (b) 16 cores; (c) 64 cores; (d) 216 cores.	69
5.8	Curvas de Precisão $\times$ Revocação para as bases: (a) <i>Partial-COIL</i> ; (b) <i>Partial-COIL-T</i> ; (c) <i>Partial-Mock</i>	70
5.9	As 10 primeiras imagens recuperadas da base <i>Partial-COIL-T</i> utilizando as abordagens: (b) CBD; (c) ACC; (d) CEDD; (e) FCTH.	72
5.10	As 10 primeiras imagens recuperadas da base <i>Partial-Mock</i> utilizando as abordagens: (b) CBD; (c) ACC; (d) CEDD; (e) FCTH.	73
5.11	Curvas de Precisão $\times$ Revocação para as bases: (a) <i>Partial-COIL-T</i> ; (b) <i>Partial-Mock</i>	75
5.12	As 10 primeiras imagens recuperadas da base <i>Partial-COIL-T</i> utilizando as abordagens: (b) <i>Color Spectrum</i> ; (c) BoVF/SIFT; (d) Espectro de Adj. + ACC; (e) Espectro de Adj.; (f) ACC.	77
5.13	As 10 primeiras imagens recuperadas da base <i>Partial-Mock</i> utilizando as abordagens: (b) <i>Color Spectrum</i> ; (c) BoVF/SIFT; (d) Espectro de Adj. + FCTH; (e) Espectro de Adj.; (f) FCTH.	78

Lista de Tabelas

5.1	Influência do nível de quantização na recuperação.	68
5.2	MAP para 100% de revocação (CBD).	70
5.3	MAP para 100% de revocação (<i>Color Spectrum</i>).	75

Lista de Abreviaturas e Siglas

ACC	<i>Auto Color Correlograms</i>
ARG	<i>Attributed Relation Graph</i>
BoVF	<i>Bag of Visual Feature</i>
CBD	<i>Color Blobs Descriptor</i>
CBIR	<i>Content-Based Image Retrieval</i>
CSS	<i>Curvature Scale Space</i>
CEDD	<i>Color and Edge Directivity Descriptor</i>
EHD	<i>Edge Histogram Descriptor</i>
FCTH	<i>Fuzzy Color and Texture Histogram</i>
GCH	<i>Global Color Histogram</i>
HSV	Modelo de cores <i>Hue-Saturation-Value</i> (Matiz, Saturação, Valor)
LCH	<i>Local Color Histogram</i>
LIRe	<i>Lucene Image Retrieval</i>
RAG	<i>Region Adjacency Graph</i>
RGB	Modelo de cores <i>Red-Green-Blue</i> (Vermelho, Verde, Azul)
SIFT	Scale Invariant Feature Transform
SGBD	Sistema de Gerenciamento de Banco de Dados
SURF	<i>Speeded-Up Robust Features</i>
TBIR	<i>Text-Based Image Retrieval</i>
TEG	Teoria Espectral de Grafos

Capítulo 1

Introdução

Na última década, houve um crescimento de 528,1% no número de usuários de Internet [8]. Nos primeiros dias do ano de 2012, 2,2 bilhão de pessoas tinham acesso à Internet em todo o mundo, o que representa 32,7% da população mundial. Em conjunto com o crescimento de usuários de Internet, presenciamos o crescente número de aplicações, serviços, portais, *websites* e sistemas computacionais disponíveis por meio da grande rede. Bancos de dados de imagens estão se tornando cada vez mais comum em diversos domínios de aplicação, tais como bibliotecas digitais, bancos de dados médicos e geográficos e investigação criminal. A evolução das técnicas para a aquisição, transmissão e armazenamento de imagens tem colaborado para o crescimento do volume de informações multimídia disponíveis. A busca por informação relevante em um cenário composto por temas tão diversos torna-se uma tarefa complexa, que tem inspirado pesquisadores da área de recuperação da informação.

Os sistemas de recuperação de informação comumente utilizados, como o *Google*, são baseados principalmente na busca textual, utilizando a similaridade entre palavras para encontrar conteúdo relevante. No caso dos sistemas de recuperação de imagens, a busca por conteúdo relevante torna-se um desafio de maior complexidade. Isso porque uma imagem pode ser descrita a partir das suas cores predominantes, dos objetos identificados na imagem, dos lugares aos quais ela se insere ou por características que dependem do indivíduo que a visualiza, sendo que dificilmente uma imagem será descrita da mesma maneira por indivíduos diferentes.

Os primeiros trabalhos relacionados ao processo de recuperação de imagens utilizavam uma técnica de pesquisa textual. Esta técnica, conhecida na literatura como *Text-Based Image Retrieval* (TBIR) [9–12], associa anotações textuais às imagens para uma posterior consulta baseada em texto utilizando Sistemas de Gerenciamento de Banco de Dados (SGBDs) tradicionais. No entanto, esse tipo de técnica requer a intervenção humana para obter as descrições textuais que serão associadas a todas as imagens, o que demanda um grande esforço. Existe ainda o problema da subjetividade da percepção humana, uma vez que a interpretação do conteúdo visual de uma imagem varia de acordo com a percepção

de cada indivíduo.

Para superar as dificuldades existentes na técnica de pesquisa textual, a técnica conhecida como *Content-Based Image Retrieval* (CBIR) [13–15] foi proposta. Nessa abordagem, em vez de serem manualmente anotadas com palavras-chave, as imagens são descritas por suas características visuais, tais como cor, forma e textura, extraídas de forma automática. O conjunto das características visuais extraídas é agrupado em um vetor de características. Assim, para o processo de busca, a distância entre dois vetores de características indica o grau de similaridade entre as respectivas imagens.

Tradicionalmente, descritores de forma como o *Scale Invariant Feature Transform* (SIFT) [16] e o *Speeded-Up Robust Features* (SURF) [17], ou descritores baseados em cor e textura, como por exemplo o *Color and Edge Directivity Descriptor* (CEDD) [18], o *Auto Color Correlograms* (ACC) [19] e o *Fuzzy Color and Texture Histogram* (FCTH) [20], são utilizados para descrever o conteúdo visual das imagens. Embora essas abordagens apresentem bons resultados, elas não são capazes de diferenciar adequadamente imagens que contêm objetos com as mesmas cores, mas com organização espacial diferente ou não suportam a pesquisa parcial de imagens. Neste trabalho descrevemos duas novas abordagens para o problema da pesquisa parcial, isto é, encontrar objetos individuais dentro de imagens contendo vários objetos, não deixando de lado a pesquisa total (encontrar imagens similares à original). Dado que a cor é uma das propriedades visuais mais discriminativas, propomos um descritor baseado somente em cor capaz de encontrar imagens de objetos que estão contidos em outras imagens. Embora tenha apresentado melhores resultados quando comparado a trabalhos correlatos, esse novo descritor de cor não é capaz de discriminar objetos topologicamente diferentes mas que possuam as mesmas cores. Com o intuito de resolver esse problema propomos uma nova abordagem para a recuperação parcial de imagens que combina características topológicas e de cor em um único descritor.

1.1 Contribuições

Considere uma base de imagens com diversas imagens, cada uma contendo vários objetos, como mostra a Figura 1.1. Um dos problemas que tem inspirado pesquisadores da área de recuperação de imagens por conteúdo nas últimas duas décadas é o problema da pesquisa parcial em imagens. Isto é, dada uma imagem de consulta, encontrar as imagens em que essa imagem de consulta fornecida esteja contida.

Embora várias soluções suportem a pesquisa parcial de imagens, algumas delas apresentam resultados ruins e outras, apesar de apresentarem bons resultados, não são capazes de diferenciar imagens que contêm o objeto de consulta mas com cores diferentes, ou objetos com as mesmas cores, mas com uma organização espacial diferente. Neste trabalho focamos em um caso particular da pesquisa parcial, que é encontrar objetos em imagens

é realizada uma decomposição do grafo principal em diversos subgrafos, e também são calculados descritores para esses subgrafos.

Como principais contribuições deste trabalho podemos destacar:

1. Proposta de um novo descritor de cor, denominado *Color Blobs Descriptor (CBD)*;
2. Avaliação da combinação entre descritores de cor e de topologia;
3. Proposta de um novo descritor que combina cor e topologia, denominado *Color Spectrum*;
4. Criação de bases de imagens destinadas a avaliação de métodos de pesquisa parcial de imagens;

1.2 Organização da Dissertação

Para expor com clareza os conceitos teóricos envolvidos, as técnicas e os resultados obtidos, essa dissertação foi organizada em seis capítulos. Esse capítulo realiza uma breve introdução sobre os sistemas de recuperação de imagens por conteúdo, apresenta as principais contribuições e a organização dessa dissertação. O Capítulo 2 apresenta os conceitos teóricos necessários para o entendimento deste trabalho juntamente com alguns dos principais trabalhos relacionados. Os Capítulos 3 e 4 descrevem de forma detalhada as abordagens propostas. O Capítulo 5 apresenta as avaliações e os experimentos realizados para cada abordagem proposta juntamente com as bases de imagens criadas para avaliar a qualidade no processo de recuperação parcial de imagens por conteúdo. Por fim, o Capítulo 6 relata as conclusões e propostas para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Este capítulo apresenta os conceitos teóricos necessários para o desenvolvimento deste trabalho.

2.1 Imagens Digitais

Uma imagem pode ser definida como uma função bidimensional, $f(x, y)$, onde x e y são coordenadas espaciais, e a amplitude de f em qualquer ponto (x, y) é chamada de intensidade ou cor da imagem neste ponto. Quando os valores da amplitude de f e de qualquer ponto (x, y) são todos finitos (discretizados), dizemos que esta imagem é uma imagem digital [22]. Dessa forma, podemos interpretar imagens digitais como sendo uma matriz cujos índices de linhas e colunas identificam um elemento na imagem, e o valor do elemento corresponde a intensidade ou cor naquele ponto. Cada um dos elementos dessa matriz é chamado de *pixel*, uma abreviação do termo em inglês *picture element*, sendo que a quantidade total de *pixels* de uma imagem é denominada resolução. A Figura 2.1 ilustra esses conceitos.

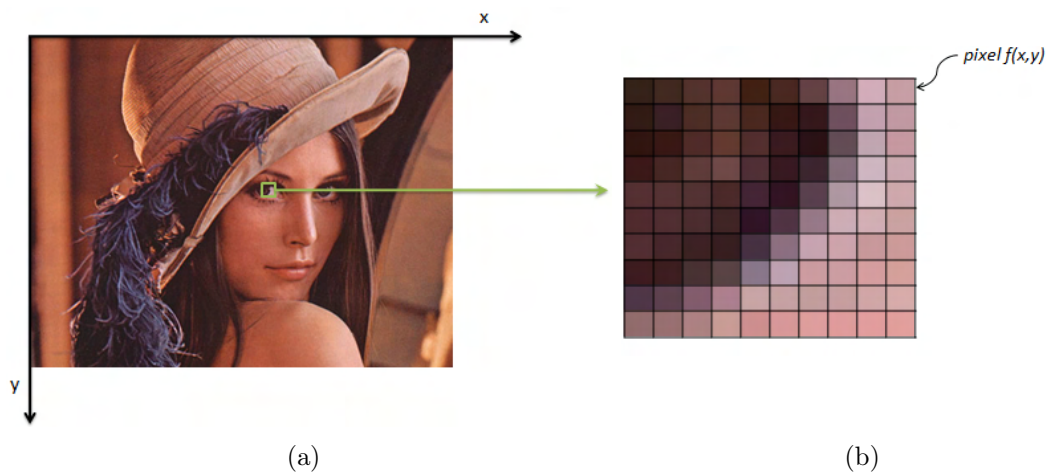


Figura 2.1: Representação matricial de uma região de interesse de 10×10 *pixels* (b) de uma imagem digital (a).

No caso de imagens digitais coloridas, cada *pixel* é descrito por meio de um conjunto de propriedades como por exemplo matiz, saturação e brilho. Assim, a cor de cada *pixel* é representada por um ponto no espaço definido pelo modelo de cor [22]. Os modelos de cor comumente utilizados são discutidos em maiores detalhes na Seção 2.1.1.

2.1.1 Modelos de cor

Um modelo de cor, também conhecido como espaço de cor ou sistema de cor, é um sistema de coordenadas 3D onde cada cor é representada por um ponto nesse espaço tridimensional [22]. Existem diversos modelos de cores, entre eles os mais utilizados atualmente são o RGB (*Red*, *Green*, *Blue*) e o HSV (*Hue*, *Saturation*, *Value*). O modelo RGB é um modelo aditivo no qual as características físicas do dispositivo de exibição são mapeadas diretamente. Já o modelo HSV reflete de forma mais precisa a maneira como os seres humanos descrevem e interpretam as cores.

Modelo RGB

No modelo RGB, as cores são descritas pela intensidade de cada uma das três cores primárias: Vermelho (*Red*), Verde (*Green*) e Azul (*Blue*). Assim, para representar uma cor no sistema RGB atribui-se um valor entre 0 e 255 para cada um dos componentes vermelho, verde e azul que o formam. A combinação dos valores RGB em quantidades iguais leva ao branco e, portanto, quanto mais alto são os valores RGB mais clara será a cor resultante. A cor vermelho puro, por exemplo, é especificada como (255, 0, 0), enquanto que a cor rosa claro é dada por (255, 220, 220). Quando duas das três cores primárias são sobrepostas, uma cor secundária é formada (veja a Figura 2.2). Por exemplo, quando as cores vermelho e verde se sobrepõem, a cor amarelo é criada. Variando os valores RGB, é possível obter cerca de 16 milhões de cores diferentes.

O modelo de cor RGB é dependente do dispositivo, isto significa que, a cor exibida não depende apenas dos valores RGB, mas também das especificações do dispositivo. Nesse caso, diferentes dispositivos podem exibir as cores de maneira diferente, causando, por exemplo, variações sensíveis ao olho humano na representação de determinado valor RGB em diferentes monitores.



Figura 2.2: Modelo de cor RGB.

Modelo HSV

O modelo de cor HSV é definido em função de três propriedades: Tonalidade (*Hue*), Saturação (*Saturation*) e Valor (*Value*). A tonalidade ou matiz, representa a cor dominante percebida por um observador. Assim, quando identificamos um objeto como sendo vermelho, laranja ou amarelo, estamos especificando sua tonalidade. Já a saturação refere-se à pureza relativa ou a quantidade de luz branca misturada à matiz. Quanto maior o valor da saturação, mais pura é a cor, sendo o grau de saturação inversamente proporcional a quantidade de luz branca adicionada [22]. Por fim, o valor representa a luminosidade ou brilho de uma cor.

O modelo HSV, é representado por um cone hexagonal conforme ilustrado na Figura 2.3. O eixo vertical (V) representa a variação do brilho ou intensidade da cor, o ângulo em torno do eixo vertical define a tonalidade (H), e a saturação (S) é medida pela distância em relação ao eixo vertical. Os valores para a tonalidade variam de 0 a 360 graus, iniciando no vermelho (0°), passando pelo amarelo (60°), verde (120°), ciano (180°), azul (240°), magenta (300°) e retornando ao vermelho (360° = 0°).

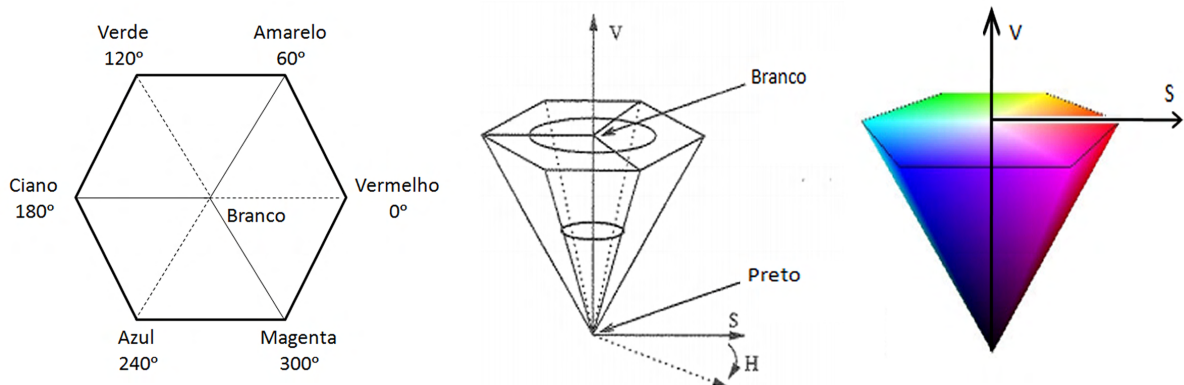


Figura 2.3: Modelo de cor HSV.

2.1.2 Quantização de cores

Em imagens digitais, as informações de cor podem ser representadas a nível de cada pixel, mas esta abordagem pode se tornar impraticável em sistemas CBIR. Suponha por exemplo que cada canal do espaço de cor RGB seja representado utilizando-se 8 *bits*, isto implica na possibilidade de representar 256 (2^8) níveis distintos para cada componente, resultando em 16.777.216 ($256 \times 256 \times 256$) cores distintas. A redução do número de cores de uma imagem é uma tarefa importante para a representação, segmentação e compressão de imagens coloridas. Esse processo é conhecido na literatura como quantização do espaço de cores. A Figura 2.4 ilustra o resultado da quantização de uma imagem em diferentes níveis.

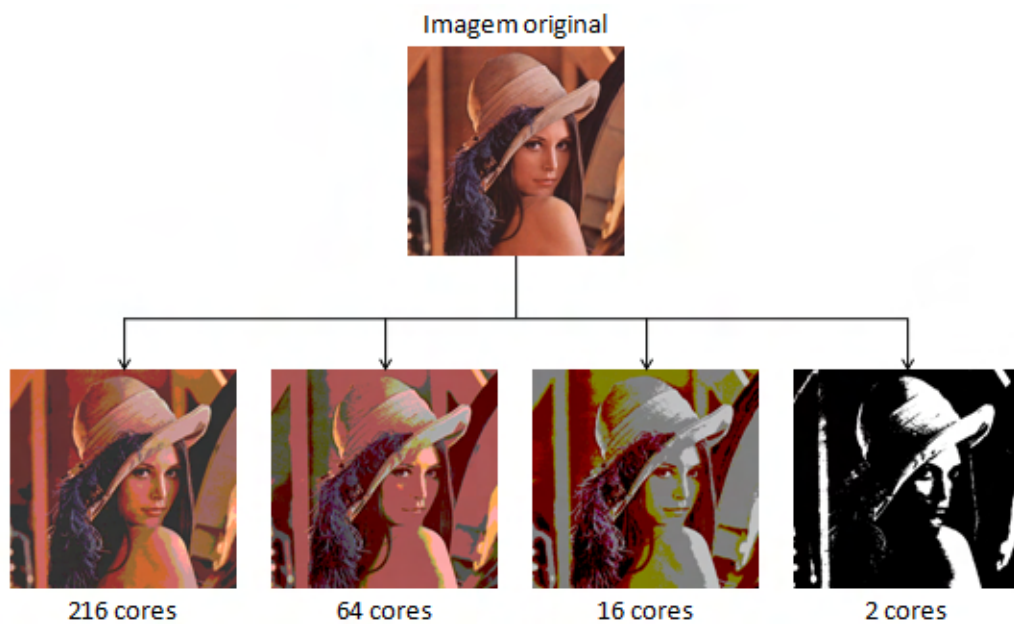


Figura 2.4: Imagem original e suas representações quantizadas em 216, 64, 16 e 2 cores.

Formalmente, o processo de quantização transforma um conjunto de cores com M elementos em um conjunto de cores com N elementos, onde $M > N$. Considere, por exemplo, o problema de quantizar uma determinada imagem $f(x,y)$, representada utilizando-se um espaço monocromático de cores com 256 níveis de cinza, para uma imagem $f'(x, y)$ a ser representada em um espaço de apenas 2 níveis de cinza. Uma solução possível seria para cada *pixel* da imagem $f(x,y)$:

$$f'(x, y) = \begin{cases} 0 & \text{se } f(x, y) < 127 \\ 255 & \text{se } f(x, y) \geq 127 \end{cases}$$

2.1.3 Segmentação de imagens

A segmentação de imagens é uma importante etapa de processamento utilizada em diversas aplicações nas áreas de imagem, video e visão computacional. Este processo consiste na divisão de uma determinada imagem em várias regiões de acordo com alguns critérios, a fim de identificar objetos de interesse presentes na cena. A segmentação de imagens também pode ser utilizada para identificar regiões de interesse (*foreground*) e diferenciá-las das demais regiões (*background*) [23, 24].

Em um processo de segmentação, os *pixels* das imagens são agrupados segundo alguns critérios de similaridade, como a cor ou a textura. As regiões identificadas devem ser uniformes e homogêneas em relação a estes critérios. Portanto, dois *pixels* de regiões adjacentes possuem uma diferença significativa no critério escolhido [23].

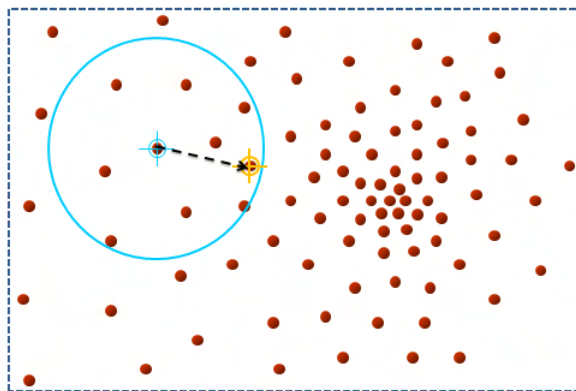
Existem diversas técnicas utilizadas para a segmentação de imagens, como por exemplo, crescimento de regiões [25], algoritmos de agrupamento [26] e de detecção de bordas [22], transformações *watershed* [27], limiarização [22], dentre outras. A seguir, descrevemos brevemente o método de segmentação baseado em cores que usaremos no desenvolvimento desse trabalho, denominado *Mean Shift* [1, 28].

Mean Shift

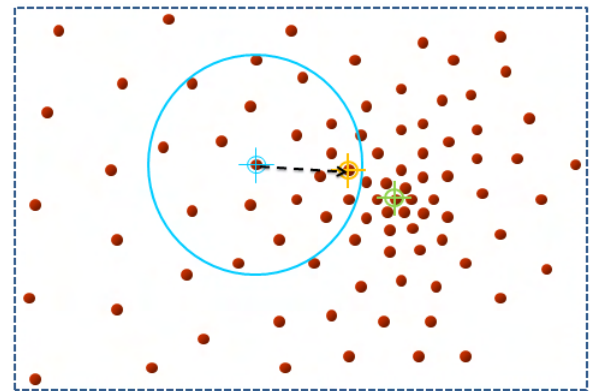
O *Mean Shift* é um algoritmo de agrupamento de dados comumente aplicado nas áreas de visão computacional e processamento de imagens. Ele foi originalmente proposto em 1975 por Fukunaga e Hostetler [28], tendo sido esquecido por cerca de 20 anos até que o trabalho de Cheng [29], explorando a sua utilização na área de visão computacional, reacendeu o interesse na sua utilização. A abordagem *mean shift* foi aplicada pela primeira vez à segmentação de imagens coloridas em 2002 por Comaniciu e Meer [1], onde os *pixels* são tratados como os objetos a serem classificados e as suas componentes de cor como as características usadas no processo de classificação. Ao final do processo, as classes são identificadas sem que seja necessário a supervisão do usuário. Cada classe é formada por um conjunto de *pixels* representados pela cor média dos *pixels* envolvidos, os quais se espera que estejam associados espacialmente formando as áreas segmentadas.

O algoritmo *Mean Shift* realiza um agrupamento de um conjunto de dados n-dimensional, associando cada ponto aos máximos de uma função densidade. O *Mean Shift* parte de cada ponto (ou *pixel* já que estamos trabalhando com imagens), delimita uma janela esférica em torno dele (*kernel*) e calcula o seu centro de massa. Encontrado o centro de massa, ele move o centro do *kernel* para o centro de massa calculado, e calcula um novo centro de massa para o conjunto de dados que agora está dentro do *kernel*. Encontrado o novo centro de massa, ele move o *kernel* novamente, e assim segue até a convergência do algoritmo, ou seja, até que o centro de massa passe a cair sempre no mesmo lugar. Ao final do processo, cada centro de massa representará uma região diferente da imagem. A

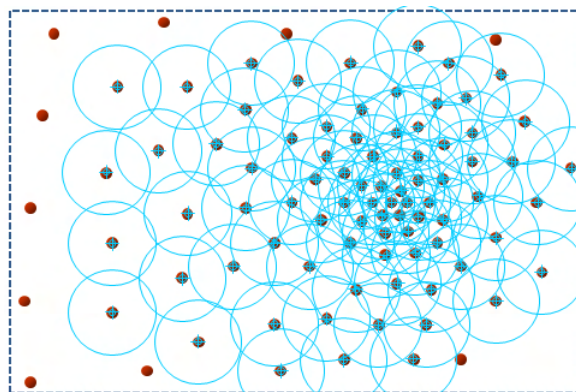
Figura 2.5 mostra exemplos de trajetórias percorridas por um *kernel* até a convergência do algoritmo.



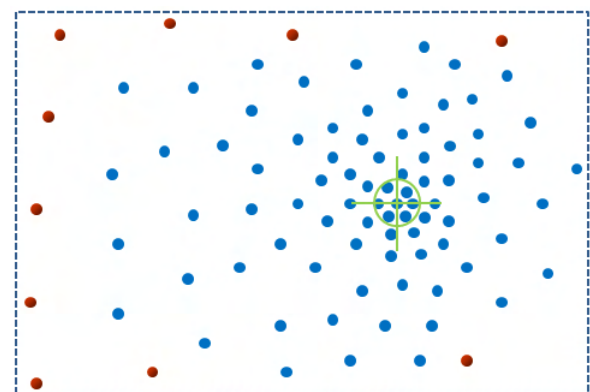
(a) Marcação de um *kernel* e do seu centro de massa.



(b) Movimentação do *kernel* para o centro de massa identificado no passo anterior. Ao final do processo o *kernel* inicial irá convergir para o centro de massa identificado.



(c) Marcação de diversos *kernels* que irão convergir para um mesmo centro de massa.



(d) Os pontos marcados em azul representam os pontos cujos *kernels* iniciais convergiram para o mesmo centro de massa indicado.

Figura 2.5: Exemplo do processo realizado pelo algoritmo *Mean Shift* até a convergência para um centro de massa.

O algoritmo *Mean Shift* necessita que a influência do *kernel* para cada *pixel* seja definida como parâmetro. Para isso, o *kernel* define uma medida de distância entre *pixels*, que engloba informações espaciais (*spatial radius*) e de cor (*color radius*). Dependendo da escolha desses parâmetros, a imagem pode ser subsegmentada (valores altos) ou supersegmentada (valores baixos). A título de exemplo, a Figura 2.6 mostra uma imagem colorida segmentada utilizando diferentes valores para os parâmetros *spatial radius* e *color radius*.



(a) Imagem original de um cachorro.



(b) Imagem após o processo de segmentação utilizando os parâmetros $spatial\ radius = 1$ e $color\ radius = 1$.



(c) Imagem após o processo de segmentação utilizando os parâmetros $spatial\ radius = 10$ e $color\ radius = 10$.



(d) Imagem após o processo de segmentação utilizando os parâmetros $spatial\ radius = 20$ e $color\ radius = 20$.

Figura 2.6: Segmentação de uma fotografia utilizando a abordagem *Mean shift* [1].

2.2 Recuperação de Imagens por Conteúdo

O principal objetivo do processo de recuperação de imagens por conteúdo é encontrar imagens relevantes à necessidade do usuário, a partir das características visuais extraídas a priori. Estas características visuais, geralmente cor, forma e textura, são extraídas por meio de descritores de características. O conjunto das características visuais extraídas pelos descritores dão origem aos chamados vetores de características. O processo de busca consiste basicamente em, dado uma imagem de consulta, calcular a sua similaridade em relação às imagens armazenadas na base e exibir as mais similares. Esta similaridade é obtida comparando os vetores de características da imagem de consulta com os vetores de características associados às imagens da base utilizando funções de distância.

Grande parte dos sistemas CBIR possui uma arquitetura similar, e envolve algumas etapas básicas como o armazenamento, a extração de características, e o cálculo de similaridade. A Figura 2.7 apresenta a arquitetura típica de um sistema de recuperação de imagens por conteúdo.

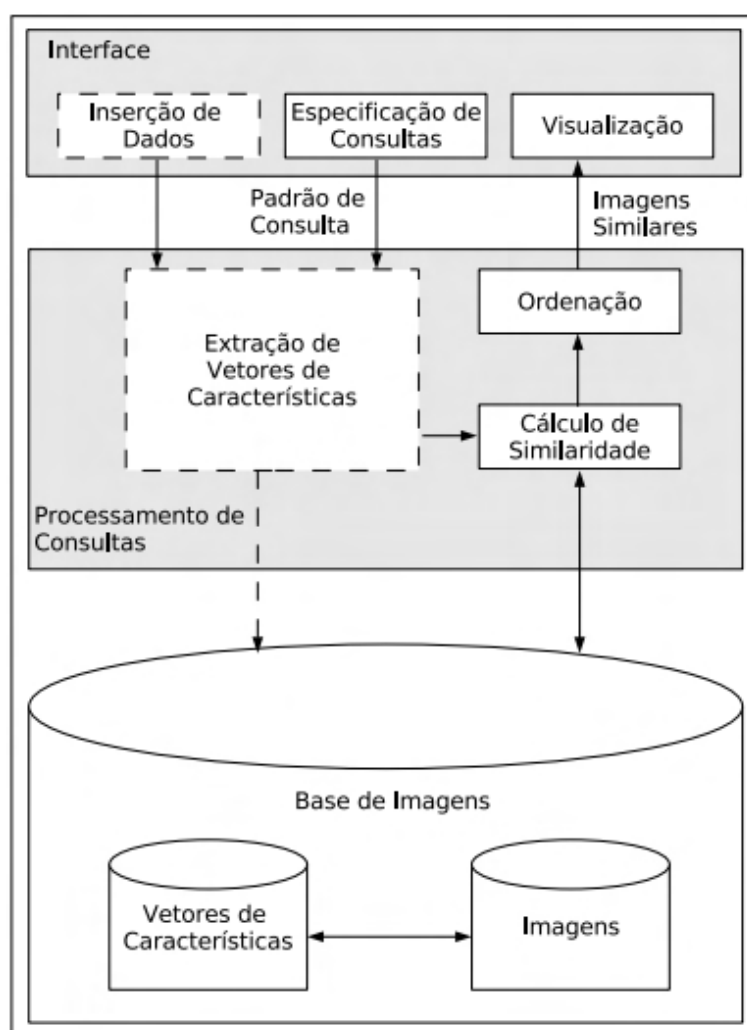


Figura 2.7: Arquitetura típica de um sistema de recuperação de imagens por conteúdo [2].

Inicialmente, os vetores de características para todas as imagens inseridas na base são calculados e armazenados. O usuário então fornece ao sistema uma imagem de consulta, cujo vetor de características também é extraído. Em seguida o sistema realiza o cálculo de similaridade entre o vetor de características da imagem de consulta e todos os vetores armazenados na base, preferencialmente utilizando algum método de indexação para reduzir o tempo de busca. Alguns sistemas ainda utilizam o processo conhecido como *relevance feedback* com o intuito de melhorar os resultados. Nesse processo o usuário tem a possibilidade de refazer a consulta refinando o resultado pela seleção das imagens identificadas como mais relevantes no resultado obtido anteriormente [30].

As seções a seguir apresentam os principais componentes de um sistema CBIR: os métodos utilizados para a extração das características visuais das imagens, alguns dos principais trabalhos relacionados, duas medidas de similaridade bastante utilizadas e algumas das medidas de avaliação de resultados comumente encontradas na literatura.

2.2.1 Extração de características

Em sistemas CBIR, chamamos de extração de características [31] o processo responsável por extrair uma medida numérica que tem por objetivo representar o conteúdo visual das imagens. As características extraídas por esse processo, também chamadas de características visuais, podem ser classificadas como globais (para uma imagem como um todo) ou locais (limitadas a regiões da imagem). Essas características são então agrupadas em vetores de características. Os vetores de características representam o conteúdo visual das imagens e são utilizados no processo de recuperação por conteúdo no lugar das imagens propriamente ditas.

As características visuais mais utilizadas para descrever imagens são: cor, forma e textura. Embora as características obtidas a partir das cores de uma imagem, de certa forma obtenha resultados satisfatórios em sistemas CBIR, existem situações onde o uso apenas dessa característica não é suficiente para uma boa recuperação. Sendo assim, pode ser necessário o uso de outras características como por exemplo informações espaciais e topológicas, textura e/ou forma com o intuito de refinar o processo de recuperação de imagens. A seguir apresentamos uma descrição mais detalhada de algumas das principais técnicas utilizadas para a extração de características.

Cor

Dentre as características visuais de baixo nível que podem ser utilizadas na recuperação de imagens por conteúdo, a cor é uma das mais utilizadas [32, 33]. Dentre os principais motivos que impulsionam o uso da cor, podemos destacar [34]: (1) os conceitos envolvidos são simples de serem entendidos e implementados; (2) a informação de cor está presente na grande maioria dos domínios de imagens; (3) a informação de cor pode ser processada de forma automática; (4) os resultados obtidos utilizando a informação de cor, em geral, são satisfatórios.

Os descritores de cor podem ser classificados em três grandes grupos de abordagens: abordagens globais, abordagens baseadas em partições regulares ou abordagens baseadas em regiões. As abordagens globais descrevem a informação de cor para a imagem como um todo, ignorando a distribuição espacial das cores. Grande parte dos descritores que seguem a abordagem global codificam as informações de cor em histogramas, como por exemplo o *Global Color Histogram* (GCH) [35]. Esse histograma é obtido contando-se, para cada uma das cores possíveis, o número de *pixels* da imagem com essas cores. A Figura 2.8 mostra um exemplo de uma imagem colorida e seu respectivo histograma global para cada um dos canais RGB.

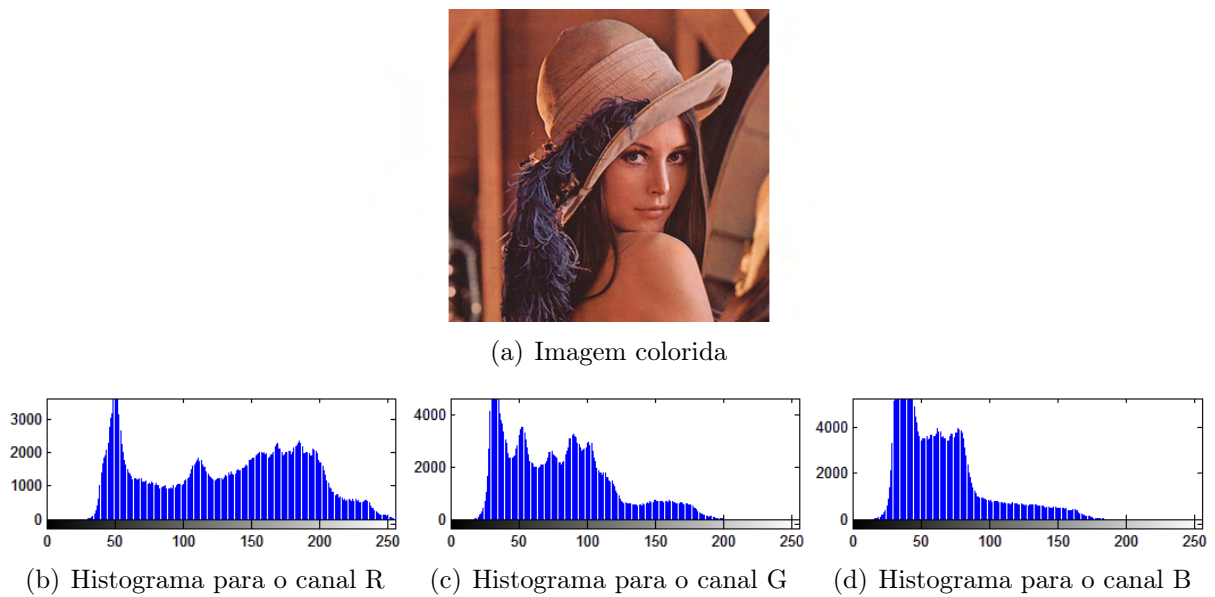


Figura 2.8: Uma imagem colorida (a) e seus histogramas globais para cada canal RGB.

As abordagens baseadas em partições regulares dividem as imagens em células de tamanho fixo, extraindo as informações de cor de cada célula separadamente. Por exemplo, cada imagem pode ser dividida em quatro (2×2) células retangulares de mesmo tamanho como mostra a Figura 2.9. A partir daí, a distribuição de cores de cada célula é descrita individualmente e representada por um histograma de cor, como faz o descritor *Local Color Histogram* (LCH) [35]. O objetivo do particionamento é tentar adicionar informação de como as cores estão espacialmente distribuídas dentro da imagem.



Figura 2.9: Exemplo de uma imagem em níveis de cinza e seus histogramas locais.

A abordagem baseada em região divide as imagens em regiões de acordo com o seu conteúdo visual. Essa divisão normalmente é realizada por meio de um algoritmo de segmentação ou agrupamento de cores (veja a Seção 2.1.3). A quantidade de regiões, assim como seu tamanho, localização e forma podem variar de uma imagem para outra. Em geral, os descritores baseados em região apresentam melhores resultados, embora quase sempre apresentem maior complexidade computacional. Os descritores *Color-Based Clustering* [36] e *Blobworld* [37] são exemplos de descritores que seguem a abordagem baseada em região.

Forma

A forma é uma das características visuais dos objetos que possibilita o seu reconhecimento. De fato, é possível identificar diferentes objetos em uma imagem analisando apenas as formas dos mesmos. Apesar da importância, a recuperação baseada em forma não é simples, especialmente devido à necessidade de segmentar os objetos de interesse presentes na imagem. Os descritores de forma podem ser classificados em dois grupos distintos [38]: descritores baseados em contorno e descritores baseados em região. Os descritores de forma baseados em contorno expressam as propriedades da forma somente pelo seu contorno (fronteira/borda). O contorno de uma forma pode sofrer efeitos negativos na presença de ruído na imagem. Como exemplos de descritores de forma baseados em contorno podemos citar os Descritores de Fourier [39] e o *Curvature Scale Space* (CSS) [40].

Os descritores de forma baseados em região consideram todos os *pixels* dentro de uma região da imagem, ao invés de considerar apenas a informação do contorno. Esse tipo de descritor é capaz de descrever adequadamente objetos complexos formados por múltiplas regiões desconectadas, bem como objetos simples com ou sem buracos, considerando as formas internas do objeto. Como exemplo desse tipo de descritor podemos citar os Momentos de Zernike [41].

A Figura 2.10 mostra as diferenças de similaridade ao utilizarmos os descritores de forma baseados em contorno e baseados em região. Os objetos da primeira linha possuem uma distribuição espacial de *pixels* similar entre eles. Portanto, serão considerados similares para descritores de forma baseados em região. Entretanto, os objetos da primeira linha possuem contornos diferentes si. Quando a similaridade baseada em contorno é utilizada, os objetos mostrados em cada coluna passam a ser os mais similares. Caso fosse realizada uma consulta pelo objeto localizado na primeira linha e segunda coluna, as imagens consideradas semelhantes seriam as da primeira linha (se a similaridade fosse baseada em região) ou da segunda coluna (se a similaridade fosse baseada no contorno) [3].

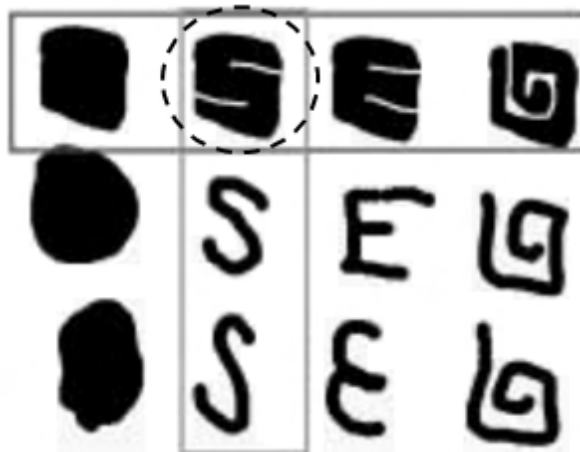


Figura 2.10: Exemplo de similaridade de formas baseada em região e contorno [3].

Textura

A textura é uma propriedade de uma determinada região da imagem suficientemente grande para demonstrar sua natureza [42]. O entendimento de textura depende do ponto de vista do observador, uma folha pode ser considerada um objeto, no entanto, a folhagem de uma árvore pode ser vista como uma textura. Textura é uma característica de praticamente todas as superfícies como, por exemplo, nuvens, árvores, tijolos, tecidos, entre outros. Ela contém informações importantes sobre a estrutura das superfícies e sua relação com o ambiente ao seu redor.

Texturas são caracterizadas pela relação entre os *pixels* da imagem, diferentemente da cor, que é uma característica própria de cada *pixel*. De uma maneira geral, a textura pode ser entendida como um conjunto de variações de intensidades que se repetem de maneira regular ou aleatória ao longo de um objeto ou imagem. A textura é uma característica fortemente discriminante e apresenta algumas propriedades como rugosidade, contraste, suavidade, regularidade e aspereza. A Figura 2.11 apresenta exemplos de diferentes texturas. Como exemplo de descritores de textura podemos citar os filtros de Gabor [43,44] e as transformadas *wavelets* [45].



Figura 2.11: Exemplos de texturas.

2.2.2 Medidas de similaridade

Para encontrar imagens relevantes à necessidade do usuário, em sistemas CBIR são utilizadas medidas de similaridade capazes de definir a semelhança entre as imagens da base e uma determinada imagem de referência fornecida pelo usuário. Como vimos na seção anterior, em sistemas CBIR, as imagens são representadas pelos seus vetores de características. Dessa forma, a semelhança entre imagens é obtida por meio de uma métrica capaz de comparar os seus vetores de características, retornando um valor numérico que quantifica o quão similares essas imagens são. Por definição esses valores são sempre maiores ou iguais a zero, sendo zero para imagens iguais. Na realidade, essas métricas indicam a dissimilaridade entre duas imagens, sendo que a similaridade é medida pelo seu inverso. Para simplificar a compreensão do texto, a maioria dos autores estabelecem que essas métricas representam a similaridade, uma vez que esta é facilmente obtida pela dissimilaridade.

Os modelos matemáticos comumente utilizados para suportar consultas por similaridade são conhecidos como espaços métricos. Um espaço métrico é composto por um conjunto de elementos (neste caso os vetores de características extraídos das imagens) e uma medida de similaridade para comparar esses elementos. Existem diversos modelos matemáticos utilizados como função de distância para cálculos de dissimilaridade.

Medidas de similaridade baseadas na função de distância

Uma função de distância d , utilizada para comparar imagens, é considerada métrica se, para quaisquer imagens x , y e z , as seguintes propriedades sejam satisfeitas:

1. Identidade: $d(x, x) = 0$, para todo x
2. Não negatividade: $d(x, y) \geq 0$, para todo x e y
3. Simetria: $d(x, y) = d(y, x)$, para todo x e y
4. Desigualdade triangular: $d(x, y) \leq d(x, z) + d(z, y)$, para todo x , y e z

A Figura 2.12 mostra uma função de distância métrica imaginária para comparar fotos de animais ilustrando as três propriedades citadas. Cada seta representa a distância entre as imagens ligadas. As distâncias são simétricas (setas duplas) e não negativas (zero quando são iguais), sendo que a distância direta entre os animais é sempre menor que as distâncias indiretas (desigualdade triangular).

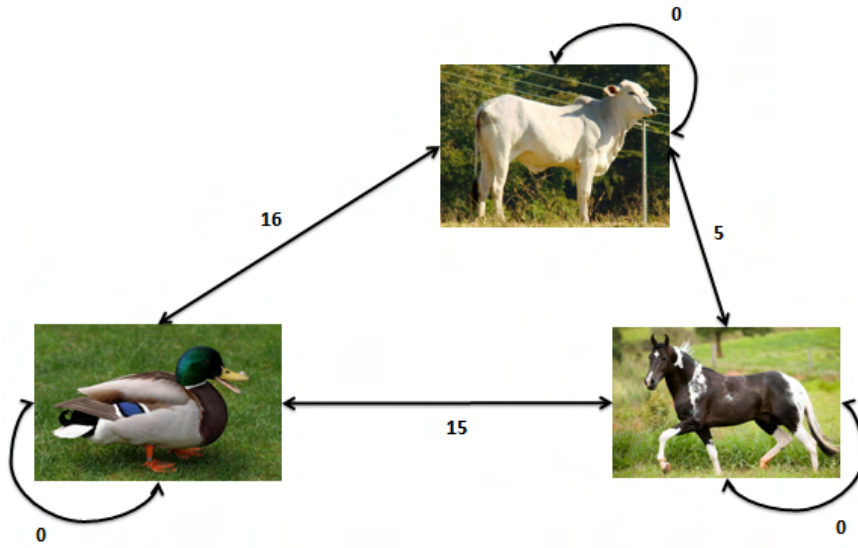


Figura 2.12: Exemplo de uma função de distância métrica para comparar a similaridade entre fotos de animais.

Um exemplo de medida de distância é a família Minkowski. As distâncias dessa família, também conhecida com L_p , estão entre as mais utilizadas na literatura. Suponha que $\bar{a} = (a_1, \dots, a_n)$ e $\bar{b} = (b_1, \dots, b_n)$ sejam dois vetores de características de dimensão n . Então a família de distâncias Minkowski pode ser definida pela Equação 2.1.

$$L_p(\bar{a}, \bar{b}) = \sqrt[p]{\sum_{i=1}^n |\bar{a}_i - \bar{b}_i|^p} \quad (2.1)$$

Note que, de acordo com o valor atribuído a p obtém-se as variações da família Minkowski. Para $p = 1$ temos a **Distância Manhattan** ou *City-Block* (L_1), para $p = 2$ temos a **Distância Euclidiana** (L_2) e para $p = \infty$ temos a **Distância Chebychev** ou *Infinity*. A Figura 2.13 ilustra as formas geradas a partir do conjunto de pontos que estão a uma mesma distância r , a partir de um centro o , para cada variação da Família Minkowski apresentada.

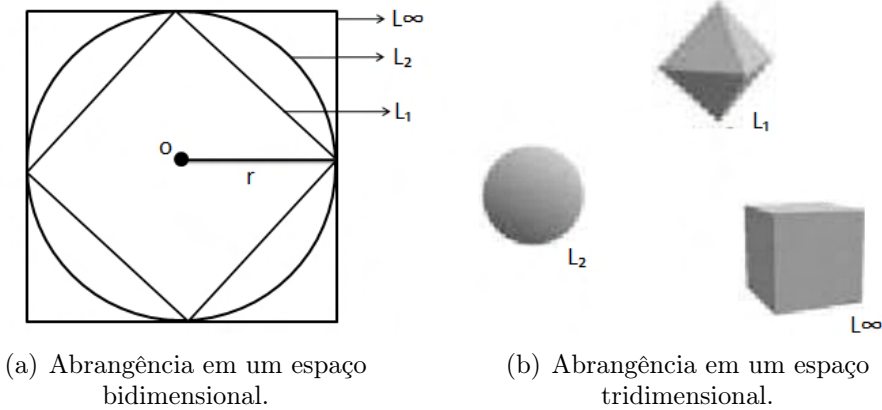


Figura 2.13: Abrangência geométrica das funções da Família Minkowski.

Medida de similaridade baseada no cosseno do ângulo entre vetores

Outra métrica comumente utilizada para o cálculo de similaridade é o cosseno do ângulo entre vetores. Para essa métrica, a similaridade entre duas imagens é medida utilizando o cosseno do ângulo formado por seus vetores de características. Nesse caso, quanto menor for o ângulo entre os vetores de características $\bar{a}$ e $\bar{b}$, maior será o grau de similaridade entre $\bar{a}$ e $\bar{b}$. Suponha que $\bar{a} = (a_1, \dots, a_n)$ e $\bar{b} = (b_1, \dots, b_n)$ sejam dois vetores de características de dimensão n . Então a distância cosseno pode ser formalmente definida pela Equação 2.2.

$$d(\bar{a}, \bar{b}) = \frac{\sum_{i=1}^n \bar{a}_i \bar{b}_i}{\sqrt{\sum_{i=1}^n \bar{a}_i^2 \sum_{i=1}^n \bar{b}_i^2}} \quad (2.2)$$

O cosseno do ângulo é independente da magnitude dos vetores de características, uma vez que o cálculo da distância considera apenas o cosseno do ângulo (θ) entre dois vetores [4], como mostra a Figura 2.14.

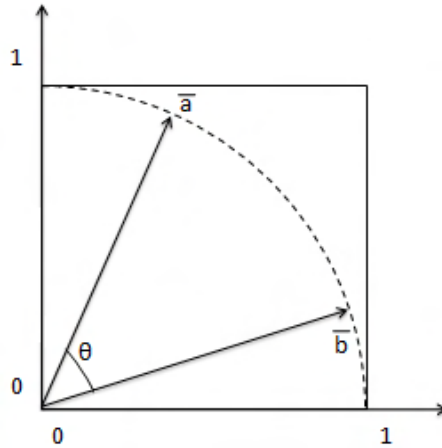


Figura 2.14: Cálculo da distância cosseno entre dois vetores $\bar{a}$ e $\bar{b}$ [4]. $d(\bar{a}, \bar{b}) = \cos\theta$

2.3 Medidas de avaliação

A qualidade de um sistema de recuperação de imagens por conteúdo está relacionada à satisfação do usuário com a lista de elementos retornados como resposta. Esta qualidade pode ser medida pela proporção de elementos relevantes na lista retornada como resposta, pela posição desses elementos relevantes com relação aos não relevantes ou ainda pela ausência de elementos relevantes. Um bom sistema CBIR deve encontrar todos os elementos relevantes e o mínimo possível de não relevantes [46].

As métricas utilizadas para comparar sistemas de recuperação de imagens por conteúdo são comumente chamadas de medidas de avaliação. As seções a seguir apresentam as principais medidas de avaliação utilizadas para medir a eficácia de sistemas CBIR.

2.3.1 Precision and Recall

Precisão e Revocação (*Precision and Recall*) são as duas medidas mais utilizadas para a avaliação da eficácia em sistemas de recuperação da informação [4]. O ponto fraco dessas medidas é que elas não contemplam a noção de *ranking*, ou seja, a ordenação dos resultados pela similaridade com a imagem utilizada como consulta. De maneira geral, os usuários de sistemas CBIR desejam que as imagens mais relevantes estejam nas primeiras posições da lista, o que é indiferente para essas medidas.

Considere a Figura 2.15. Dada uma consulta, definiremos R_{el} como sendo o conjunto das imagens relevantes em uma base de imagens D , e R_{ec} como sendo o conjunto das imagens recuperadas para tal consulta. A intersecção dos conjuntos R_{el} e R_{ec} compreende as imagens relevantes que foram recuperadas para a consulta realizada. Considere também $|R_{el}|$ e $|R_{ec}|$ como sendo a quantidade de elementos nos conjuntos R_{el} e R_{ec} , respectivamente, e $|R_{el} \cap R_{ec}|$ a quantidade de elementos na intersecção dos conjuntos R_{el} e R_{ec} , ou seja, a quantidade de elementos relevantes que foram recuperados. Note que, quanto maior for a intersecção entre os conjuntos R_{el} e R_{ec} , maior será a precisão para esta consulta.

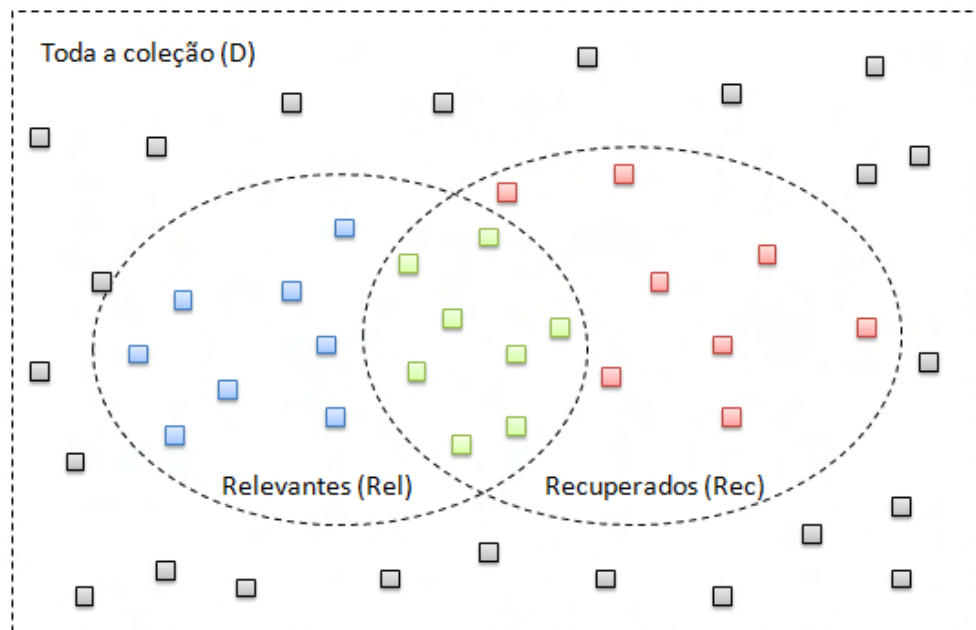


Figura 2.15: Relacionamento entre os possíveis resultados para uma operação de busca.

A **precisão** para uma consulta corresponde à capacidade do sistema recuperar somente elementos relevantes. Deste modo, temos que a precisão (P) é definida como sendo a quantidade de elementos relevantes recuperados, dividido pelo número total de documentos recuperados, como mostra a Equação 2.3. Por exemplo, se para uma consulta realizada forem recuperados 16 elementos e destes apenas 8 forem realmente relevantes, a precisão para esta consulta será de 50% (independentemente da posição em que os

relevantes ocorrem).

$$P = \frac{ElementosRelevantes \cap ElementosRecuperados}{ElementosRecuperados} = \frac{|R_{el} \cap R_{ec}|}{|R_{ec}|} \quad (2.3)$$

A **revocação** para uma consulta corresponde à capacidade do sistema recuperar todos os elementos relevantes existentes na base. Deste modo, temos que a taxa de revocação (R) é definida como sendo a quantidade de elementos relevantes recuperados, dividido pela quantidade total de elementos relevantes existentes na base de dados, como mostra a Equação 2.4. Seguindo o exemplo, caso fossem recuperados 16 elementos e destes apenas 8 forem realmente relevantes e a quantidade total de elementos relevantes da base fossem 16, a revocação para esta consulta será de 50%.

$$R = \frac{ElementosRelevantes \cap ElementosRecuperados}{ElementosRelevantes} = \frac{|R_{el} \cap R_{ec}|}{|R_{el}|} \quad (2.4)$$

As curvas **Precisão x Revocação** mostram a relação entre as duas medidas para cada nível de revocação, facilitando a análise comparativa entre diferentes sistemas. A Figura 2.16 mostra as curvas de precisão x revocação para uma mesma consulta realizada em três sistemas distintos. Neste exemplo, o Sistema 1 é mais preciso que os demais quando as taxas de revocação são baixas (até 50%), entretanto, a medida que as taxas de revocação vão subindo (acima dos 50%), o Sistema 2 passa a apresentar melhores taxas de precisão que os demais. Note que o gráfico atinge a marca de 100% de revocação quando todos os elementos relevantes existentes na base são recuperados.

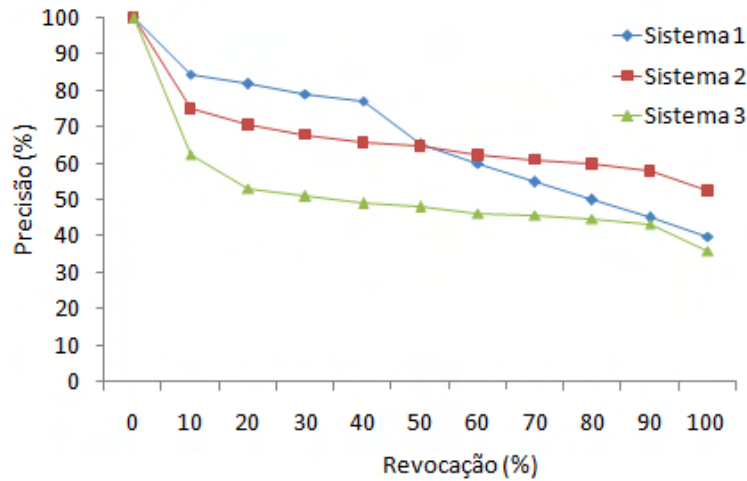


Figura 2.16: Exemplo de uma curva Precisão x Revocação.

2.3.2 Average Precision

A análise das curvas de Precisão x Revocação pode trazer informações relevantes sobre a eficácia dos sistemas avaliados, mas por vezes, existe a necessidade de se representar a qualidade dos sistemas avaliados por meio de um valor único. Uma das formas tradicionais

de se fazer isso é por meio da medida **Precisão Média** (*Average Precision*) [4]. A precisão média é obtida pela média aritmética dos valores de precisão ($P(k)$) até um determinado ponto, onde k representa a posição na lista retornada e $P(k)$ indica a precisão obtida no ponto k . Essa medida pode ser formalmente definida pela Equação 2.5.

$$A_{ve}P = \frac{\sum_{k=1}^n P(k)rel(k)}{|R_{el}|} \quad (2.5)$$

onde n representa o número de elementos recuperados, $|R_{el}|$ representa o total de elementos relevantes para a consulta e $rel(k)$ é uma função que indica a relevância do elemento presente na posição k , retornando 1 para relevante e 0 para não relevante.

2.3.3 Mean Average Precision (MAP)

A **Média dos Valores de Precisão Média** (*Mean Average Precision - MAP*) fornece uma medida de qualidade única para os diversos níveis de revocação [4]. Por meio dessa medida é possível obter o desempenho médio de um sistema de recuperação para um determinado número de consultas realizadas. Essa medida pode ser formalmente definida pela Equação 2.6.

$$MAP = \frac{\sum_{q=1}^Q A_{ve}P_q}{Q} \quad (2.6)$$

onde $A_{ve}P_q$ é a precisão média para a consulta q e Q é o número de consultas realizadas.

2.4 Grafos

A representação de problemas por meio de grafos foi proposto pela primeira vez pelo matemático suíço Leonard Euler, em 1736. Euler realizou o estudo de um problema existente na época conhecido como “As sete pontes de Königsberg” [47]. Modelando o problema por meio de grafos, Euler provou que seria impossível cruzar as sete pontes da cidade de Königsberg, hoje Kalingrado na Rússia, sem repetir nenhuma. Desde então, esse trabalho é considerado o marco zero da pesquisa em grafos, cuja solução originou a Teoria dos Grafos [48].

Os grafos podem ser utilizados para representar imagens através das regiões que a compõem e seus respectivos relacionamentos. Diversas soluções para a recuperação de imagens por conteúdo que utilizam grafos podem ser encontradas na literatura [49–53]. As seções a seguir apresentam os conceitos básicos da teoria dos grafos necessários para o entendimento deste trabalho. Maiores detalhes sobre as notações e conceitos básicos sobre grafos utilizados nesta e nas próximas seções podem ser encontrados em [21, 48, 54, 55].

2.4.1 Conceitos básicos

Conceitualmente, um grafo é um diagrama composto por um conjunto de elementos no espaço conectados entre si. A essa conexão entre os elementos pode ser atribuído um valor, ou seja, um peso. Tal peso define, portanto, se o grafo é ponderado ou não. Um **grafo** pode ser definido pelo par $G = (V, E)$, onde V é o conjunto de elementos no espaço e E é o conjunto de conexões entre os elementos de V . Os elementos de V são chamados de **vértices** ou nós, enquanto os elementos de E são chamados de **arestas**, arcos ou ligações.

Dois vértices v e w são considerados **conectados**, adjacentes ou vizinhos quando existe uma aresta ligando esses dois vértices. As arestas de um grafo podem ou não ter o seu sentido direcionado. Grafos que possuem arestas com sentido direcionado são chamados de dígrafos, grafos orientados ou grafos direcionados.

Para grafos não ponderados, o **grau** de um vértice u é o número de arestas que incidem no vértice u . Já para grafos ponderados, o grau de um determinado vértice é a soma dos pesos das arestas que incidem nesse vértice. A Figura 2.17 ilustra alguns dos conceitos abordados até o momento.

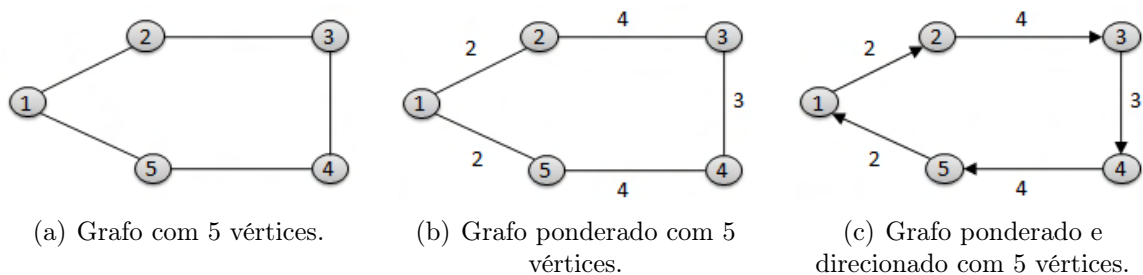


Figura 2.17: Exemplos de tipos de grafo.

A Figura 2.17(a) mostra um grafo $G = (V, E)$ não ponderado e não direcionado contendo 5 vértices. O conjunto dos vértices do grafo G é dado por $V = \{1, 2, 3, 4, 5\}$ e o conjunto das suas arestas é dado por $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{5, 1\}\}$. Além disso, todos os 5 vértices do grafo ilustrado na Figura 2.17(a) possuem grau 2.

Dizemos que o grafo $G_s(V_s, E_s)$ é um **subgrafo** do grafo $G(V, E)$, quando o conjunto dos seus vértices (V_s) é um subconjunto do conjunto de vértices do grafo G (V) e o conjunto das suas arestas (E_s) é um subconjunto do conjunto das arestas do grafo G (E). Assim, o grafo $G_s(V_s, E_s)$ é dito ser subgrafo de $G(V, E)$ quando $V_s \subseteq V$ e $E_s \subseteq E$. O grafo ilustrado na Figura 2.18(b), por exemplo, é um subgrafo do grafo ilustrado na Figura 2.18(a).

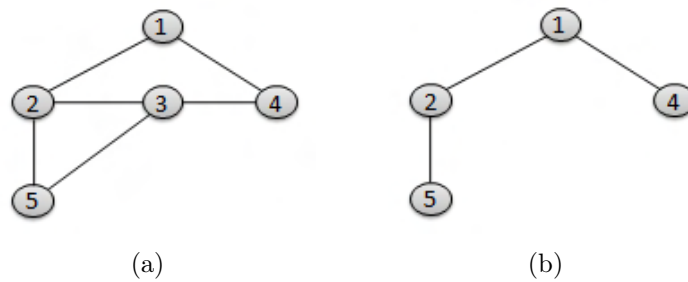


Figura 2.18: Exemplo de um grafo (a) e um dos seus subgrafos (b).

Dois grafos $G_1 = (V_1, E_1)$ e $G_2 = (V_2, E_2)$ são considerados **isomorfos** quando existe uma correspondência entre os seus vértices, de forma que as suas arestas sejam preservadas. Em outras palavras, dois grafos são isomorfos se é possível alterar os nomes dos vértices de um deles de tal modo que os dois grafos fiquem iguais. Assim, um isomorfismo de G_1 sobre G_2 é um mapeamento bijetivo $f : V_1 \leftrightarrow V_2$ tal que $(x, y) \in E_1$ se e somente se $(f(x), f(y)) \in E_2$, para todo $x, y \in V_1$. A Figura 2.19 mostra um exemplo de dois grafos isomorfos.

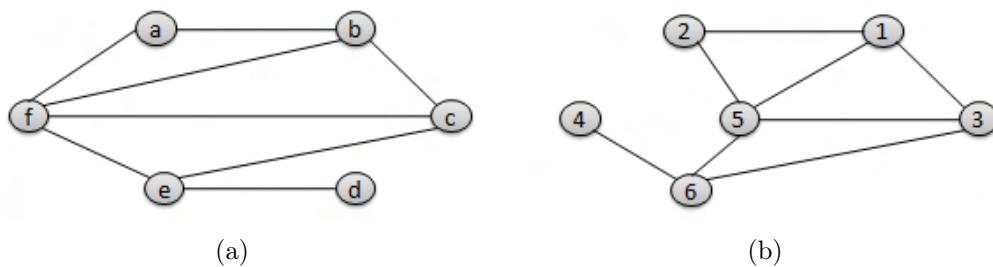


Figura 2.19: Exemplo de grafos isomorfos $\{(a, 2), (b, 1), (c, 3), (d, 4), (e, 6), (f, 5)\}$.

2.4.2 Representação matricial

A representação de grafos por meio de diagramas pode ser útil para se analisar a estrutura do grafo como um todo, bem como as relações entre seus vértices. No entanto, a representação de grafos por meio de matrizes pode ser indicada para grafos com um grande número de vértices e arestas. Além disso, a utilização de matrizes pode facilitar a representação e o processamento de grafos em computadores. Existem diversas formas de se representar grafos por meio de matrizes. Nas seções a seguir, apresentamos algumas das representações mais utilizadas na literatura.

Matriz de adjacência

Como vimos na Seção 2.4.1, dois vértices são conectados quando existe uma aresta ligando esses dois vértices. A matriz de adjacência de um grafo $G = (V, E)$ é uma matriz quadrada A de $|V| \times |V|$, tal que cada elemento A_{ij} indica se os vértices i e j são conectados

ou não [55]. Para grafos não ponderados, cada elemento A_{ij} tem valor 1 sempre que i e j são conectados, e 0 caso contrário. Esta matriz é formalmente definida pela Equação 2.7:

$$A(i, j) = \begin{cases} 1 & \text{se } i \text{ e } j \text{ são conectados} \\ 0 & \text{caso contrário} \end{cases} \quad (2.7)$$

Para grafos ponderados, a matriz de adjacência, denotada por W , indica a distribuição dos pesos das arestas entre vértices adjacentes. Tal matriz é formalmente definida pela Equação 2.8:

$$W(i, j) = \begin{cases} p(i, j) & \text{se } i \text{ e } j \text{ são conectados} \\ 0 & \text{caso contrário} \end{cases} \quad (2.8)$$

onde $p(i, j)$ indica o peso da aresta que liga os vértices i e j . A Figura 2.20 mostra um exemplo de um grafo e sua respectiva matriz de adjacência.

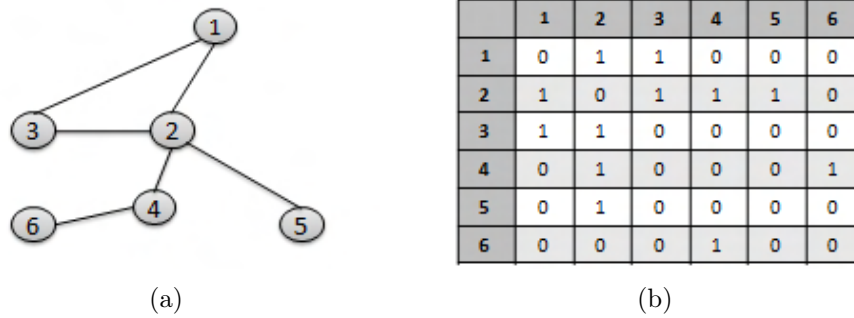


Figura 2.20: Um grafo qualquer (a) e sua respectiva matriz de adjacência (b).

Matriz de graus

Conforme visto na Seção 2.4.1, o grau de um vértice corresponde ao número de arestas que incidem nesse vértice (para grafos não ponderados) ou à somatória dos pesos das suas arestas (para grafos ponderados). A matriz de graus de um grafo $G = (V, E)$, denotada por D , é uma matriz diagonal $|V| \times |V|$ que indica o número de conexões de cada vértice do grafo G [55]. Esta matriz é formalmente definida pela Equação 2.9:

$$D(i, j) = \begin{cases} grau(i) & \text{se } i = j \\ 0 & \text{caso contrário} \end{cases} \quad (2.9)$$

onde $grau(i)$ indica o grau do vértice i . A Figura 2.21 mostra um exemplo de um grafo e sua respectiva matriz de graus.

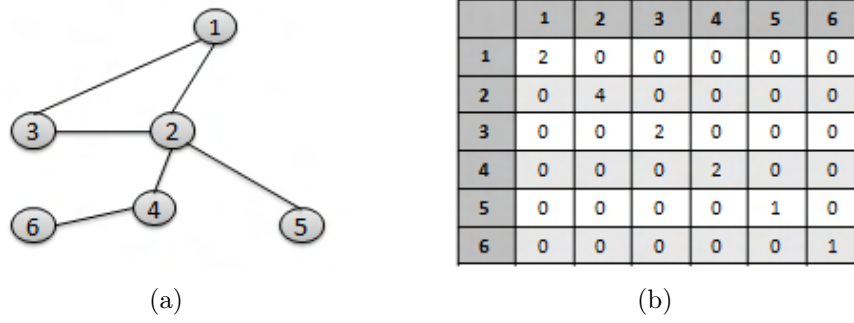


Figura 2.21: Um grafo não ponderado (a) e sua respectiva matriz de graus (b).

Matriz Laplaciana

A matriz Laplaciana de um grafo $G = (V, E)$, denotada por L , é uma matriz quadrada $|V| \times |V|$ bastante utilizada na literatura para a representação de grafos [21]. Essa matriz, também conhecida como matriz de Kirchhoff, é formalmente definida pela Equação 2.10:

$$L(i, j) = \begin{cases} grau(i) & \text{se } i = j \\ -1 & \text{se } i \text{ e } j \text{ são conectados} \\ 0 & \text{caso contrário} \end{cases} \quad (2.10)$$

onde $grau(i)$ indica o grau do vértice i . A matriz Laplaciana também pode ser obtida por meio da subtração entre a matriz de graus e a matriz de adjacência [54], como mostra a Equação 2.11.

$$L = D - A \quad (2.11)$$

Uma variação da matriz Laplaciana, também muito utilizada na literatura para a representação e análise das características dos grafos, foi proposta por Chung [21]. Essa representação, conhecida como matriz Laplaciana normalizada, é definida pela Equação 2.12:

$$L_N(i, j) = \begin{cases} 1 & \text{se } i = j \\ -\frac{1}{\sqrt{grau(i) \times grau(j)}} & \text{se } i \text{ e } j \text{ são conectados} \\ 0 & \text{caso contrário} \end{cases} \quad (2.12)$$

onde $grau(i)$ indica o grau do vértice i e $grau(j)$ indica o grau do vértice j . A seguir, a Figura 2.22 mostra um grafo qualquer e suas representações matriciais laplacianas.

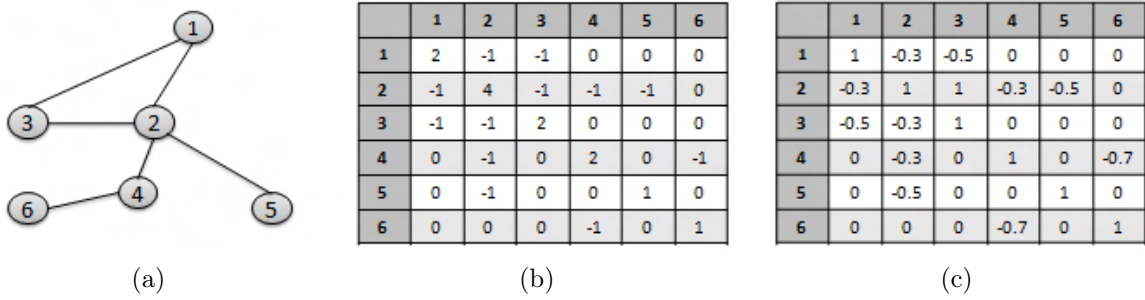


Figura 2.22: Um grafo qualquer (a) e suas matrizes Laplaciana (b) e Laplaciana normalizada (c).

2.4.3 Teoria Espectral de Grafos

Devido ao fato de que uma série de propriedades de um grafo podem ser deduzidas a partir do espectro da sua representação matricial, tal ramo da Teoria dos Grafos é denominada Teoria Espectral de Grafos (TEG). A origem da TEG se deu na Química Quântica quando, em 1931, Huckel produziu um modelo teórico para um problema a partir de moléculas de hidrocarbonetos não saturadas em que os níveis de energia de certos elétrons eram representados por autovalores de um grafo [56]. No entanto, a fundamentação teórica da Teoria Espectral de Grafos iniciou-se somente em 1980, com a publicação do primeiro livro sobre o assunto [57], tendo se popularizado nas últimas duas décadas [56].

O **espectro de um grafo** $G = (V, E)$ é definido como sendo o conjunto dos autovalores da sua representação matricial [58]. Considere M como sendo uma matriz quadrada de ordem n . Um **autovalor** de M é um escalar λ tal que $Mx = \lambda x$, sendo $x \neq 0$. Para o cálculo dos autovalores da matriz M deve-se encontrar as raízes do polinômio característico de M , denotado pela Equação 2.13:

$$p_M(\lambda) = \det(M - \lambda I_n) \quad (2.13)$$

onde I_n é a matriz identidade de ordem n .

A multiplicidade algébrica de um autovalor λ é o número de vezes que λ ocorre como raiz do polinômio característico $p_M(\lambda)$. Para um grafo com n vértices, existem n autovalores, não necessariamente distintos, representados pelas raízes do seu polinômio característico [56]. O espectro de um grafo é então definido como sendo o conjunto dos autovalores apresentados em ordem decrescente, ou seja, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \dots \geq \lambda_n$.

Tome como exemplo um grafo qualquer e sua respectiva matriz de adjacência A , como ilustra a Figura 2.23. Na sequência, detalhamos os passos necessários para o cálculo do espectro de adjacência do grafo ilustrado na Figura 2.23(a)

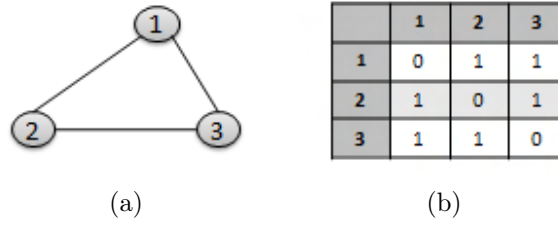


Figura 2.23: Um grafo qualquer (a) e sua respectiva matriz de adjacência (b).

Os autovalores da matriz de adjacência A são calculados por meio do seu polinômio característico. A partir da Equação 2.13, temos que o polinômio característico de uma matriz é obtido calculando-se o determinante da matriz diferença $A - \lambda I_n$, como se segue:

$$\begin{aligned}
 A &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \\
 I_n &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
 A - \lambda I_n &= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} = \begin{pmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{pmatrix}
 \end{aligned}$$

Uma das formas de se calcular o determinante da matriz diferença $A - \lambda I_n$ é utilizando-se a Regra de Sarrus:

$$\begin{aligned}
 \det(A - \lambda I_n) &= (-\lambda \cdot -\lambda \cdot -\lambda) + (1 \cdot 1 \cdot 1) + (1 \cdot 1 \cdot 1) - (-\lambda \cdot 1 \cdot 1) - (1 \cdot 1 \cdot -\lambda) - (1 \cdot -\lambda \cdot 1) \\
 \det(A - \lambda I_n) &= -\lambda^3 + 1 + 1 - (-\lambda) - (-\lambda) - (-\lambda) \\
 \det(A - \lambda I_n) &= -\lambda^3 + 3\lambda + 2
 \end{aligned}$$

Calculando-se as raízes do polinômio característico, temos os três autovalores λ_1 , λ_2 e λ_3 para a matriz de adjacência A :

$$\begin{aligned}
 p_M(\lambda) &= \det(A - \lambda I_n) \\
 0 &= -\lambda^3 + 3\lambda + 2 \\
 \lambda_1 &= 2 \\
 \lambda_2 &= -1 \\
 \lambda_3 &= -1
 \end{aligned}$$

O espectro de adjacência para o grafo da Figura 2.23(a) é obtido ordenando de maneira decrescente os autovalores calculados para a sua matriz de adjacência. Nesse caso, $\bar{e} =$

$\{2, -1, -1\}$.

Uma das questões em aberto da Teoria dos Grafos se refere à caracterização de uma lista completa de parâmetros capaz de decidir se dois grafos são isomorfos. Tais parâmetros são conhecidos como invariantes do grafo. Grafos isomorfos têm o mesmo espectro e, portanto, o espectro é um invariante do grafo [56]. Embora grafos isomorfos possuam o mesmo espectro, dois grafos com o mesmo espectro podem não ser isomorfos. Os grafos não-isomorfos que possuem o mesmo espectro são conhecidos na literatura como **grafos co-espectrais**.

Apesar da possibilidade de que grafos distintos sejam representados pelo mesmo espectro, estudos encontrados na literatura mostram que a ocorrência de grafos co-espectrais é pouco frequente [59,60]. De acordo com Cvetković [59] e Shokoufandeh [60] o espectro pode ser utilizado como representação de grafos uma vez que: (1) é capaz de representar a topologia do grafo, (2) os autovalores de um grafo são independentes da rotulação dos seus vértices, (3) é estável, uma vez que pequenas alterações no grafo produzirão pequenas alterações no seu espectro. Em Fonseca et al [50] encontramos uma avaliação experimental realizada em conjuntos de até 100.000 grafos gerados automaticamente que confirmam que as colisões entre grafos co-espectrais ocorre de forma pouco frequente.

2.4.4 Representação de Imagens por meio de Grafos

Grande parte das técnicas existentes para representação de imagens recorre as suas propriedades estatísticas. No entanto, a organização espacial de uma imagem também é uma característica importante utilizada para descrever seu conteúdo. De maneira geral, os relacionamentos espaciais extraídos das imagens são representados por árvores ou grafos. A seguir apresentamos duas das principais abordagens encontradas na literatura para se representar imagens por meio de grafos.

Grafo de adjacência de *pixels*

Nesta abordagem, cada *pixel* da imagem é associado a um vértice do grafo, e as arestas ligam os *pixels* vizinhos de acordo com alguma relação de vizinhança. As relações de vizinhança comumente utilizadas são: a vizinhança-4 e a vizinhança-8.

Dado um *pixel* p , a **vizinhança-4** para esse *pixel* corresponde aos seus dois vizinhos horizontais e aos seus dois vizinhos verticais. No caso do *pixel* p possuir as coordenadas (x, y) , então os *pixels* de coordenadas $(x + 1, y)$ e $(x - 1, y)$ serão os seus dois vizinhos horizontais. Já os *pixels* de coordenadas $(x, y + 1)$ e $(x, y - 1)$ serão os seus dois vizinhos verticais. A Figura 2.24(a) mostra a vizinhança-4 para o *pixel* p .

Para a **vizinhança-8**, além dos dois vizinhos horizontais e dos dois verticais, também são envolvidos os quatro vizinhos diagonais do *pixel*. Assim sendo, no caso de um *pixel* p possuir as coordenadas (x, y) então sua vizinhança-8 engloba todos os *pixels* indicados na

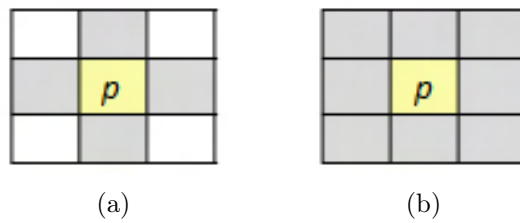


Figura 2.24: Um *pixel* p qualquer as suas relações de vizinhança: (a) Vizinhança-4; (b) Vizinhança-8.

vizinhança-4 mais os *pixels* com as seguintes coordenadas: $(x-1, y-1)$, $(x-1, y+1)$, $(x+1, y-1)$ e $(x+1, y+1)$. A Figura 2.24(b) mostra a vizinhança-8 para o *pixel* p .

Grafo de adjacência de regiões

Esta abordagem, permite descrever a estrutura espacial de uma imagem por meio das suas regiões previamente identificadas. Um grafo de adjacência de regiões (*Region Adjacency Graph* - RAG) é comumente utilizado para representar a disposição espacial das regiões de uma imagem. Nessa abordagem, cada vértice representa uma região da imagem e uma aresta entre dois vértices indica que as regiões correspondentes são vizinhas. Os grafos de adjacência de regiões fornecem uma visão espacial global da imagem, facilitando a execução de tarefas que envolvem a análise da disposição espacial das regiões na imagem. A Figura 2.25 ilustra um exemplo de uma imagem e sua respectiva representação em forma de grafo de adjacência de regiões.



Figura 2.25: Representação de uma imagem em forma de RAG [5]: (a) Imagem original; (b) Imagem segmentada; (c) RAG.

2.5 Trabalhos Relacionados

A cor é uma das características visuais de maior poder discriminante, e também uma das mais utilizadas em sistemas CBIR. Os Histogramas de Cor [35] e os Momentos de

Cor [61] são comumente utilizados como vetores de características para descrever o conteúdo visual das imagens. As duas abordagens descrevem a distribuição global das cores de uma imagem. Embora sejam fáceis de calcular, essas duas abordagens não são capazes de diferenciar imagens que possuem as mesmas cores mas que estejam espacialmente distribuídas de forma diferente pela imagem. A Figura 2.26 mostra um exemplo em que duas imagens distintas possuem o mesmo histograma. Para contornar esse problema, algumas técnicas foram propostas para incorporar informação espacial às cores das imagens [18–20].

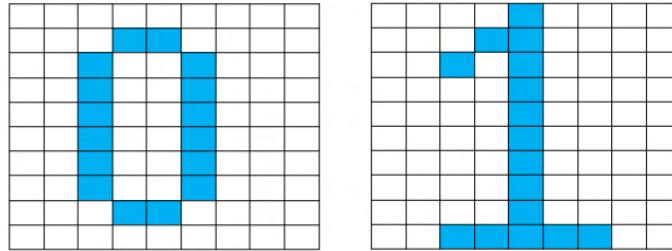


Figura 2.26: Duas imagens distintas que apresentam o mesmo histograma de cores.

O Correlograma de Cor, proposto por Huang et al. [19], expressa como a correlação espacial entre as cores varia com a distância, transmitindo a idéia de áreas maiores ou menores de uma determinada cor na imagem. O descritor *Auto Color Correlograms* (ACC) é uma versão simplificada dessa técnica que captura a correlação espacial somente entre cores idênticas. Um auto-correlograma indica a probabilidade de se encontrar na imagem dois *pixels* de cor C a uma distância d um do outro.

Chatzichristofis et al. [62] propuseram dois descritores de cor que combinam informações sobre cor e textura em um único histograma. As informações espaciais são obtidas dividindo-se a imagem em 1600 blocos regulares e extraíndo as características para cada um desses blocos. Para obter o descritor *Color and Edge Directivity Descriptor* (CEDD), as características de textura são representadas por um histograma de 6 posições que utiliza cinco filtros digitais como proposto no descritor MPEG-7 *Edge Histogram Descriptor* (EHD) [63]. Cada posição do histograma corresponde a uma forma de textura, sendo que a análise das texturas é realizada no espaço de cor YIQ. Já para as características de cor, o descritor CEDD utiliza um histograma de 24 posições produzido por um sistema *fuzzy* que caracteriza as cores Preto, Cinza, Branco, Vermelho, Laranja, Amarelo, Verde, Ciano, Azul e Magenta utilizando três predicados (claro, escuro e normal). A análise das cores é realizada no espaço de cor HSV, sendo que ao final do processo *fuzzy* o histograma de 24 posições representa o mapeamento das seguintes cores: (0) Branco, (1) Cinza, (2) Preto, (3) Vermelho claro, (4) Vermelho, (5) Vermelho escuro, (6) Laranja claro, (7) Laranja, (8) Laranja escuro, (9) Amarelo claro, (10) Amarelo, (11) Amarelo escuro, (12) Verde claro, (13) Verde, (14) Verde escuro, (15) Ciano claro, (16) Ciano, (17) Ciano escuro, (18) Azul claro, (19) Azul, (20) Azul escuro, (21) Magenta claro, (22) Magenta (23) Magenta escuro. O histograma final gerado pelo descritor CEDD possui 6 (textura) $\times$ 24 (cor)

= 144 posições. Assim, cada bloco da imagem é processado e atribuído a uma das 144 posições de acordo com as informações de cor e textura extraídas da região.

O descritor *Fuzzy Color Texture Histogram* (FCTH) também combina informações de cor e textura. As características de textura são representadas por um histograma de 8 posições produzido por um sistema *fuzzy* utilizando-se a transformada de Haar [64]. As características de cor são representadas por meio de um histograma de 24 posições obtido da mesma forma que o descritor CEDD. O histograma final gerado pelo descritor FCTH possui $8 \text{ (textura)} \times 24 \text{ (cor)} = 192$ posições. Assim, cada bloco da imagem é processado e atribuído a uma das 192 posições da mesma forma que o CEDD.

Além dos descritores de cor, existem diversos trabalhos que propõem utilizar descritores locais de forma para descrever o conteúdo visual das imagens. Os descritores *Scale Invariant Feature Transform* (SIFT) [16] e *Speeded-Up Robust Features* (SURF) [17] são bons exemplos dessa abordagem. Ambos são capazes de detectar e extrair descritores locais que são invariantes a mudanças de iluminação, rotação, escala e pequenas mudanças de perspectiva. No entanto, essas técnicas produzem uma grande quantidade de descritores, além de serem computacionalmente caras. Quando aplicadas em sistemas CBIR, seus descritores são comumente quantizados utilizando métodos de agrupamento como proposto em [65, 66].

Outra maneira de descrever o conteúdo visual das imagens é por meio dos relacionamentos espaciais entre os seus componentes. Os relacionamentos espaciais podem ser classificados em relacionamentos direcionais ou topológicos. Os relacionamentos direcionais comumente utilizados são: Norte, Sul, Leste, Oeste, Nordeste, Noroeste, Sudeste e Sudoeste. Em [6] Egenhofer apresenta o seguinte conjunto de relacionamentos topológicos entre duas regiões planas: Disjunto, Adjacente, Sobreposto, Incluso/Contido, Cobre/Coberto e Igual. A Figura 2.27 ilustra o conjunto de relacionamentos proposto por Egenhofer. Baseado nesse conjunto de relacionamentos topológicos, outros autores criaram seus próprios subconjuntos para descrever a organização espacial entre os componentes de uma imagem.



Figura 2.27: Relacionamentos topológicos definidos por Egenhofer [6].

Fonseca et al. [50] agruparam os relacionamentos topológicos de Egenhofer por similaridade e criaram um subconjunto de apenas três relacionamentos: Inclusão, Adjacência e Disjunto. Os autores utilizaram esse subconjunto de relações (veja a Figura 2.28) para descrever desenhos vetoriais. As informações topológicas extraídas dos desenhos vetoriais são representadas utilizando um grafo de topologia, que é posteriormente convertido em

um descritor utilizando a teoria espectral de grafos [59].

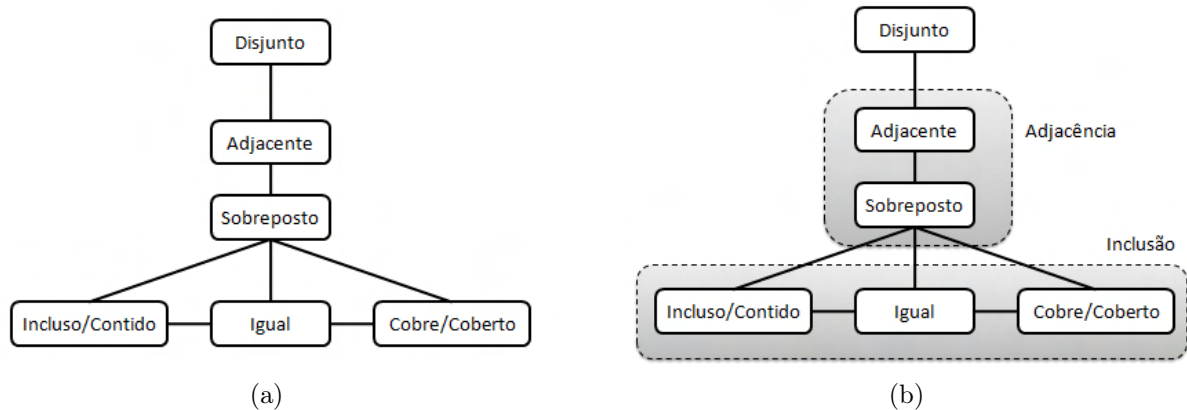


Figura 2.28: Relacionamentos topológicos originalmente definidos por Egenhofer (a) e a versão simplificada proposta por Fonseca et al. (b).

Leung e Chen [51] reduziram ainda mais o conjunto de relacionamentos topológicos proposto por Egenhofer. Os autores utilizaram apenas a relação de inclusão para descrever a organização espacial de desenhos feitos a mão. Nesse trabalho, os autores também utilizaram um grafo para representar as conexões entre os elementos dos desenhos.

Liang et al. [67] desenvolveram uma solução para a recuperação de rascunhos de imagens, baseada no trabalho de Fonseca et al., utilizando oito tipos de relacionamentos topológicos. Para descrever a organização espacial dos rascunhos, os autores criaram oito grafos topológicos, um para cada tipo de relacionamento.

Hlaoui e Wang [52] apresentaram uma abordagem para recuperação de imagens que descreve o relacionamento entre regiões resultantes de um processo de segmentação. Os autores utilizam a distância e a posição relativa entre as regiões como relacionamento espacial. Essas informações espaciais são armazenadas em grafos relacionais com atributos, ou *Attributed Relation Graph* (ARG). Para comparar os grafos, os autores utilizam a distância de edição entre eles.

2.6 Considerações Finais

Neste capítulo foram apresentados os fundamentos teóricos necessários para o entendimento deste trabalho. Em cada um dos tópicos apresentados, abordamos os conceitos fundamentais que serão úteis no decorrer desta dissertação.

Também discutimos ao longo desse capítulo, uma visão geral do funcionamento de um sistema de recuperação de imagens por conteúdo. O processo de recuperação de imagens por conteúdo foi apresentado detalhando-se as principais fases do seu fluxo de funcionamento: a extração de características, as medidas de similaridade e as métricas de avaliação comumente utilizadas. Além disso, esse capítulo apresentou alguns dos principais trabalhos relacionados.

Os trabalhos relacionados encontrados e descritos nesse capítulo utilizam apenas cor ou apenas topologia, ou quando combinam topologia e cor, são utilizados dois descritores distintos, fazendo com que o processo de recuperação seja realizado em dois passos. Além disso, eles não suportam a recuperação parcial de imagens adequadamente.

Nesse trabalho, nós propomos uma abordagem que combina características de cor e de topologia em um único descritor. Nosso novo descritor utiliza um grafo topológico para representar o relacionamento espacial e a similaridade de cor entre as regiões da imagem. Para evitar a comparação direta entre grafos, ao realizar o cálculo da similaridade entre imagens, nós utilizamos a teoria espectral dos grafos. Dessa forma, reduzimos o problema da correspondência entre grafos ao cálculo de distância entre vetores de características.

Dado que a cor é uma das características visuais de maior poder discriminativo, nós também propomos um novo descritor baseado apenas em cor para a recuperação parcial de imagens. Nosso descritor de cor, após um processo de simplificação da imagem original, identifica as manchas de cor que ocorrem na imagem e as utiliza como descritor. A partir das manchas identificadas, utilizamos um algoritmo de correspondência para identificar as imagens que contêm as manchas de cor mais similares.

Conceitos relevantes necessários para o entendimento da nossa abordagem também foram apresentados: imagens digitais, modelos de cor, quantização de cores e teoria espectral de grafos são alguns exemplos. Os próximos capítulos apresentam os dois novos descritores que propomos, as bases de imagens criadas para a validação e as avaliações experimentais realizadas.

Capítulo 3

Color Blobs Descriptor: Um novo descritor de cor para pesquisa parcial

Dado que a cor é uma das características visuais de maior poder discriminativo, um novo descritor baseado apenas em cor foi desenvolvido. Esse novo descritor de cor, denominado *Color Blobs Descriptor* (CBD), é obtido após uma sequência de passos para obter uma versão simplificada das imagens. O processo de simplificação remove detalhes desnecessários das imagens enquanto preserva as informações relevantes da cena. A partir da versão simplificada da imagem, uma abordagem local é utilizada para a extração de características e um algoritmo de correspondência que permite identificar objetos que estejam contidos em outras imagens é aplicado.

3.1 Simplificação da Imagem

O cálculo do descritor CBD inicia com um processo de simplificação das imagens. Esse processo de simplificação é realizado em três etapas:

1. **A imagem original é borrada:** Inicialmente, a imagem original é borrada por meio de um filtro Gaussiano com desvio padrão igual a 5, para eliminar ruídos que por ventura estejam presentes na imagem. O valor utilizado para desvio padrão foi obtido empiricamente.
2. **A imagem borrada é quantizada:** A imagem borrada é então submetida a um processo de quantização no espaço de cores RGB. Nessa etapa as imagens são quantizadas em 216 cores (veja a Seção 2.1.2).
3. **A imagem quantizada é segmentada:** Por último, a imagem quantizada é segmentada utilizando a abordagem *Mean Shift* [1]. Para o algoritmo *Mean Shift* utilizamos o valor para o parâmetro *spatial radius* = 7 e para o parâmetro *color radius* = 6.5, valores esses determinados empiricamente.

A imagem resultante do processo de segmentação é então utilizada como versão simplificada da imagem. A Figura 3.1 ilustra os três passos realizados para obter a simplificação de uma imagem.

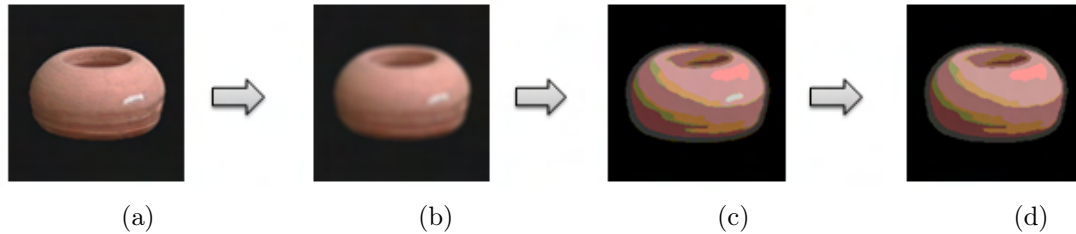


Figura 3.1: Processo de simplificação da imagem: (a) Imagem original; (b) Imagem borrada; (c) Imagem quantizada (216 cores); (d) Imagem simplificada.

O descritor CBD é então definido como o conjunto das manchas de cor que ocorrem na imagem simplificada. Esse conjunto é obtido percorrendo a imagem e identificando os valores RGB distintos que ocorrem na imagem. A etapa de simplificação é importante para que a nossa abordagem seja capaz de diferenciar dois objetos mesmo sobre pequenas variações de cores. Isso porque dificilmente dois objetos distintos terão exatamente o mesmo conjunto de cores após o processo de simplificação. A Figura 3.2 mostra um exemplo.

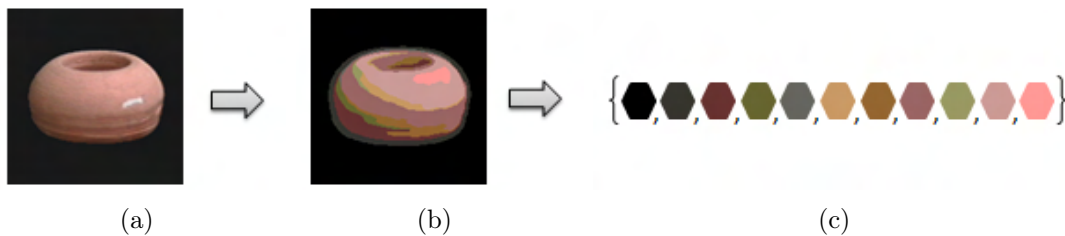


Figura 3.2: Cálculo do descritor CBD para uma imagem de consulta: (a) Imagem original; (b) Imagem simplificada; (c) CBD.

3.2 Medida de Similaridade

Para realizar uma consulta por similaridade, os descritores CBD calculados para a imagem de consulta e para as imagens da base devem ser comparados utilizando-se uma medida de similaridade. Entretanto, o CBD extraído para as imagens podem não ter a mesma dimensão, o que é um pré-requisito para a maioria das medidas de similaridade. Pensando nisso, é realizada uma verificação para identificar se as manchas de cor que ocorrem na imagem de consulta, também ocorrem na imagem que está sendo comparada. Essa verificação é realizada calculando-se a distância Euclidiana (d) entre os valores RGB de cada mancha de cor da imagem de consulta e os valores RGB de todas as manchas de cor da imagem a ser comparada. A distância final (D_f) é obtida somando-se as menores

distâncias (d) obtidas para cada mancha de cor da imagem de consulta. Essa medida de similaridade é formalmente definida pela Equação 3.1:

$$D_f = \sum_{i=1}^n \min_{j=1}^m (d(c_i, c_j)) \quad (3.1)$$

onde n e m representam a quantidade de manchas de cor encontradas na imagem de consulta e na imagem a ser comparada, respectivamente.

Tome como exemplo os descritores CBD extraídos para uma imagem de consulta e para outra imagem a ser comparada como mostra as Figuras 3.2 e 3.3, respectivamente. Note que todas as manchas de cor que ocorrem na imagem de consulta também ocorrem na imagem a ser comparada. Nesse caso, a medida de similaridade D_f entre elas será zero. Isso significa que a imagem de consulta está contida na imagem a ser comparada.

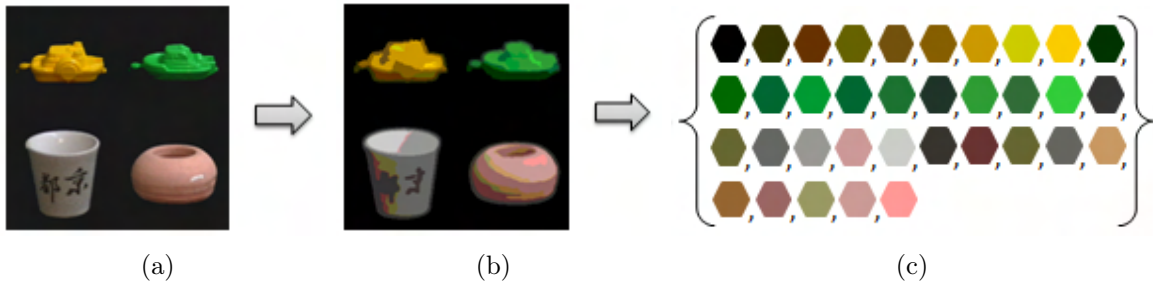


Figura 3.3: Cálculo do descritor CBD para uma imagem a ser comparada: (a) Imagem original; (b) Imagem simplificada; (c) CBD.

A medida de similaridade apresentada obtém bons resultados devido às duas principais características dessa abordagem: Primeiro, a descrição das cores dos objetos como partes individuais da imagem; e segundo, a utilização do processo de simplificação que reduz o número de cores possíveis e fornece um conjunto de cores personalizadas para cada objeto. É importante notar que a abordagem que propomos simplesmente verifica se as cores identificadas na imagem de consulta também ocorrem na imagem a ser analisada. O algoritmo de similaridade proposto juntamente com o processo de simplificação torna o CBD bastante robusto para a pesquisa parcial em cenas reais, onde diferentes objetos dificilmente serão representados com exatamente as mesmas manchas de cor, ainda que isso seja possível. Entretanto, a abordagem CBD pode não obter resultados tão bons em imagens criadas de forma artificial, onde um mesmo objeto pode ser representado com cores diferentes ou até mesmo objetos completamente diferentes contendo exatamente as mesmas cores. Uma avaliação experimental utilizando bases contendo tanto imagens com objetos de cenas reais quanto objetos criados de forma artificial foi realizada no Capítulo 5.

3.3 Considerações Finais

Nesse capítulo apresentamos um novo descritor de cor, próprio para a recuperação parcial de imagens, denominado *Color Blobs Descriptor* (CBD). O descritor apresentado identifica um conjunto de manchas de cor que ocorrem na imagem após um processo de simplificação. Para recuperar imagens que contêm uma determinada imagem fornecida como consulta, as manchas de cor identificadas na imagem de consulta são comparadas com as manchas de cor identificadas na imagem que está sendo comparada por meio de uma medida de similaridade proposta.

Uma avaliação experimental do descritor CBD foi realizada no Capítulo 5. Embora tenha obtido bons resultados, o CBD não leva em consideração a topologia dos objetos presentes na imagem. Sendo assim, objetos diferentes mas com os mesmos conjuntos de cores serão erroneamente considerados similares. Para entender a importância da topologia dos objetos, realizamos um estudo onde combinamos informações topológicas e de cor para descrever as características das imagens. Os resultados desse estudo demonstraram que a associação entre descritores topológicos e de cor melhora significativamente os resultados obtidos por cada descritor utilizado isoladamente. A partir desse estudo, propomos um novo descritor capaz de associar características topológicas e de cor, que será apresentado no próximo capítulo.

Capítulo 4

Color Spectrum: Combinando cor e topologia em um único descritor para pesquisa parcial

Embora a cor seja uma das propriedades visuais de maior poder discriminativo, por vezes, a cor por si só não é suficiente para descrever o conteúdo visual das imagens. A organização espacial das diferentes regiões de cor também desempenham um papel importante. Neste capítulo, propomos e avaliamos um novo descritor que combina informações sobre cor e sobre a sua organização espacial na imagem. Além disso, o mecanismo utilizado para calcular o descritor fornece suporte para a recuperação parcial de imagens. O novo descritor proposto, denominado *Color Spectrum*, primeiramente realiza um processo de simplificação da imagem original, onde são identificadas as regiões de cor que compõem a imagem da mesma forma como é feito para obter o CBD. Após esse processo de simplificação, a organização espacial das regiões de cor previamente identificadas é representada por meio de um grafo topológico, onde os vértices representam as regiões de cor e as arestas representam as conexões entre essas regiões. Para calcular o descritor a partir dessa representação em forma de grafo, a teoria espectral de grafos [21] é utilizada, evitando a necessidade de realizar uma comparação direta entre grafos. Para suportar a pesquisa parcial é realizada uma decomposição do grafo principal em diversos subgrafos, e também são calculados descritores para esses subgrafos.

4.1 Simplificação da Imagem

Para calcular o descritor topológico, é necessário a obtenção de um grafo de topologia para representar o relacionamento espacial entre as regiões de cor que compõem a imagem. Para isso, primeiramente é realizado um processo de simplificação da imagem original para que seja possível extrair o relacionamento espacial entre essas regiões. Esse

processo de simplificação remove detalhes desnecessários das imagens enquanto preserva as informações topológicas relevantes para a abordagem proposta.

O processo de simplificação é realizado em três passos, da mesma forma como foi feito para obter o CBD (veja a Seção 3.1): (i) A imagem original é borrada; (ii) A imagem borrada é quantizada; (iii) A imagem quantizada é segmentada. A única diferença com relação à simplificação apresentada na Seção 3.1, é o processo de quantização realizado no passo (ii). Nessa etapa as imagens são quantizadas em 16 cores, enquanto que na quantização realizada para obter o CBD as imagens são quantizadas em 216 cores.

A imagem resultante do processo de segmentação é então utilizada como a versão simplificada da imagem. Essa representação simplificada da imagem original, permite extrair as regiões de cor presentes na imagem e o relacionamento espacial entre elas. A Figura 4.1 ilustra os passos realizados para a obtenção da representação simplificada da imagem.

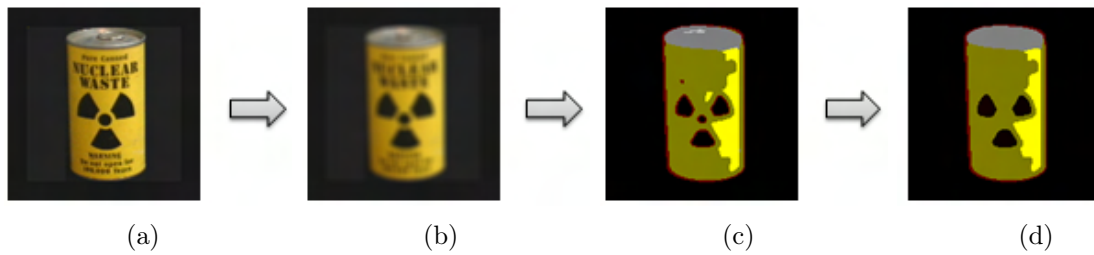


Figura 4.1: Processo de simplificação da imagem: (a) Imagem original; (b) Imagem borrada; (c) Imagem quantizada (16 cores); (d) Imagem simplificada.

4.2 Grafo Topológico

O grafo de topologia e, consequentemente, os descritores topológicos utilizados são baseados na abordagem proposta por Fonseca et al. [50,68]. Entretanto, em Fonseca et al. o domínio da aplicação é restrito a desenhos vetoriais (cliparts e desenhos técnicos), onde as regiões da imagem são facilmente identificadas, enquanto nossa proposta aplica-se também a imagens *raster*.

Como visto na Seção 2.4, definimos um grafo pelo par $G = (V, E)$, onde V representa o conjunto de vértices e $E \subseteq V \times V$ representada o conjunto de arestas. A partir da representação simplificada da imagem, um grafo topológico é criado, onde cada vértice representa uma região de cor identificada na imagem e uma aresta entre dois vértices indica o relacionamento espacial entre essas regiões. Os vértices do grafo armazenam as características de cor de cada região associada (valores RGB). Para expressar a organização espacial das regiões de cor, utilizamos o mesmo conjunto de relacionamentos espaciais utilizado por Fonseca et al.: Disjunção, Inclusão e Adjacência. Esses relacionamentos são ilustrados na Figura 4.2 e formalmente definidos a seguir.

Definição 4.1. (Relação de Disjunção) Se não há interseções entre duas regiões $R1$ e $R2$, essas regiões são consideradas **Disjuntas**.

Definição 4.2. (Relação de Inclusão) Se uma região $R1$ contém completamente uma outra região $R2$, ou $R1$ é igual a $R2$, então a região $R1$ **inclui** a região $R2$.

Definição 4.3. (Relação de Adjacência) Se duas regiões se encontram ou se existe uma interseção entre elas, essas regiões são consideradas **Adjacentes**.

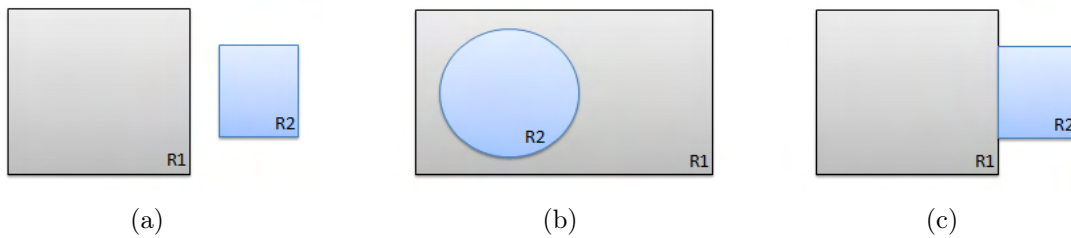


Figura 4.2: Relações topológicas entre regiões da imagem: (a) Disjuntos; (b) Inclusão; (c) Adjacência.

Os relacionamentos espaciais extraídos a partir da imagem simplificada são então representados em um grafo topológico, onde **arestas verticais** representam a relação de inclusão e **arestas horizontais** representam a relação de adjacência. Além disso, dois vértices não conectados por uma aresta indica que esses vértices são disjuntos. A Figura 4.3 ilustra uma imagem simplificada e a sua respectiva representação em forma de grafo topológico. Para facilitar o entendimento, utilizaremos uma imagem fictícia, onde os relacionamentos podem ser facilmente identificados, para ilustrar todo o processo de extração dos descritores topológicos. A título de curiosidade, o grafo topológico extraído a partir da Figura 4.1 é ilustrado na Figura 4.4.

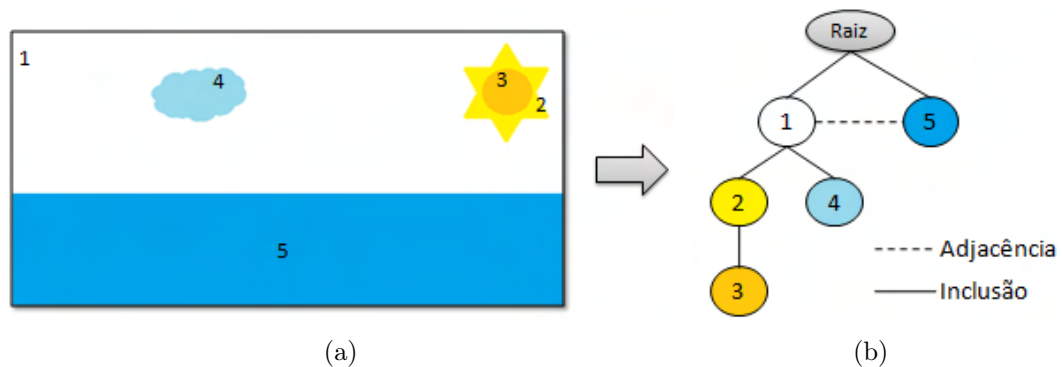


Figura 4.3: Imagem simplificada (a) e seu respectivo grafo topológico (b).

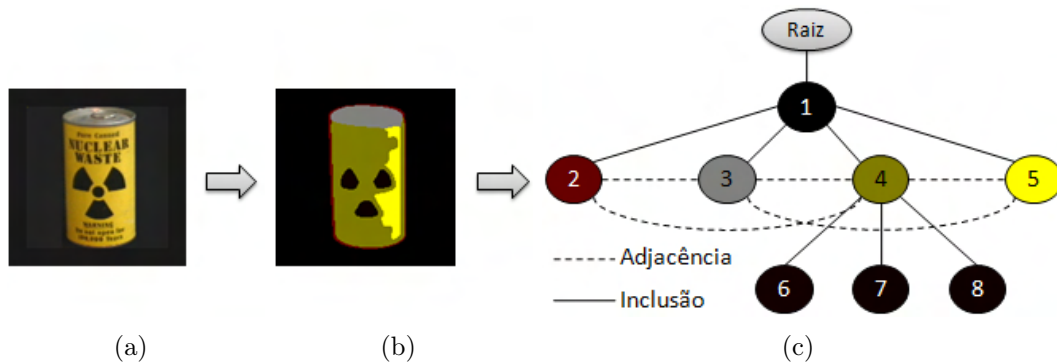


Figura 4.4: Extração do grafo topológico da Figura 4.1: (a) Imagem original; (b) Imagem simplificada; (c) Grafo topológico.

Como podemos ver na Figura 4.3, o grafo topológico utilizado possui uma estrutura bem definida sendo similar a uma árvore. Essa estrutura de dados sempre possui um vértice raiz, representando toda a imagem. Os vértices filhos do vértice raiz representam as regiões de cor dominantes da imagem, isto é, as regiões que não estão incluídas em nenhuma outra região. O próximo nível do grafo representa as regiões que estão incluídas nas regiões representadas no nível anterior. Esse processo é aplicado recursivamente até completarmos toda a hierarquia das regiões de cor.

4.3 Medida de Similaridade

Dado que o cálculo da correspondência exata entre grafos é um problema NP-completo [69], utilizamos a teoria espectral dos grafos [21, 59, 70] para obter a similaridade entre grafos. Dessa forma, evitamos a necessidade de realizar uma comparação direta entre grafos, reduzindo um problema NP-completo ao cálculo de similaridade entre vetores multidimensionais. Para obter o espectro de um grafo, primeiramente é necessário obter a sua representação em forma de matriz. A partir daí, os autovalores dessa matriz são calculados e organizados em ordem decrescente. O vetor formado por esses autovalores ordenados de forma decrescente é chamado de espectro do grafo. Existe uma grande variedade de formas de representar grafos utilizando matrizes, como por exemplo a matriz de adjacência ou a matriz Laplaciana (veja a Seção 2.4). Nos experimentos realizados nesse capítulo, utilizamos a matriz de adjacência e seu respectivo espectro de adjacência como descritor topológico em uma das abordagens avaliadas. A Figura 4.5 ilustra o processo realizado para se obter o espectro de adjacência de um determinado grafo.

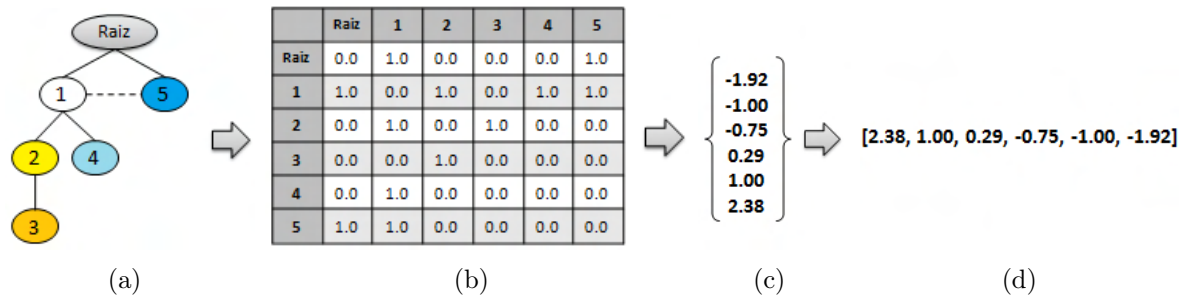


Figura 4.5: Cálculo do espectro de adjacência para a imagem da Figura 4.3. (a) Grafo topológico; (b) Matriz de adjacência; (c) Autovalores; (d) Espectro de adjacência.

Os descritores topológicos resultantes são vetores multidimensionais, cuja dimensão depende da complexidade do grafo. Imagens muito complexas irão produzir descritores de alta dimensionalidade, enquanto que imagens simples irão produzir descritores de baixa dimensionalidade. Isso se deve ao fato de que a dimensionalidade do espectro coincide com a quantidade de vértices do grafo [70]. A similaridade topológica entre duas imagens é obtida calculando-se a distância Euclidiana entre seus descritores topológicos. Entretanto, como os descritores topológicos da imagem de consulta e da imagem a ser analisada podem não ter a mesma dimensionalidade, o descritor de menor dimensionalidade é preenchido com zeros respeitando a ordenação dos elementos do descritor. Isso equivale a adicionar vértices não conectados ao menor grafo de forma que os dois grafos que serão comparados passem a ter a mesma cardinalidade [70]. Como os valores do espectro são organizados em ordem decrescente, essa estratégia de preencher com zeros o menor espectro não corrompe a representação da imagem.

Tome como exemplo o descritor extraído para uma imagem de consulta conforme mostra a Figura 4.6. Vamos comparar essa imagem com a imagem ilustrada na Figura 4.3, cujo descritor está expresso na Figura 4.5(d).

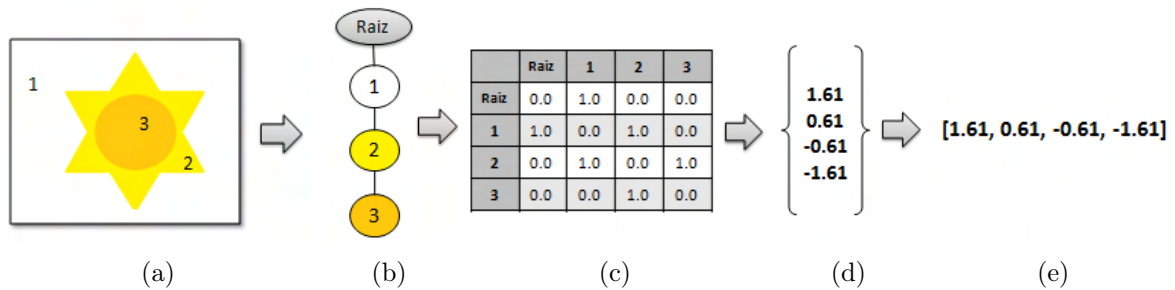


Figura 4.6: Cálculo do espectro de adjacência para uma imagem de consulta. (a) Imagem simplificada; (b) Grafo topológico; (c) Matriz de adjacência; (d) Autovalores; (e) Espectro de adjacência.

Para realizar a comparação entre as imagens da Figuras 4.6(a) e 4.3(a), como os descritores não possuem a mesma dimensionalidade, o descritor da Figura 4.6(e) será completado com zeros. Assim, a similaridade entre eles será obtida calculando-se a dis-

tância Euclidiana entre os descritores $\bar{d}_1 = [1.61, 0.61, \mathbf{0.00}, \mathbf{0.00}, -0.61, -1.61]$ e $\bar{d}_2 = [2.38, 1.00, 0.29, -0.75, -1.00, -1.92]$.

4.4 Descritor *Color Spectrum*

Para combinar informações espaciais e de cor em um mesmo descritor, informações sobre as cores das regiões são acrescentadas ao grafo topológico. Para isso, são adicionados pesos as arestas do grafo topológico com a similaridade de cor entre duas regiões conectadas. Esse novo grafo topológico permite que a solução proposta seja capaz de diferenciar, por exemplo, um objeto que contenha uma região amarela dentro de uma região verde, de um objeto que contenha uma região verde dentro de uma região amarela. A Figura 4.7 mostra um diagrama de blocos com os passos necessários para calcular o descritor *Color Spectrum*.

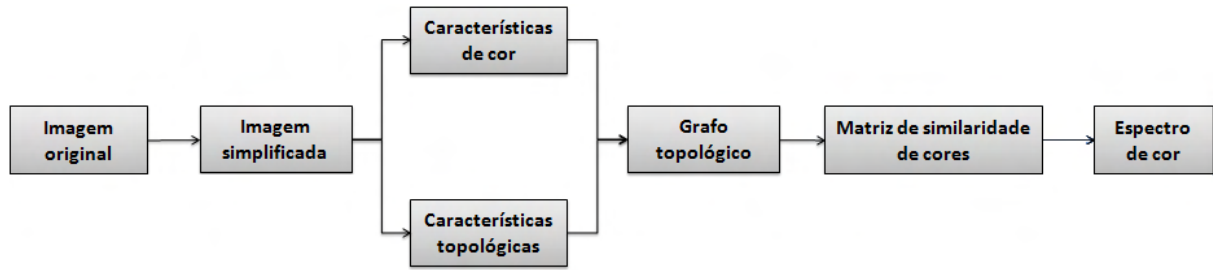


Figura 4.7: Diagrama de blocos para a obtenção do descritor *Color Spectrum*.

Após o processo de simplificação, descrito anteriormente, um grafo topológico é criado a partir da imagem simplificada, como mostramos na seção anterior. As características de cor de cada região (os valores RGB) são extraídas e armazenadas como atributos dos vértices do grafo. A partir desse grafo topológico, com as características de cor armazenadas nos vértices, uma nova representação matricial $\mathbf{M}$ é calculada. Essa matriz representa a similaridade de cor entre os vértices do grafo, sendo formalmente definida pela Equação 4.1,

$$M_{uv} = \begin{cases} r & \text{se } u \text{ ou } v \text{ é o vértice raiz e } u \neq v \\ r - d(u, v) & \text{se } u \text{ e } v \text{ são vértices conectados} \\ 0 & \text{caso contrário} \end{cases} \quad (4.1)$$

onde $r = \sqrt[2]{\sum_{i=1}^3 255^2}$, u e v são vértices do grafo, e $d(u, v)$ é a distância Euclidiana entre os valores RGB associados aos vértices u e v , respectivamente. Como o vértice raiz representa toda a imagem, ele não possui informação de cor associada a ele. Assim, os vértices que estão ligados ao vértice raiz por uma aresta possuem o valor de similaridade máxima (r). Todos os outros vértices conectados possuem similaridade de cor inversamente proporcional a distância Euclidiana entre os valores RGB das suas regiões conectadas. Além

disso, vértices não conectados possuem o valor de similaridade mínima (0).

Finalmente, os autovalores são calculados a partir da representação matricial $\mathbf{M}$ e ordenados em ordem decrescente, resultando no descritor *Color Spectrum*. A Figura 4.8 mostra um exemplo com os passos necessários para calcular o *Color Spectrum* para a imagem da Figura 4.3.

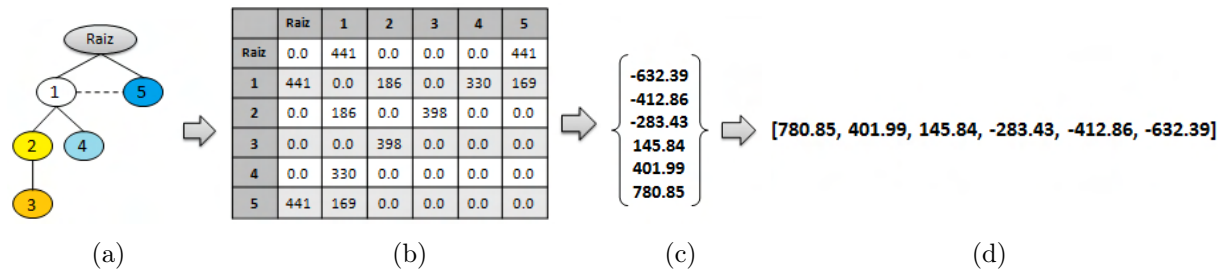


Figura 4.8: Cálculo do *Color Spectrum* para a imagem da Figura 4.3. (a) Grafo topológico; (b) Matriz de similaridade de cor $\mathbf{M}$; (c) Autovalores; (d) *Color Spectrum*.

4.5 Decomposição em Subgrafos

Para que a solução proposta seja capaz de suportar a pesquisa parcial de imagens, ao invés de calcular um único descritor para todo o grafo topológico, é realizada uma decomposição do grafo principal em diversos subgrafos, e também são calculados descritores para esses subgrafos. A partir do grafo principal são identificados os níveis de detalhe da imagem descendo pelas arestas verticais, onde a quantidade de arestas verticais percorridas para se chegar em um determinado vértice indica o nível de detalhe. Para cada nível de detalhe identificado, são extraídos os seus ramos principais (subgrafos), sendo que cada ramo representa um objeto distinto presente na imagem. Para cada um desses subgrafos encontrados o descritor *Color Spectrum* é calculado. Ao final do processo, é extraído um descritor por nível de detalhe para cada objeto distinto presente na imagem. A Figura 4.9 ilustra os subgrafos extraídos a partir da imagem da Figura 4.3 e sua parte correspondente na imagem.

Após a aplicação dessa descrição multinível, é obtida uma coleção de descritores topológicos para diversas partes da imagem, como mostra a Figura 4.9. Dessa forma, é possível recuperar imagens fornecendo como consulta toda a imagem ou qualquer uma das suas subpartes. Para calcular a similaridade entre duas imagens, um único descritor é calculado para a imagem fornecida como consulta. O descritor obtido para a imagem de consulta é então comparado com todos os descritores armazenados na base. O cálculo de similaridade entre esses descritores é realizado utilizando-se a distância Euclidiana, como visto na Seção 4.3.

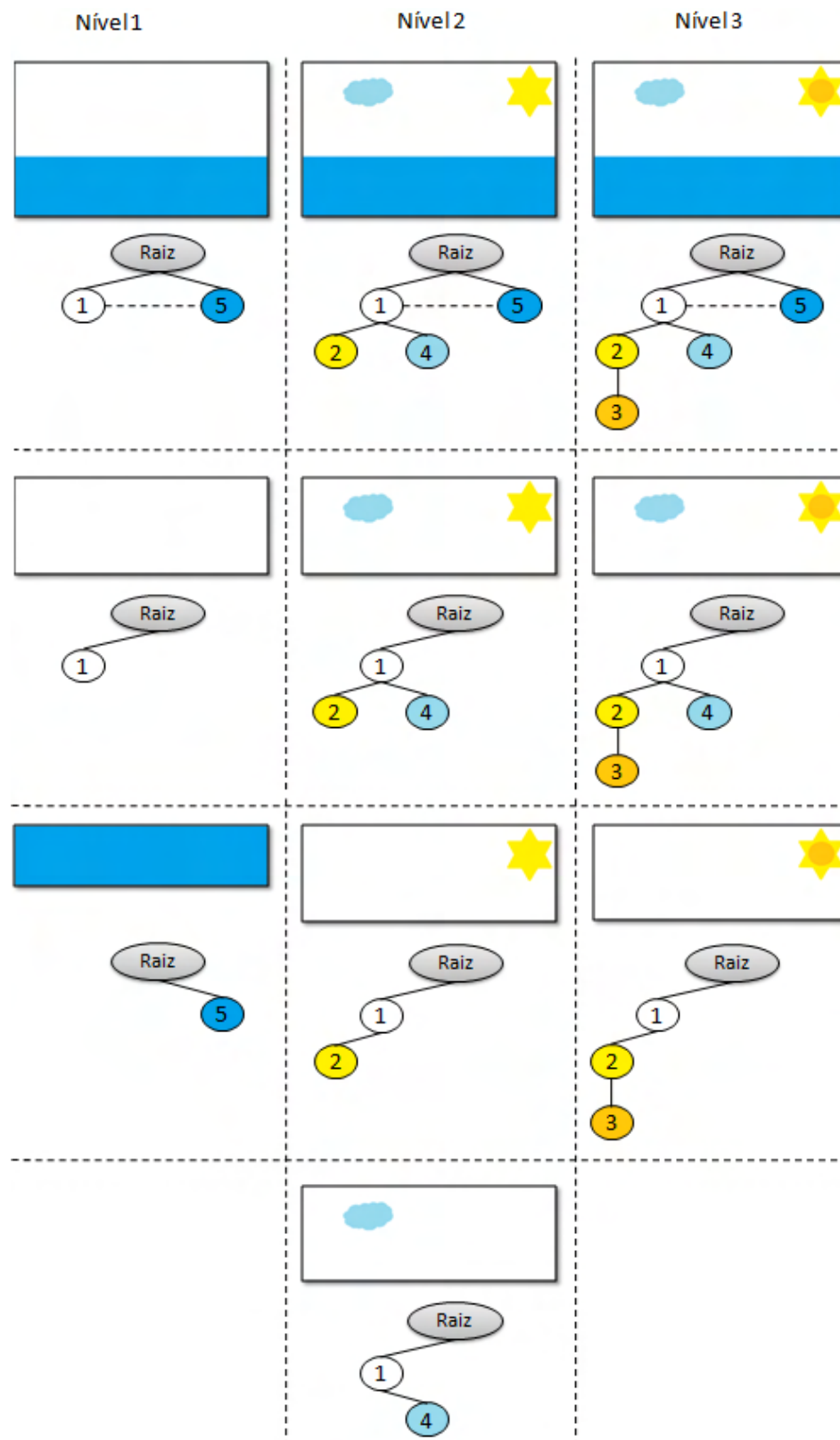


Figura 4.9: Subgrafos extraídos e suas partes correspondentes na imagem.

4.6 Considerações Finais

Nesse capítulo apresentamos um novo descritor para a recuperação parcial de imagens, chamado *Color Spectrum*. Esse novo descritor adiciona características de cor em um grafo de topologia para descrever o conteúdo das imagens. Dessa forma, combinamos em um único descritor informações sobre cor e sobre a organização espacial das imagens. Além disso, também apresentamos um mecanismo de descrição multinível utilizado para calcular diversos descritores a partir do grafo topológico, tornando possível a descrição de várias subpartes da imagem, e assim possibilitando a recuperação parcial.

Para evitar a comparação direta entre grafos, uma vez que este é um problema NP-completo, utilizamos a teoria espectral de grafos para converter os grafos topológicos em descritores. Dessa forma, reduzimos um problema NP-completo ao cálculo de distâncias entre vetores multidimensionais. Embora esse mapeamento de grafos para descritores não seja único, uma vez que diferentes grafos podem ter o mesmo descritor, estudos na literatura mostram que essas colisões ocorrem de forma pouco frequente [60].

O próximo capítulo apresenta uma avaliação experimental dos descritores propostos (CBD e *Color Spectrum*) utilizando diferentes bases de imagens.

Capítulo 5

Experimentos

Neste capítulo apresentamos as avaliações experimentais para os descritores propostos e detalhados nos capítulos anteriores. Primeiramente, apresentamos as três bases de imagens utilizadas nos experimentos. Em seguida apresentamos a avaliação dos resultados experimentais e as considerações finais do capítulo.

5.1 Bases de Imagens

Para avaliar a qualidade no processo de recuperação parcial de imagens por conteúdo, foram criadas três bases de imagens distintas. À estas bases demos os nomes de: *Partial-COIL*, *Partial-COIL-T* e *Partial-Mock*. Essas bases foram criadas com a intenção de avaliar a capacidade de encontrar objetos que estejam contidos em outras imagens. Para isso, todas as três bases foram formadas a partir de um conjunto de imagens que contêm um ou mais objetos na mesma cena. Duas das três bases foram criadas a partir de imagens de objetos reais. A terceira base foi criada a partir de imagens de objetos fictícios criados de forma artificial por meio de editores de imagem. As seções a seguir descrevem as três bases de imagens criadas em maiores detalhes.

5.1.1 *Partial-COIL*

A base de imagens *Partial-COIL* é formada por imagens contendo uma composição de até quatro objetos originários da base *Columbia University Image Library* (COIL-100) [7]. A COIL-100 é uma base de imagens bastante utilizada na literatura, sendo composta por imagens de 100 objetos rotacionados de 5 em 5 graus, totalizando 720 imagens (de 0° a 355°) para cada um dos 100 objetos. A Figura 5.1 mostra os objetos originais utilizados para a construção da base COIL-100.



Figura 5.1: Objetos utilizados para criar a base de imagens COIL-100 [7].

Para construir a base de imagens *Partial-COIL*, primeiramente foram escolhidos de forma aleatória 30 dos 100 objetos presentes na base COIL-100. As imagens de ângulo 0° para os objetos escolhidos serão posteriormente utilizadas para criar as imagens que irão compor a base. Além disso, essas imagens serão utilizadas como consulta esperando-se que todas as imagens em que o objeto aparece sejam consideradas como relevantes no processo de recuperação. A Figura 5.2 mostra os 30 objetos selecionados aleatoriamente.



Figura 5.2: Os 30 objetos aleatoriamente selecionados da base COIL-100.

Esse processo resultou em uma base com um total de 339 imagens de tamanho 320×320 *pixels*. A Figura 5.3 mostra algumas das imagens criadas para compor a base *Partial-COIL*. A base completa está disponível para *download* no seguinte endereço: http://www.4shared.com/zip/xlp9_a_4/Partial-COIL-30.html.



Figura 5.3: Exemplos de imagens da base *Partial-COIL*.

A base de imagens *Partial-COIL-T* foi criada seguindo o mesmo processo utilizado para a criação da base *Partial-COIL*. A diferença é que para essa base todos os 100 objetos da base COIL-100 foram utilizados, e não somente 30. Seguindo o processo descrito na Seção 5.1.1, foram criadas 768 imagens não relevantes que adicionadas às imagens da base *Partial-COIL* formaram a base *Partial-COIL-T*. Essas 768 imagens foram criadas a partir

A base *Partial-COIL-T* foi criada para avaliar a consistência das abordagens em bases com grande quantidade de imagens não relevantes (lixos). A Figura 5.4 mostra algumas das imagens criadas para compor a base *Partial-COIL-T*. A base completa está disponível para *download* no seguinte endereço: <http://www.4shared.com/zip/obSdxWDV/Partial-COIL-30-T.html>.



Figura 5.4: Exemplos de imagens da base *Partial-COIL-T*.

5.1.3 *Partial-Mock*

Para validar a capacidade discriminativa das abordagens com relação às características topológicas dos objetos e suas respectivas cores, foi criada uma outra base contendo o mesmo tipo de imagens descritas nas Seções 5.1.1 e 5.1.2. A diferença desta base para as bases anteriores é que foram utilizados objetos artificiais para criar as imagens, enquanto

que para a criação das outras bases foram utilizados objetos reais. Dessa forma, os objetos presentes nas imagens da base *Partial-Mock* não sofrem influência das condições de iluminação nem das pequenas transições de cor e textura. Além disso, as características topológicas e de cor para este tipo de objeto são facilmente reconhecidas.

Para construir a base *Partial-Mock*, primeiramente foram criados 30 objetos fictícios. A Figura 5.5 mostra os 30 objetos fictícios criados. A partir desses 30 objetos fictícios, foi realizada uma combinação entre eles de forma que cada objeto estivesse presente em exatamente 30 imagens diferentes. Essa combinação foi realizada seguindo o mesmo procedimento utilizado para a criação da base *Partial-COIL*. Este processo resultou em uma base contendo 314 imagens de tamanho 400×400 pixels. A Figura 5.6 mostra exemplos de imagens da base *Partial-Mock*. A base completa está disponível para *download* no seguinte endereço: http://www.4shared.com/zip/YM1fyXz_/Partial-Mock.html.



Figura 5.5: Os 30 objetos criados para construir a base *Partial-Mock*.

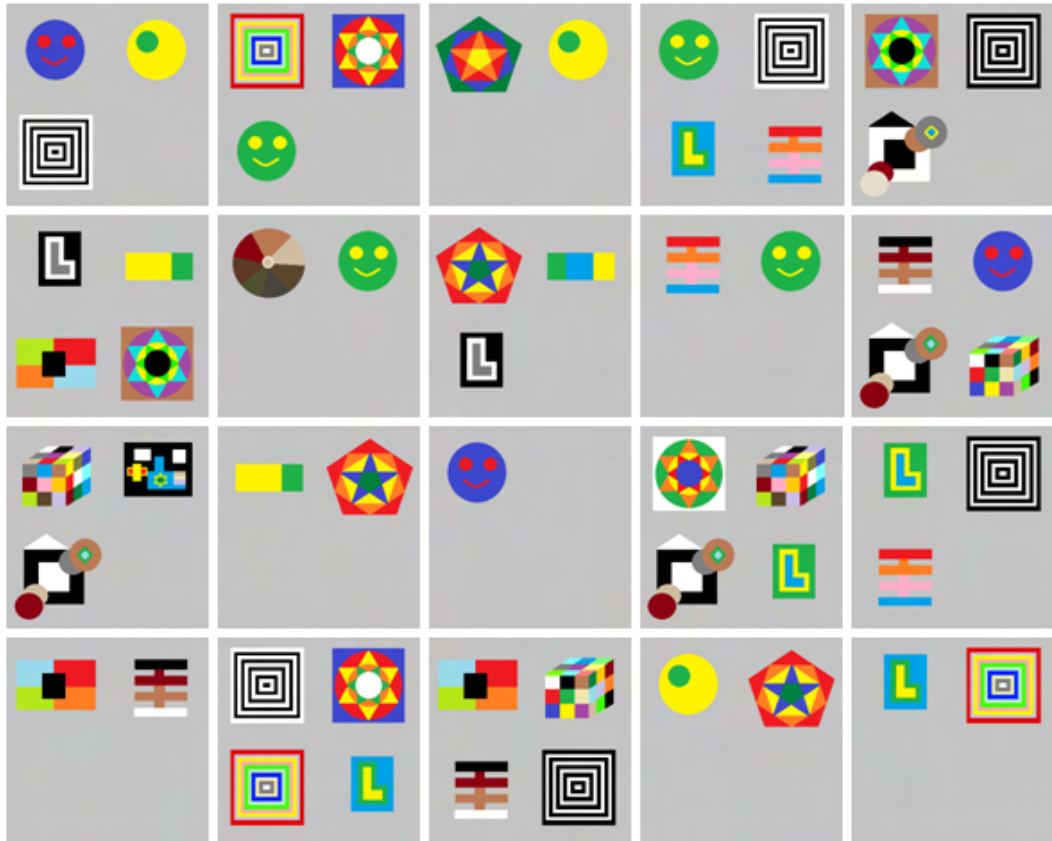


Figura 5.6: Exemplos de imagens da base *Partial-Mock*.

5.2 Influência do nível de quantização

Para encontrarmos o melhor nível de quantização a ser realizado no processo de simplificação das imagens nas abordagens CBD e *Color Spectrum* (veja as Seções 3.1 e 4.1), alguns experimentos foram realizados utilizando a base *Partial-COIL-T*. Nesses experimentos, as imagens presentes na Figura 5.2 foram submetidas como consulta à base *Partial-COIL-T* utilizando as abordagens CBD e *Color Spectrum*.

A Tabela 5.1 apresenta os resultados obtidos em termos das taxas de MAP para imagens quantizadas em 16, 64 e 216 cores. A partir dos resultados apresentados na Tabela 5.1, foi possível concluir que para a abordagem CBD o processo de simplificação deve realizar uma quantização de 216 cores para que se obtenha os melhores resultados. Já para a abordagem *Color Spectrum*, o nível de quantização ideal deve ser de apenas 16 cores.

Tabela 5.1: Influência do nível de quantização na recuperação.

Quantização	MAP (CBD)	MAP (<i>Color Spectrum</i>)
16 cores	88.06%	100%
64 cores	98.76%	90.13%
216 cores	100%	92.53%

A Figura 5.7 ilustra uma imagem qualquer submetida ao processo de simplificação

utilizando diferentes níveis de quantização. Com base nos resultados da Tabela 5.1, os experimentos realizados nas Seções 5.3 e 5.4 consideram uma quantização de 216 cores para a abordagem CBD e uma quantização de 16 cores para a abordagem *Color Spectrum*.

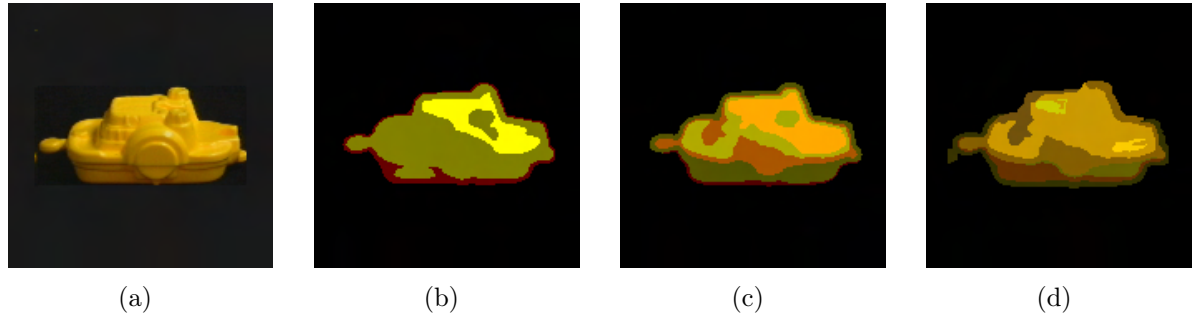


Figura 5.7: Resultado do processo de simplificação aplicado à uma imagem original (a) utilizando diferentes níveis de quantização: (b) 16 cores; (c) 64 cores; (d) 216 cores.

5.3 Experimentos para o descritor CBD

Para avaliar a eficácia da abordagem CBD, nós a comparamos com outras abordagens utilizando as bases de imagens apresentadas na Seção 5.1: *Partial-COIL*, *Partial-COIL-T* e *Partial-Mock*. Cada um dos 30 objetos originais utilizados para a criação das bases foram utilizados como consulta. As Figuras 5.2 e 5.5 mostram essas imagens. As 30 imagens da Figura 5.2 foram utilizadas como consulta nas bases *Partial-COIL* e *Partial-COIL-T*, enquanto que as 30 imagens da Figura 5.5 foram utilizadas como consulta na base *Partial-Mock*.

Os resultados foram avaliados utilizando as curvas de Precisão x Revocação e os valores MAP para 100% de revocação. Além disso, também disponibilizamos os 10 primeiros resultados retornados para uma consulta realizada em cada uma das bases. A partir desses resultados, comparamos a abordagem CBD com outras três abordagens:

- ***Auto Color Correlograms - ACC***: Esse descritor expressa a correlação espacial entre as cores de uma imagem. Os experimentos com esse descritor foram realizados por meio do *framework* LIRe [71].
- ***Color and Edge Directivity Descriptor - CEDD***: Esse descritor incorpora cor e textura em um único histograma levando em conta características locais da imagem. Os experimentos com esse descritor foram executados por meio do *framework* LIRe [71].
- ***Fuzzy Color Texture Histogram - FCTH***: O descritor FCTH também utiliza informações de cor e textura levando em consideração a sua localização espacial. Os experimentos também foram realizados por meio do *framework* LIRe [71].

5.3.1 Avaliação dos Resultados

Os resultados em termos das curvas de Precisão $\times$ Revocação para as bases *Partial-COIL*, *Partial-COIL-T* e *Partial-Mock* são exibidos nas Figuras 5.8(a), 5.8(b) e 5.8(c), respectivamente. As Figuras 5.9 e 5.10 mostram as 10 primeiras imagens recuperadas para cada abordagem nas bases *Partial-COIL-T* e *Partial-Mock*, respectivamente. Todas as abordagens também foram avaliadas em termos das taxas de MAP para as 30 consultas realizadas. A Tabela 5.2 sintetiza os resultados.

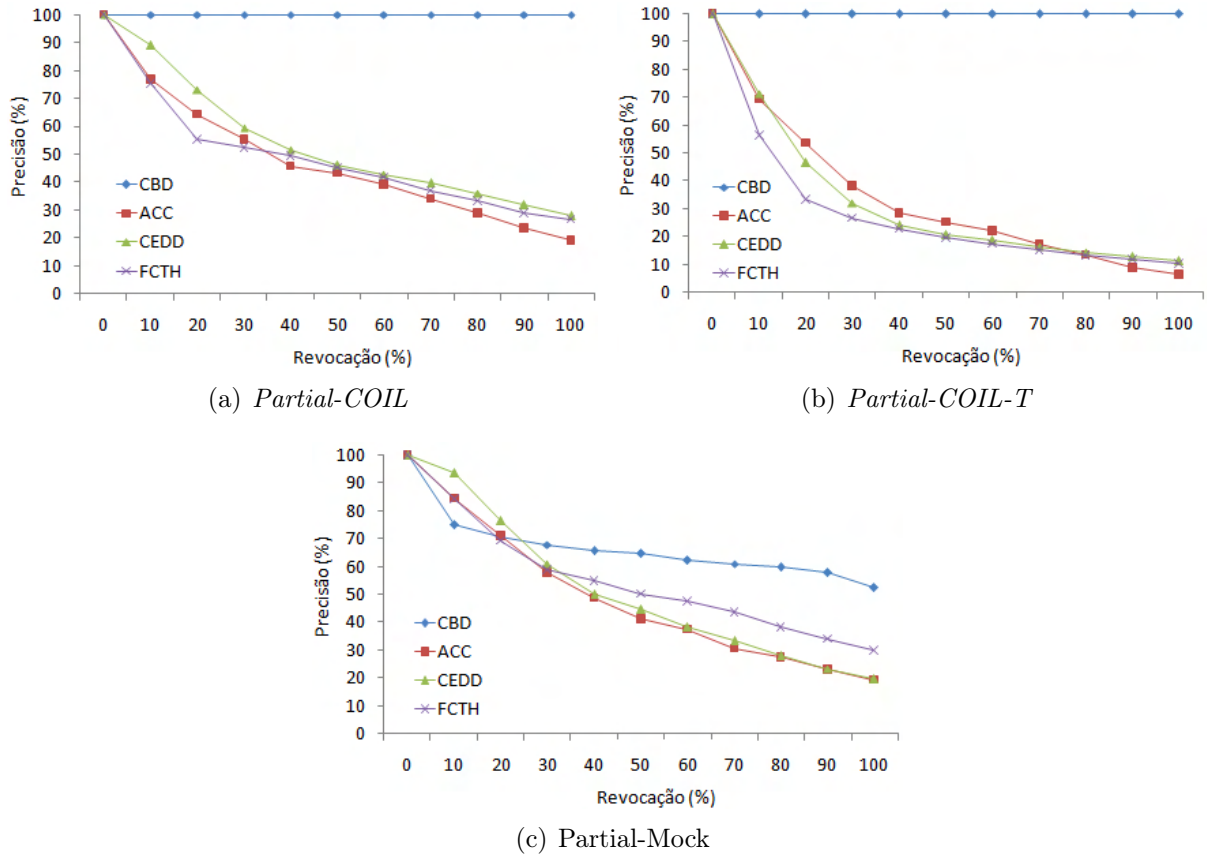


Figura 5.8: Curvas de Precisão $\times$ Revocação para as bases: (a) *Partial-COIL*; (b) *Partial-COIL-T*; (c) *Partial-Mock*.

Tabela 5.2: MAP para 100% de revocação (CBD).

Abordagem	<i>Partial-COIL</i>	<i>Partial-COIL-T</i>	<i>Partial-Mock</i>
CBD	100%	100%	65.9%
ACC	40.5%	26.1%	41.3%
CEDD	47.3%	25.1%	43.7%
FCTH	42.4%	21.4%	48.2%

O descritor CBD apresentou 100% de MAP para as bases *Partial-COIL* e *Partial-COIL-T*, mostrando que é pouco sensível à presença de elementos não relevantes, como pode ser visto nas Figuras 5.8(a) e 5.8(b). Atribuímos esse resultado principalmente ao processo de simplificação utilizado, que garante que diferentes objetos dificilmente terão as mesmas manchas de cor. Além disso, o algoritmo que propomos como medida de similaridade assegura que objetos com manchas de cor muito próximas serão considerados objetos similares, mas não serão considerados o mesmo objeto.

Já para a base *Partial-Mock*, o descritor CBD obteve taxa de precisão mais baixa quando comparado as outras bases. Acreditamos que a precisão foi menor para a base *Partial-Mock*, devido a presença de imagens contendo objetos completamente diferentes mas que possuem exatamente as mesmas cores e de imagens do mesmo objeto representado com diferentes cores. Apesar disso, o descritor CBD continuou sendo superior aos outros descritores de cor tendo alcançado taxa de precisão superior a 52% para todos os valores de revocação, conforme mostra a Figura 5.8(c).

Como podemos ver na Figura 5.8 e na Tabela 5.2, os descritores ACC e CEDD foram superiores ao descritor FCTH nas bases que contêm objetos reais, enquanto que na base composta por objetos criados artificialmente o FCTH foi superior a ambos. O descritor CBD superou todos os descritores (ACC, CEDD, FCTH) utilizados nesse experimento. Esses descritores analisam todas as cores e texturas presentes na imagem, enquanto o descritor CBD leva em conta apenas as cores presentes na imagem de consulta. É interessante notar também que todos os descritores analisados obtiveram valores de MAP diferentes para as bases *Partial-COIL* e *Partial-COIL-T*, enquanto o CBD não foi afetado pela inclusão de elementos não relevantes na base de imagens original.



(a) Imagem utilizada como consulta



(b) CBD



(c) ACC



(d) CEDD



(e) FCTH

Figura 5.9: As 10 primeiras imagens recuperadas da base *Partial-COIL-T* utilizando as abordagens: (b) CBD; (c) ACC; (d) CEDD; (e) FCTH.

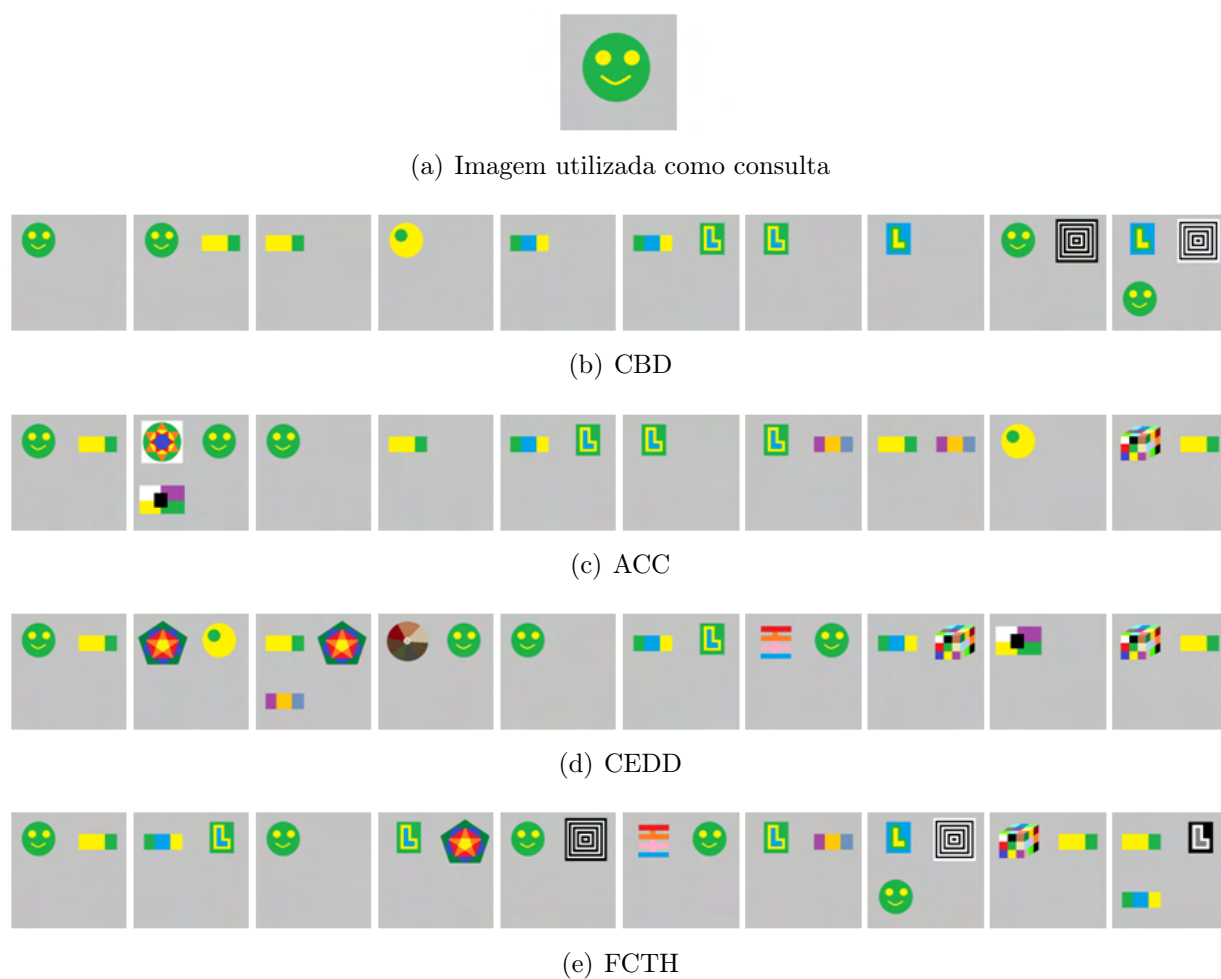


Figura 5.10: As 10 primeiras imagens recuperadas da base *Partial-Mock* utilizando as abordagens: (b) CBD; (c) ACC; (d) CEDD; (e) FCTH.

5.4 Experimentos para o descritor *Color Spectrum*

Para avaliar a precisão da abordagem *Color Spectrum* para a recuperação parcial de imagens, foram realizadas avaliações experimentais utilizando duas bases de imagens. As bases de imagens utilizadas nos experimentos foram apresentadas ao Seção 5.1: *Partial-COIL-T* e *Partial-Mock*. Cada um dos 30 objetos originais utilizados para a criação das bases foram utilizados como consulta. As Figuras 5.2 e 5.5 mostram essas imagens. As 30 imagens da Figura 5.2 foram utilizadas como consulta na base *Partial-COIL-T*, enquanto que as 30 imagens da Figura 5.5 foram utilizadas como consulta na base *Partial-Mock*.

Os resultados foram avaliados utilizando as curvas de Precisão $\times$ Revocação e os valores MAP para 100% de revocação. Além disso, também disponibilizamos os 10 primeiros resultados retornados para uma consulta realizada em cada uma das bases. A partir desses resultados, comparamos a abordagem *Color Spectrum* com as seguintes abordagens:

- ***Auto Color Correlograms - ACC***: Esse descritor foi escolhido por ter obtido os melhores resultados na base *Partial-COIL-T* quando comparado aos outros descritores de cor avaliados na Seção 5.3. O descritor ACC expressa a correlação espacial entre as cores de uma imagem. Os experimentos com esse descritor foram realizados por meio do *framework* LIRe [71].
- ***Fuzzy Color Texture Histogram - FCTH***: O descritor FCTH incorpora cor e textura em um único histograma levando em consideração a sua localização espacial. Esse descritor foi escolhido por ter obtido os melhores resultados na base *Partial-Mock* quando comparado aos outros descritores de cor avaliados na Seção 5.3. Os experimentos com esse descritor também foram realizados por meio do *framework* LIRe [71].
- ***Bag of Visual Feature with SIFT - BoVF/SIFT***: O descritor SIFT foi escolhido devido à sua habilidade em descrever características espaciais locais. Para obter as palavras visuais foi utilizada a mesma metodologia proposta por Csurka et al. [66]. O tamanho do dicionário utilizado foi de 1000 palavras visuais e a distância cosseno foi utilizada como medida de similaridade.
- **Espectro de adjacência**: Esse descritor extrai apenas características espaciais das imagens. Para o cálculo do espectro de adjacência foi utilizada a mesma metodologia para o cálculo do descritor *Color Spectrum*, diferindo apenas da representação matricial do grafo topológico. Enquanto a abordagem *Color Spectrum* utiliza a matriz de similaridade de cor, o espectro de adjacência utiliza a matriz de adjacência.
- **Combinação do espectro de adjacência com descritores de cor**: Essa abordagem realiza uma combinação entre um descritor topológico e descritores de cor. Para que não haja distorção nos resultados, as similaridades topológicas e de cor

foram previamente normalizadas no intervalo $[0, 1]$. A distância final D_w é calculada pela soma ponderada da similaridade de cor D_c e da similaridade topológica D_t , conforme definida na Equação 5.1:

$$D_w = \frac{(W_c \times D_c + W_t \times D_t)}{2} \quad (5.1)$$

onde W_t e W_c são os pesos para os descritores topológicos e de cor, respectivamente. Os experimentos foram realizados atribuindo-se o mesmo peso para os descritores (0.5 para W_t e 0.5 para W_c).

5.4.1 Avaliação dos Resultados

Os resultados em termos das curvas de Precisão $\times$ Revocação para as bases *Partial-COIL-T* e *Partial-Mock* são exibidos nas Figuras 5.11(a) e 5.11(b), respectivamente. As Figuras 5.12 e 5.13 mostram as 10 primeiras imagens recuperadas para cada abordagem nas bases *Partial-COIL-T* e *Partial-Mock*, respectivamente, para uma determinada consulta. Todas as abordagens também foram avaliadas em termos das taxas de MAP para as 30 consultas realizadas. A Tabela 5.3 sintetiza os resultados.

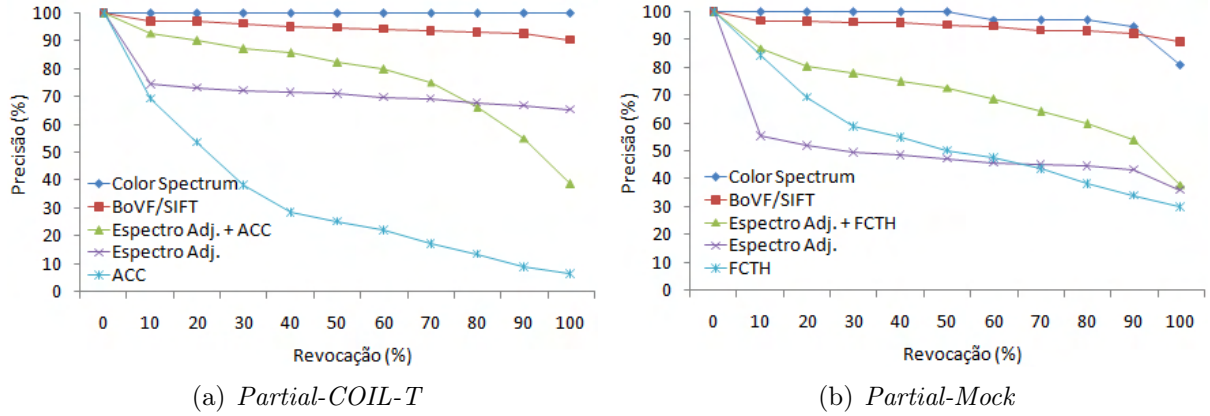


Figura 5.11: Curvas de Precisão $\times$ Revocação para as bases: (a) *Partial-COIL-T*; (b) *Partial-Mock*.

Tabela 5.3: MAP para 100% de revocação (*Color Spectrum*).

Abordagem	<i>Partial-COIL-T</i>	<i>Partial-Mock</i>
Color Spectrum	100%	96.6%
BoVF/SIFT	94.5%	93.4%
Espectro de Adj. + FCTH	-	67.7%
Espectro de Adj. + ACC	75.3%	-
Espectro de Adj.	70.1%	46.7%
FCTH	21.4%	48.2%
ACC	26.1%	41.3%

Nesse experimento comparamos a abordagem *Color Spectrum* para a recuperação parcial de imagens com os melhores descritores de cor identificados nos experimentos anteriores (veja a Seção 5.3), com o espectro de adjacência, com uma combinação entre o espectro de adjacência e os melhores descritores de cor, e com uma abordagem utilizando o algoritmo SIFT.

O descritor *Color Spectrum* apresentou os melhores valores de precisão/revocação, alcançando 100% de precisão para todos os valores de revocação, na base *Patial-COIL-T*, e uma taxa de precisão superior a 80% para todos os valores de revocação na base *Partial-Mock*. Os valores de MAP confirmam esses resultados, mostrando que o *Color Spectrum* obteve 100% de MAP para a base *Partial-COIL-T* e 96% de MAP para a base *Partial-Mock*. Embora a precisão para o *Color Spectrum* diminua para a base *Partial-Mock*, onde os objetos possuem uma estrutura mais complexa em termos de organização espacial, ele ainda continua obtendo os melhores resultados.

A abordagem BoVF/SIFT apresentou bons resultados por explorar características locais de forma dos objetos. No entanto, por ser invariante à cor, essa abordagem obtém bons resultados em bases de imagens contendo objetos diferentes com cores similares, mas não alcança resultados tão bons em bases com objetos com forma similar mas com cores diferentes.

Os descritores de cor apresentaram os piores resultados, embora eles também descrevam alguma informação espacial. Provavelmente esse resultado ruim seja devido ao fato de que eles são capazes de descrever a distribuição espacial das cores, mas apenas considerando a imagem como um todo e não as suas subpartes.

O espectro de adjacência, que descreve apenas o relacionamento espacial entre as regiões da imagem, apresentou um bom resultado para a base *Partial-COIL-T*, mas não foi tão bom para a base *Partial-Mock*. Acreditamos que a precisão foi menor na base *Partial-Mock*, devido a presença de imagens contendo objetos com exatamente a mesma topologia mas com cores diferentes.

A partir da Figura 5.11 podemos ver que a combinação entre descritores topológicos e descritores de cor melhoram a precisão dos resultados, alcançando valores de precisão superiores aos descritores topológicos ou de cor sozinhos. Os valores de MAP apresentados na Tabela 5.3 também confirmam esses resultados.

Para uma análise qualitativa, as Figuras 5.12 e 5.13 mostram as 10 primeiras imagens retornadas para uma mesma consulta realizada utilizando todos os algoritmos analisados, usando as bases *Partial-COIL-T* e *Partial-Mock*. Note que o descritor *Color Spectrum* é capaz de diferenciar objetos com a mesma topologia e diferentes cores, de objetos com diferentes topologias e as mesmas cores.



Figura 5.12: As 10 primeiras imagens recuperadas da base *Partial-COIL-T* utilizando as abordagens: (b) *Color Spectrum*; (c) BoVF/SIFT; (d) Espectro de Adj. + ACC; (e) Espectro de Adj.; (f) ACC.

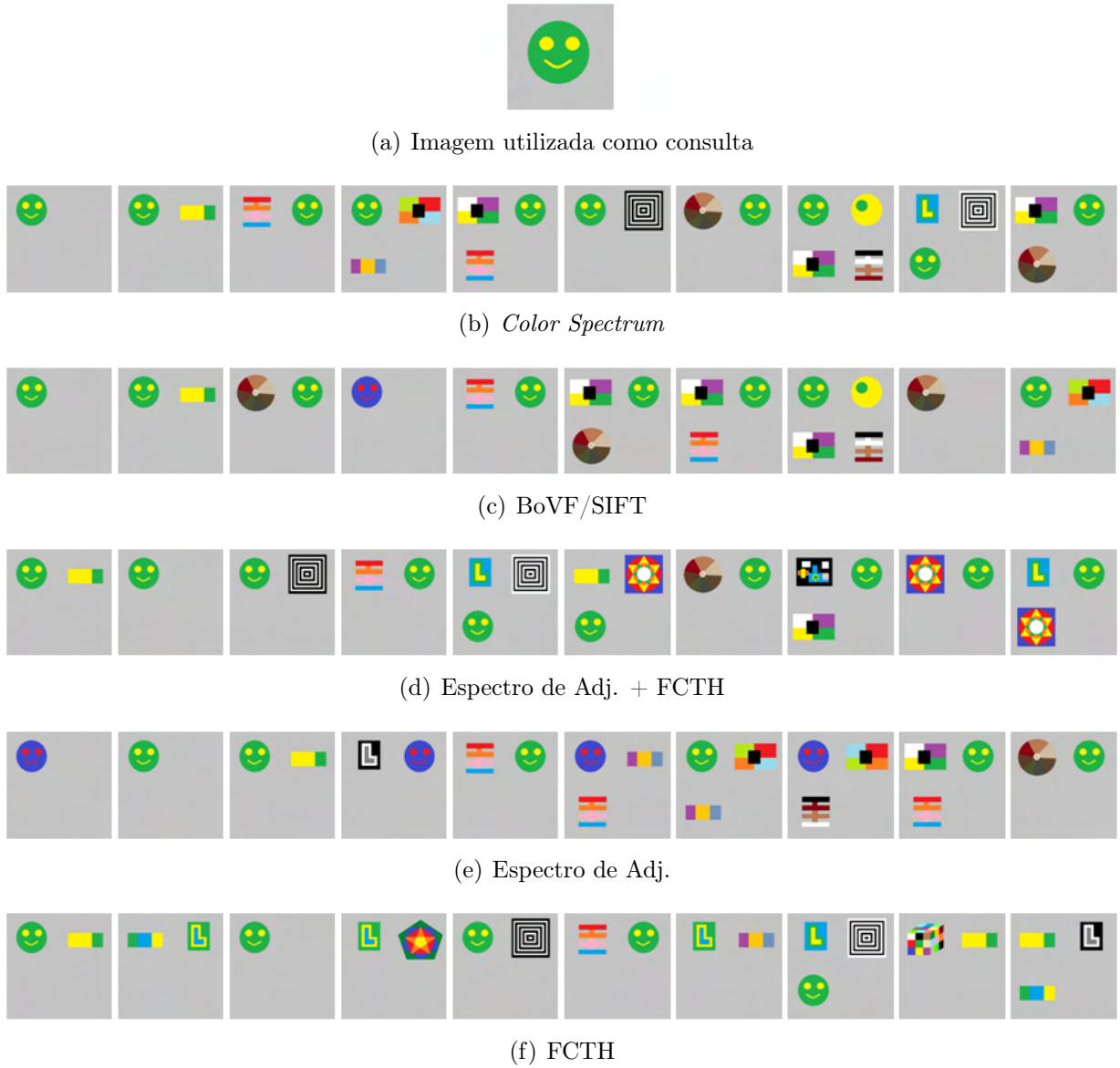


Figura 5.13: As 10 primeiras imagens recuperadas da base *Partial-Mock* utilizando as abordagens: (b) *Color Spectrum*; (c) BoVF/SIFT; (d) Espectro de Adj. + FCTH; (e) Espectro de Adj.; (f) FCTH.

5.5 Considerações Finais

Neste capítulo apresentamos as avaliações experimentais realizadas para os dois novos descritores propostos. Também apresentamos as três bases de imagens criadas para avaliar a qualidade das abordagens no processo de recuperação parcial de imagens por conteúdo. Essas bases possuem tanto objetos extraídos de cenas reais quanto objetos criados artificialmente. Os objetos extraídos de cenas reais sofrem influência das condições de iluminação bem como das transições de cor e diferenças de textura. Já os objetos criados artificialmente permitem reconhecer mais facilmente as suas características topológicas e de cor.

As bases *Partial-COIL*, *Partial-COIL-T* e *Partial-Mock* permitem avaliar a precisão na recuperação de objetos individuais em imagens que contenham vários objetos. Além disso, essas bases incluem imagens de diferentes objetos com as mesmas cores e o mesmo objeto com diferentes cores. Imagens com essas características permitem avaliar quais abordagens conseguem diferenciar objetos com mesma topologia mas com cores diferentes, e vice versa.

Uma avaliação experimental foi realizada para o descritor CBD em que foi possível mostrar os pontos fracos e fortes para esse novo descritor proposto. Nessa avaliação experimental, o descritor CBD se mostrou robusto quanto à presença de elementos não relevantes na base de imagens, alcançando altas taxas de precisão. Além disso, o CBD se mostrou superior a outros três descritores avaliados: ACC, CEDD e FCTH.

Apesar dos bons resultados, o CBD não leva em consideração a topologia dos objetos presentes na imagem. Sendo assim, objetos diferentes mas com os mesmos conjuntos de cores serão erroneamente considerados similares. Os resultados dos experimentos realizados com a base *Partial-Mock* retratam isso.

O descritor *Color Spectrum* combina em um único descritor informações sobre cor e sobre a organização espacial das imagens. Uma avaliação experimental utilizando diferentes bases de imagens, mostrou que abordagem *Color Spectrum* foi superior aos descritores de cores avaliados, ao espectro de adjacência que utiliza apenas informação topológica, ao algoritmo SIFT que utiliza informações locais espaciais e de forma, e uma combinação de informações topológicas e de cor, alcançando valores de MAP superiores a 96% para todos os valores de revocação.

Capítulo 6

Conclusão e Trabalhos Futuros

Este capítulo apresenta as conclusões e resultados obtidos neste trabalho de mestrado, bem como algumas sugestões possíveis para trabalhos futuros.

6.1 Conclusão

Embora diversas estratégias utilizadas para a recuperação de imagens por conteúdo apresentem bons resultados, muitas dessas estratégias não são capazes de diferenciar adequadamente imagens que contêm objetos com as mesmas cores mas com organização espacial diferente, ou simplesmente não suportam a pesquisa parcial de imagens.

Neste trabalho apresentamos dois novos descritores para a recuperação parcial de imagens. O primeiro descritor, denominado *Color Blobs Descriptor* (CBD), baseia-se apenas na característica de cor. O descritor CBD utiliza uma abordagem local para identificar um conjunto de manchas de cor na imagem, após um processo de simplificação da imagem original. Para recuperar as imagens que contêm o objeto de interesse, é realizada uma verificação para descobrir se as manchas de cor identificadas na imagem de consulta também ocorrem na imagem a ser analisada. Os experimentos realizados mostraram que o CBD é robusto quanto à presença de grande quantidade de imagens não relevantes na base de imagens, tendo obtido taxas de precisão superiores aos três outros descritores avaliados: ACC, CEDD, FCTH.

Apesar dos bons resultados, o descritor CBD não leva em consideração a organização espacial dos objetos presentes na imagem. Por esse motivo, objetos diferentes mas que possuem os mesmos conjuntos de cores podem ser erroneamente considerados similares. Para melhor entender a importância da topologia dos objetos, foi realizado um estudo onde informações topológicas e de cor dos objetos foram combinadas para descrever as características das imagens. Os resultados desse estudo demonstraram que a associação entre descritores topológicos e de cor melhora significativamente os resultados obtidos por cada descritor utilizado isoladamente.

O segundo descritor proposto, denominado *Color Spectrum*, combina informações topológicas e de cor em um único descritor. O descritor *Color Spectrum* utiliza uma abordagem global, onde, primeiramente é criado um grafo de topologia que descreve o relacionamento espacial entre as regiões de cor identificadas após um processo de simplificação da imagem original. O descritor topológico é então calculado a partir do grafo de topologia utilizando a teoria espectral de grafos. Dessa forma, o problema da correspondência entre grafos é reduzido ao cálculo de similaridade entre vetores de características. Para suportar a pesquisa parcial de imagens, uma decomposição do grafo principal em diversos subgrafos é realizada. Os experimentos realizados mostraram que o *Color Spectrum* foi superior as outras abordagens avaliadas, obtendo valores de MAP superiores a 96%.

O processo de simplificação das imagens, utilizado tanto pelo descritor CBD quanto pelo descritor *Color Spectrum*, é uma etapa de grande importância dentro das abordagens propostas. Pensando nisso, novas estratégias para a simplificação de imagens estão sendo estudadas, como por exemplo, a utilização de autômatos celulares.

Outra contribuição deste trabalho foi a criação de bases de imagens próprias para avaliar a qualidade das abordagens na recuperação parcial de imagens. Ao todo foram criadas três bases de imagens denominadas: *Partial-COIL*, *Partial-COIL-T* e *Partial-Mock*. Essas bases de imagens possuem tanto objetos extraídos de cenas reais quanto objetos criados artificialmente. Além disso, as bases de imagens criadas incluem imagens de diferentes objetos com as mesmas cores e o mesmo objetos com diferentes cores. Imagens com essas características permitem avaliar quais abordagens conseguem diferenciar objetos com a mesma topologia mas com cores diferentes, e vice versa.

6.2 Trabalhos Futuros

Alguns trabalhos que podem ser realizados a partir deste são:

1. Avaliar novas estratégias para a simplificação das imagens, como por exemplo, a utilização de autômatos celulares, experimentar diferentes abordagens de segmentação, dentre outras.
2. Alterar o descritor topológico de forma que seja possível descrever a posição espacial dos objetos na cena, em termos de esquerda, direita, cima, baixo.
3. Alterar o descritor topológico de forma que seja possível descrever a proximidade entre os objetos na cena. Nesse caso, objetos mais próximos poderiam receber um peso maior do que objetos mais afastados.
4. Avaliar o comportamento dos descritores propostos em outras bases de imagens.

Referências Bibliográficas

- [1] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603–619, 2002.
- [2] R. da Silva Torres and A. X. Falcão, “Content-based image retrieval: Theory and applications,” *Revista de Informática Teórica e Aplicada*, vol. 13, no. 2, pp. 161–185, 2006.
- [3] M. Bober, “Mpeg-7 visual shape descriptors,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 716–719, 2001.
- [4] C. D. Manning, P. Raghavan, and H. Schtze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.
- [5] M. de Carvalho, A. Ferreira, and A. Costa, “Image segmentation using quadtree-based similarity graph and normalized cut,” in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ser. Lecture Notes in Computer Science, I. Bloch and R. Cesar, Eds. Springer Berlin / Heidelberg, 2010, vol. 6419, pp. 329–337. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-16687-7_45
- [6] M. J. Egenhofer, “A formal definition of binary topological relationships,” in *3rd International Conference, FODO 1989 on Foundations of Data Organization and Algorithms*. New York, NY, USA: Springer-Verlag New York, Inc., 1989, pp. 457–472. [Online]. Available: <http://dl.acm.org/citation.cfm?id=68413.68447>
- [7] S. A. Nene, S. K. Nayar, and H. Murase, “Columbia Object Image Library (COIL-100),” Tech. Rep., Feb 1996.
- [8] “Internet World Stats.” [Online]. Available: <http://www.internetworldstats.com/am/us.htm>
- [9] N. Chang and K. Fu, “A relational database system for images,” in *Pictorial Information Systems*, ser. Lecture Notes in Computer Science, S. Chang and K. Fu, Eds. Springer Berlin / Heidelberg, 1980, vol. 80, pp. 288–321. [Online]. Available: http://dx.doi.org/10.1007/3-540-09757-0_11
- [10] N.-S. Chang and K.-S. Fu, “Query-by-pictorial-example,” *IEEE Transactions on Software Engineering*, vol. 6, no. 6, pp. 519–524, Nov. 1980. [Online]. Available: <http://dx.doi.org/10.1109/TSE.1980.230801>
- [11] S.-K. Chang and T. L. Kunil, “Pictorial data-base systems,” *IEEE Computer Magazine*, vol. 14, no. 11, pp. 13–21, Nov. 1981. [Online]. Available: <http://dx.doi.org/10.1109/C-M.1981.220243>

- [12] S. K. Chang, C. W. Yan, D. C. Dimitroff, and T. Arndt, “An intelligent image database system,” *IEEE Transactions on Software Engineering*, vol. 14, no. 5, pp. 681–688, May 1988. [Online]. Available: <http://dx.doi.org/10.1109/32.6147>
- [13] B. V. Funt and G. D. Finlayson, “Color constant color indexing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 522–529, 1995.
- [14] W.-Y. Ma and B. S. Manjunath, “Texture-based pattern retrieval from image databases,” *Multimedia Tools and Applications*, vol. 2, no. 1, pp. 35–51, 1996.
- [15] H. V. Jagadish, “A retrieval technique for similar shapes,” in *Proceedings of the 1991 ACM SIGMOD international conference on Management of data*, ser. SIGMOD '91. New York, NY, USA: ACM, 1991, pp. 208–217. [Online]. Available: <http://doi.acm.org/10.1145/115790.115821>
- [16] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004. [Online]. Available: <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- [17] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, Jun. 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.cviu.2007.09.014>
- [18] S. A. Chatzichristofis and Y. S. Boutalis, “Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval,” in *Proceedings of the 6th international conference on Computer vision systems*, ser. ICVS'08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 312–322. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1788524.1788559>
- [19] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, “Image indexing using color correlograms,” in *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, ser. CVPR '97. Washington, DC, USA: IEEE Computer Society, 1997, pp. 762–. [Online]. Available: <http://dl.acm.org/citation.cfm?id=794189.794514>
- [20] S. A. Chatzichristofis and Y. S. Boutalis, “Fcth: Fuzzy color and texture histogram - a low level feature for accurate image retrieval,” *Image Analysis for Multimedia Interactive Services, International Workshop on*, vol. 0, pp. 191–196, 2008.
- [21] F. Chung, *Spectral Graph Theory*, ser. Regional Conference Series in Mathematics. Conference Board of the Mathematical Sciences, 1997, no. N° 92. [Online]. Available: http://books.google.com.br/books?id=YUc38_MCuhAC
- [22] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
- [23] G. Stockman and L. G. Shapiro, *Computer Vision*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.
- [24] N. R. Pal and S. K. Pal, “A review on image segmentation techniques,” *Pattern Recognition*, no. 9, pp. 1277–1294, 1993.

- [25] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, 1989. [Online]. Available: <http://doi.wiley.com/10.1002/cpa.3160420503>
- [26] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognition*, vol. 40, no. 3, pp. 825–838, Mar. 2007. [Online]. Available: <http://dx.doi.org/10.1016/j.patcog.2006.07.011>
- [27] S. Beucher and F. Meyer, "The morphological approach to segmentation: the watershed transformation," in *Mathematical Morphology in Image Processing*, E. R. Dougherty, Ed. New York: Marcel Dekker, 1993, ch. 12, pp. 433–481.
- [28] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *Information Theory, IEEE Transactions on*, vol. 21, no. 1, pp. 32–40, 1975.
- [29] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, Aug. 1995. [Online]. Available: <http://dx.doi.org/10.1109/34.400568>
- [30] D. Feng, W. C. Siu, and H. J. Zhang, *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, 1st ed. Springer Publishing Company, Incorporated, 2010.
- [31] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," in *FnT Comp. Graphics and Vision*, 2008, pp. 177–280. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.157.4126>
- [32] G. Lu, *Multimedia Database Management Systems*. Artech House Publishers, 1999, ISBN: 0-89006-342-7.
- [33] A. Del Bimbo, *Visual information retrieval*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999.
- [34] R. Stehling, M. Nascimento, and A. Falcao, *Multimedia Mining - a High Way to Intelligent Multimedia Document*. C. Djeraba (Eds) - Kluwer Academic Publishers, 2002, ch. 4 - Techniques for Color-Based Image Retrieval.
- [35] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, pp. 11–32, 1991.
- [36] R. O. Stehling, M. A. Nascimento, and A. X. Falcão, "An adaptive and efficient clustering-based approach for content-based image retrieval in image databases," in *Proceedings of the International Database Engineering & Applications Symposium*, ser. IDEAS '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 356–365. [Online]. Available: <http://dl.acm.org/citation.cfm?id=646290.687049>
- [37] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, "Blobworld: A system for region-based image indexing and retrieval," in *In Third International Conference on Visual Information Systems*. Springer, 1999, pp. 509–516.

- [38] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, 2004.
- [39] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. Chapman & Hall, 1998.
- [40] F. Mokhtarian, S. Abbasi, and J. Kittler, "Efficient and robust retrieval by shape content through curvature scale space," 1996, pp. 35–42.
- [41] L. Kotoulas and I. Andreadis, "Image analysis using moments," *Information Systems Journal*, vol. 1, no. 1, pp. 360–364, 2005. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.2664&rep=rep1&type=pdf>
- [42] J. R. Parker, *Algorithms for Image Processing and Computer Vision*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1996.
- [43] S. Santini and R. Jain, "Gabor space and the development of preattentive similarity," in *In International Conference on Pattern Recognition*, 1996.
- [44] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, Aug. 1996. [Online]. Available: <http://dx.doi.org/10.1109/34.531803>
- [45] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1513–1521, jun 2003. [Online]. Available: [http://dx.doi.org/10.1016/S0167-8655\(02\)00390-2](http://dx.doi.org/10.1016/S0167-8655(02)00390-2)
- [46] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc., 1986.
- [47] L. Euler, "Solutio problematis ad geometriam situs pertinentis," *Commentarii academiae scientiarum Petropolitanae*, vol. 8, pp. 128–140, 1736. [Online]. Available: <http://www.math.dartmouth.edu/~euler/pages/E053.html>
- [48] R. Diestel, *Graph Theory*, 3rd ed., ser. Graduate Texts in Mathematics. Springer-Verlag, Heidelberg, 2005, vol. 173. [Online]. Available: <http://vg00.met.vgwort.de/na/ddfc84df913d6ef96f8f?l=http://www.math.uni-hamburg.de/home/diestel/books/graph.theory/GraphTheoryIII.pdf>
- [49] T. Tan and P. Mulhem, "Image query system using object probes," in *Proceedings of the International Conference on Image Processing*, 2001, pp. 701–704.
- [50] M. J. Fonseca, A. Ferreira, and J. A. Jorge, "Sketch-based retrieval of complex drawings using hierarchical topology and geometry," *Computer Aided Design*, vol. 41, no. 12, pp. 1067–1081, dec 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.cad.2009.09.004>
- [51] W. H. Leung and T. Chen, "Hierarchical matching for retrieval of hand-drawn sketches," in *Proceedings of the 2003 International Conference on Multimedia and Expo - Volume 1*, ser. ICME '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 29–32. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1153922.1154421>

- [52] A. Hlaoui and S. Wang, “Using graph representation in content-based image retrieval,” in *Proceedings of AICCSA*, Tunis, Tunisia, 2003.
- [53] H. Aboulmagd, N. E. Gayar, and H. Onsi, “A new approach in content-based image retrieval using fuzzy,” *Telecommunication Systems*, vol. 40, no. 1-2, pp. 55–66, 2009.
- [54] L. Hogben, *Handbook of Linear Algebra*, ser. Discrete Mathematics And Its Applications. Chapman & Hall/CRC, 2007. [Online]. Available: <http://books.google.com.br/books?id=n2g-x1OIbvYC>
- [55] R. Wilson and J. Watkins, *Graphs: an introductory approach : a first course in discrete mathematics*. Wiley, 1990. [Online]. Available: <http://books.google.com.br/books?id=cevuAAAAMAAJ>
- [56] N. M. M. de ABREU, “Teoria espectral dos grafos: Um híbrido entre a Álgebra linear e a matemática discreta e combinatória com origens na química quântica,” *TEMA, Tendências em Matemática Aplicada e Computacional*, vol. 6, pp. 1–10, 2005.
- [57] D. Cvetković, M. Doob, and H. Sachs, *Spectra of graphs: theory and application*, ser. Pure and applied mathematics. Academic Press, 1980. [Online]. Available: <http://books.google.com.br/books?id=4u7uAAAAMAAJ>
- [58] K. Rosen, *Discrete mathematics and its applications*, ser. Random House/Birkhäuser mathematics series. Random House, 1988. [Online]. Available: <http://books.google.com.br/books?id=9X9QAAAAMAAJ>
- [59] D. Cvetkovic, P. Rowlinson, and S. Simic, *Eigenspaces of Graphs*, ser. Encyclopedia of Mathematics and its Applications. Cambridge: Cambridge University Press, Encyclopedia of Mathematics, 1997, vol. 66.
- [60] A. Shokoufandeh, S. J. Dickinson, K. Siddiqi, and S. W. Zucker, “Indexing using a spectral encoding of topological structure,” *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 2, p. 2491, 1999.
- [61] M. A. Stricker and M. Orengo, “Similarity of Color Images,” in *Storage and Retrieval for Image and Video Databases (SPIE)*, 1995, pp. 381–392. [Online]. Available: citeseer.nj.nec.com/stricker95similarity.html
- [62] S. A. Chatzichristofis, K. Zagoris, Y. S. Boutalis, and N. Papamarkos, “Accurate image retrieval based on compact composite descriptors and relevance feedback information,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 24, no. 2, pp. 207–244, 2010.
- [63] C. S. Won, D. K. Park, and S. J. Park, “Efficient use of mpeg-7 edge histogram descriptor,” *Electronics and Telecommunications Research Institute Journal*, vol. 24, no. 1, pp. 23–30, 2002.
- [64] E. Stollnitz, T. DeRose, and D. Salesin, *Wavelets for Computer Graphics: Theory and Applications*, ser. Morgan Kaufmann Series in Computer Graphics and Geometric Modeling. Morgan Kaufmann Publishers, 1996. [Online]. Available: <http://books.google.com.br/books?id=KbplQroKrTsC>

- [65] J. Sivic and A. Zisserman, “Video Google: A text retrieval approach to object matching in videos,” in *Proceedings of the International Conference on Computer Vision*, vol. 2, Oct. 2003, pp. 1470–1477. [Online]. Available: <http://www.robots.ox.ac.uk/~vgg>
- [66] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” in *Workshop on Statistical Learning in Computer Vision, ECCV*, 2004, pp. 1–22.
- [67] S. Liang, Z.-X. Sun, B. Li, and G.-H. Feng, “Effective sketch retrieval based on its contents,” in *Proceedings of Machine Learning and Cybernetics - Volume 9*, Guangzhou, China, 2005, pp. 5266–5272.
- [68] P. M. A. Sousa and M. J. Fonseca, “Sketch-based retrieval of drawings using spatial proximity,” *Journal of Visual Languages and Computing*, vol. 21, no. 2, pp. 69–80, 2010.
- [69] D. Conte, P. Foggia, C. Sansone, and M. Vento, “Thirty years of graph matching in pattern recognition,” *International Journal of Pattern Recognition and Artificial Intelligence*, 2004. [Online]. Available: <http://www.worldscinet.com/ijprai/18/1803/S0218001404003228.html>
- [70] R. C. Wilson and P. Zhu, “A study of graph spectra for comparing graphs and trees,” *Pattern Recognition*, vol. 41, no. 9, pp. 2833–2841, Sep. 2008. [Online]. Available: <http://dx.doi.org/10.1016/j.patcog.2008.03.011>
- [71] M. Lux and S. A. Chatzichristofis, “Lire: lucene image retrieval: an extensible java cbir library,” in *ACM*, ser. MM ’08, 2008, pp. 1085–1088. [Online]. Available: <http://doi.acm.org/10.1145/1459359.1459577>