

Universidade Federal do Amazonas  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Programa de Pós-Graduação em Informática

PETRINA DE ASSIS DA SILVA KIMURA

## **Um Descritor Baseado em Análise Local de Cor para Busca de Imagens em Grandes Coleções**

Manaus  
2011

**Petrina de Assis da Silva Kimura**

**Um Descritor Baseado em Análise Local de Cor para  
Busca de Imagens em Grandes Coleções**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Informática do Departamento de Ciência da Computação da Universidade Federal do Amazonas, como parte dos requisitos necessários para a obtenção do título de Mestre em Informática.

**Orientador: Prof. Dr. João Marcos Bastos Cavalcanti**

Manaus

2011

**Petrina de Assis da Silva Kimura**

**Um Descritor baseado em Análise Local de Cor para Busca de  
Imagens em Grandes Coleções**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Informática do Departamento de Ciência da Computação da Universidade Federal do Amazonas, como parte dos requisitos necessários para a obtenção do título de Mestre em Informática.

Banca Examinadora

---

Prof. Dr. João Marcos Bastos Cavalcanti (Orientador)  
Universidade Federal do Amazonas

---

Prof. Dr. Edleno Silva de Moura  
Universidade Federal do Amazonas

---

Prof. Dr. Ricardo da Silva Torres  
Universidade Estadual de Campinas

Manaus – 2011

Aos meus pais pelo incentivo e apoio nesta difícil caminhada.

# Agradecimentos

Primeiramente, a Deus pelo dom da vida e por todas as minhas conquistas.

À minha mãe pelo incentivo, motivação e apoio, e por ter me guiado pelo caminho certo, educando e ajudando a superar os obstáculos encontrados ao longo desta caminhada.

Ao meu pai por todo sacrifício realizado ao longo de todos esses anos para que eu pudesse concluir meus estudos.

Aos meus familiares por toda a confiança depositada em mim, fazendo com que eu não desistisse do meu propósito.

Ao meu orientador Prof. Dr. João Marcos, pela orientação e suporte necessário para a conclusão desta dissertação.

Ao Prof. Dr. Marcos Gonçalves da UFMG por ter colaborado na elaboração do artigo resultante do trabalho desenvolvido nesta dissertação.

Ao Prof. Dr. Ricardo da Silva Torres do Instituto de Computação da UNICAMP por ter participado da minha banca e pela ajuda na revisão do artigo.

Ao Prof. Dr. Edleno Moura da UFAM pela participação na banca e pela contribuição com sugestões para o desenvolvimento desta dissertação.

À Patrícia Saraiva pela contribuição dada na elaboração do artigo resultante deste trabalho.

A Otávio Penatti pelo suporte dado ao longo do desenvolvimento desta dissertação.

Aos meus amigos que me proporcionaram muitos momentos alegres e que estiveram sempre presentes, pois minha jornada foi muito mais fácil com vocês sempre ao meu lado.

A todos que, de alguma forma, contribuíram para a conclusão deste trabalho.

A FAPEAM, pelo apoio financeiro.

*"E sem saber que era impossível, ele foi lá e fez"*

*Jean Cocteau*

# Sumário

<b>Lista de Figuras</b>	<b>9</b>
<b>Lista de Tabelas</b>	<b>10</b>
<b>Resumo</b>	<b>12</b>
<b>Abstract</b>	<b>13</b>
<b>1 Introdução</b>	<b>14</b>
1.1 Motivação . . . . .	16
1.2 Justificativa . . . . .	17
1.3 Objetivos . . . . .	17
1.4 Organização do trabalho . . . . .	18
<b>2 Fundamentação Teórica e Trabalhos Relacionados</b>	<b>19</b>
2.1 Descritor de imagens . . . . .	19
2.2 Características de imagens . . . . .	21
2.2.1 Abordagem global . . . . .	23
2.2.2 Abordagem baseada em segmentação . . . . .	23
2.2.3 Abordagem baseada em partições fixas . . . . .	25
2.3 Características das bases de imagens . . . . .	25
2.4 Trabalhos Relacionados . . . . .	27
<b>3 Local Color Pixel Classification - um Descritor Baseado em Características Locais</b>	<b>34</b>
3.1 Segmentação versus Particionamento Fixo . . . . .	35
3.1.1 Abordagem baseada em segmentação por crescimento de região . . . . .	36

---

3.1.2	Abordagem baseada em partições estratégicas . . . . .	38
3.2	Local Color Pixel Classification - LCPC . . . . .	46
3.2.1	LCPC com características de cor . . . . .	47
3.2.2	LCPC com características de cor e aresta . . . . .	47
<b>4</b>	<b>Experimentos</b>	<b>51</b>
4.1	Métricas de Avaliação Utilizadas . . . . .	51
4.1.1	Curvas de Precisão X Revocação . . . . .	52
4.1.2	Precisão na N-ésima posição do ranking - P@N . . . . .	52
4.1.3	Mean Average Precision - MAP . . . . .	53
4.1.4	Método de Pooling . . . . .	53
4.1.5	Teste de validação Wilcoxon Matched-Pairs Signed-Ranks . . . . .	54
4.2	Base de Imagens . . . . .	55
4.2.1	Base Wang . . . . .	55
4.2.2	Base Common Color Dataset - CCD . . . . .	55
4.2.3	Base Yahoo . . . . .	57
4.3	Resultados . . . . .	58
4.3.1	Estudo comparativo entre os três esquemas de particionamento . . . . .	58
4.3.2	Comparação entre o LCPC e outros descritores . . . . .	63
<b>5</b>	<b>Conclusões e Trabalhos futuros</b>	<b>74</b>
5.1	Trabalhos futuros . . . . .	75
	<b>Referências bibliográficas</b>	<b>78</b>



# Lista de Figuras

1.1	Imagem com diferentes interpretações. . . . .	15
2.1	Processo realizado por um descritor de imagem. . . . .	20
2.2	Imagem (a) com tons de cores mais claros e imagem (b) com tons escuros. . . . .	21
2.3	Exemplo de segmentação por região. . . . .	24
2.4	Exemplo de segmentação baseada em contorno. . . . .	24
2.5	Exemplos de particionamento estático . . . . .	25
2.6	(a) subimagem e (b) direções das 5 arestas . . . . .	29
3.1	Pontos sementes para segmentação. . . . .	36
3.2	Agrupamento de pixels em regiões. . . . .	37
3.3	As Figuras <i>a</i> , <i>b</i> , <i>c</i> e <i>d</i> são exemplos de segmentações bem sucedidas. As Figuras <i>e</i> e <i>f</i> mostram um exemplo mal sucedido . . . . .	38
3.4	Tamanho dos três esquemas de partições. . . . .	39
3.5	Imagem dividida em duas partições. . . . .	41
3.6	Esquema 1: duas partições com pixels classificados como borda ou interior. . . . .	41
3.7	Imagem dividida em 3 partições. . . . .	42
3.8	Esquema 2: três partições com pixels classificados como borda ou interior. . . . .	43
3.9	Imagem dividida em cinco partições. . . . .	44
3.10	Esquema 3: cinco partições com pixels classificados como borda ou interior. . . . .	44
3.11	Esquema de partições usado pelo LCPC. . . . .	46
3.12	Detecção de arestas através do gradiente. . . . .	48
3.13	Operador Sobel para computar os gradientes na horizontal e vertical. . . . .	49
3.14	Imagens originais e após a aplicação das máscaras de Sobel. . . . .	50

---

4.1	Imagens usadas como consulta para experimentos na base Wang. . . . .	56
4.2	Dois exemplos de consulta na base CCD e 3 respostas para cada consulta. . . . .	56
4.3	30 imagens de consulta usadas na base Yahoo. . . . .	57
4.4	Curvas de Precision x Recall para o teste de peso na base Wang. . . . .	64
4.5	Curvas de Precision x Recall para os descritores na base Wang. . . . .	65
4.6	Curvas de Precisão x Revocação para os descritores na base CCD. . . . .	67
4.7	Curvas de Precisão x Revocação para os descritores na base Yahoo. . . . .	69
4.8	30 imagens de resposta do descritor LCPC+Aresta na base Yahoo. . . . .	70
4.9	30 imagens de resposta do descritor LCPC na base Yahoo. . . . .	70
4.10	Exemplos de consultas em que o LCPC+Aresta retornou os melhores resultados. .	71
4.11	30 imagens de resposta do descritor LCPC na base Yahoo. . . . .	72
4.12	30 imagens de resposta do descritor BIC na base Yahoo . . . . .	72

# Lista de Tabelas

2.1	Bases comumente usadas para avaliação de descritores de imagens. . . . .	26
2.2	Resumo dos trabalhos relacionados. . . . .	33
3.1	Espaço necessário para o armazenamento do vetor de características. . . . .	45
4.1	$P@10$ para 5 avaliações de 50 consultas. . . . .	59
4.2	$P@20$ para 5 avaliações de 50 consultas. . . . .	59
4.3	$P@30$ para 5 avaliações de 50 consultas. . . . .	59
4.4	Média do tempo de extração de característica em $ms$ . . . . .	60
4.5	Média do tempo de cálculo de similaridade em $\mu s$ . . . . .	61
4.6	Nível de significância em $P@10$ . . . . .	62
4.7	Nível de significância em $P@20$ . . . . .	62
4.8	Nível de significância em $P@30$ . . . . .	62
4.9	Valores de MAP para os descritores na base Wang. . . . .	65
4.10	Validação dos descritores na base Wang. . . . .	66
4.11	Valores de MAP para os descritores na base CCD. . . . .	67
4.12	Validação dos descritores na base CCD. . . . .	68
4.13	MAP e porcentagem de ganho dos descritores na base Yahoo. . . . .	71
4.14	Validação dos descritores na base Yahoo. . . . .	72

# Resumo

Os avanços em tecnologia multimídia ocasionou um grande crescimento da quantidade de imagens digitais, em consequência disso, cresceu também a necessidade de métodos mais eficazes e eficientes para armazenar e recuperar esse conteúdo multimídia. A maioria dos métodos propostos na literatura alcançam altos níveis de eficiência e eficácia (a cima de 70% de precisão), entretanto grande parte delas executam experimentos usando bases de imagens pequenas (menos de 10.000 imagens), previamente classificadas em categorias bem definidas, facilitando assim a tarefa de busca e, conseqüentemente aumentando os níveis de precisão dos descritores avaliados. Por outro lado, quando esses métodos são avaliados em grandes coleções heterogêneas, o nível de precisão é relativamente baixo. Pensando nesse problema, esta dissertação propõe o descritor Local Color Pixel Classification (LCPC), um método baseado em análise local para busca em grandes bases de imagens. A abordagem proposta extrai características de cor, classificando os pixels como borda ou interior, usando o mesmo esquema de classificação do método Border/Interior Pixel Classification (BIC), através de um esquema de particionamento simples, mas muito eficiente e eficaz para incorporar informações espaciais sobre o conteúdo visual da imagem.

Experimentos foram conduzidos usando três bases de imagens, incluindo uma com mais de 100.000 imagens coletadas da Web. Os resultados obtidos mostram que a abordagem proposta é bastante superior quando comparado com outros descritores visuais previamente apresentados na literatura, com ganhos em precisão média de 51% até 105%.

**PALAVRAS-CHAVE:** Recuperação de imagem, Extração de características, Particionamento de imagem, Descritor de imagem

# Abstract

Advances in multimedia technology have increased the availability of digital images. This huge amount of visual information is driving the need for more effective and efficient methods to store and retrieve multimedia content. Most of the proposed techniques in the literature achieve high levels of efficiency and accuracy, however most of them run experiments using small image databases (less than 10,000 images). Most of these image databases are composed of well defined categories, which facilitates the search task resulting in high precision levels. Therefore, when those methods are applied to large and heterogeneous databases, the overall effectiveness is relatively low. This dissertation presents Local Color Pixel Classification (LCPC), a partition-based image descriptor for searching in large and heterogeneous image databases. The proposed descriptor extracts color features of images and it classifies pixels as border or interior through the same classification scheme of BIC Border/Interior Pixel Classification (BIC), and it also uses a simple, but very effective partitioning scheme to incorporate spatial information about their visual content.

Experiments were carried out using three image databases, including one with over 100,000 images collected from the Web. The results obtained show that the proposed descriptor achieves superior retrieval performance when compared to other visual descriptors previously presented in the literature, with gains in *mean average precision* ranging from 51% up to 105%.

KEYWORDS: Image retrieval, Feature extraction, Image decomposition, Image descriptor

# Capítulo 1

## Introdução

A popularidade da World-Wide Web (WWW) e dos dispositivos de aquisição de imagens e conteúdo multimídia como câmeras digitais, celulares, scanners e outros, contribuíram para o rápido crescimento da quantidade de grandes coleções de imagens. De fato, existe uma grande quantidade de informações visuais disponíveis ao usuário na Web. Por exemplo, um milhão de novas imagens são armazenadas diariamente somente no Flickr <sup>1</sup> [1]. A mais popular máquina de busca, Google Image Search <sup>2</sup>, possui mais de 390 milhões de imagens indexadas para a busca.

Atualmente, existem duas abordagens para recuperação de imagens: uma baseada em texto e outra baseada em conteúdo. Na abordagem baseada em texto, as imagens são associadas com palavras-chave usadas para indexar e buscar as imagens. Essa abordagem tem como principal vantagem a capacidade de descrever de forma mais precisa a semântica da imagem. Entretanto, é impraticável descrever cada imagem existente na Web. Além disso, devido à multiplicidade de conteúdos em uma única imagem e da subjetividade de percepção humana e da compreensão, também é difícil fazer exatamente a mesma descrição para a mesma imagem por diferentes usuários.

Considerando os grandes repositórios de imagens presentes na Web que não possuem qualquer descrição textual e as desvantagens da busca baseada em texto, uma outra abordagem tem sido amplamente estudada nos últimos anos: a recuperação de imagens baseada em conteúdo (Content-Based Image Retrieval - CBIR). Em CBIR, as imagens são indexadas usando suas propriedades visuais como cor, textura e forma, representadas por vetores de características. Tais propriedades são extraídas da imagem automaticamente, assim, um grande esforço humano pode ser evitado na construção manual de bases de imagens indexadas por textos.

---

<sup>1</sup><http://www.flickr.com/> acessado em 08/2006

<sup>2</sup><http://images.google.com/> acessado em 08/2006

Muitos métodos em CBIR têm sido propostos ultimamente a fim de resolver o problema de recuperação de imagem, e embora as pesquisas nessa área tenham alcançado alguns resultados promissores [42, 33], o uso de tecnologias de CBIR ainda é bastante limitado em sistemas comerciais, principalmente em sistemas voltados para buscas de imagens em grandes coleções heterogêneas, onde a quantidade e diversidade das imagens aumentam a complexidade do problema.

Dada a imensa quantidade de conteúdo visual disponível, é imprescindível encontrar métodos de CBIR eficientes e eficazes que possam manipular esse tipo de informação. Nesse cenário, é essencial escolher técnicas rápidas e compactas para representar as características visuais sem degradar o desempenho dos sistemas de busca.

Escolher um método eficiente e eficaz para grandes repositórios de imagens não é uma tarefa simples, dada a diversidade de imagens existentes e as diferentes interpretações que elas podem possuir, como por exemplo, a Figura 1.1 mostra uma imagem que pode assumir diferentes significados dependendo da interpretação dos usuários.



Figura 1.1: Imagem com diferentes interpretações.

Se a imagem presente na Figura 1.1 fosse usada como consulta em sistemas de busca, diferentes usuários poderiam ter diferentes interesses, por exemplo, um usuário poderia estar interessado em imagens que possuam pôr do sol, outro poderia estar interessado em barcos, e outros em mar e assim por diante. Dessa forma, é difícil retornar imagens relevantes que atendam aos interesses específicos de cada usuário.

Para tentar atender às expectativas de diversos usuários, os sistemas de busca trabalham com os descritores de imagens que são responsáveis por extrair propriedades visuais (cor, textura e forma) de uma imagem de consulta, e compará-las com as propriedades visuais das imagens presentes na

base, esse processo de comparação é feito através de uma função de similaridade que irá retornar as imagens mais semelhantes à consulta de acordo com a técnica e com as evidências visuais usadas pelo descritor. Entretanto, tratar uma grande coleção de imagens e interpretá-las de acordo com os interesses dos usuários é uma tarefa bastante complexa. A dificuldade desse processo aumenta ainda mais quando consideramos os grandes repositórios de imagens que não possuem qualquer descrição textual. Dessa forma, encontrar um descritor para interpretar uma imagem de consulta e retornar respostas semelhantes com um nível de qualidade aceitável por parte dos usuários torna-se um desafio na área de CBIR.

## 1.1 Motivação

Dada a quantidade de conteúdo multimídia disponível para os usuários, a demanda por métodos capazes de processar e recuperar as imagens em sistemas de busca do mundo real é extremamente grande. Muitas pesquisas nessa área já foram realizadas, entretanto a eficácia dos descritores de busca para coleções heterogêneas ainda é baixa [46]. Alguns sistemas são projetados para domínios específicos, dessa forma é possível aproveitar as vantagens que imagens pertencentes a categorias específicas possuem, como por exemplo, sistema de detecção facial possuem características específicas inerentes a esse tipo de imagem o que pode facilitar o seu reconhecimento.

Por outro lado, métodos que possam interpretar imagens pertencentes a qualquer domínio tornam-se muito mais complexos, havendo assim, a necessidade de muitas pesquisas nessa área. Por esse motivo, esta dissertação concentra-se no desenvolvimento de um método capaz de realizar busca em grandes coleções de imagens heterogêneas.

A maioria dos descritores existentes na literatura alcançam ótima eficácia quando realizam buscas em bases bem comportadas e pequenas (menos de 10 mil imagens) [19, 48, 24] De acordo com os experimentos realizados em [46], em geral, a eficácia dos descritores de imagens é baixa em grandes bases heterogêneas, enquanto que nas bases pequenas, a maioria deles possuem altos valores de precisão. Recentemente, alguns autores começaram a estudar o problema de CBIR em grandes coleções de imagens [23, 16, 45], entretanto, esses trabalhos são focados, principalmente em aspectos de eficiência, como, por exemplo, tempo de indexação, tempo de consulta e espaço para armazenamento, e a maioria desses trabalhos não avaliam a eficácia de suas técnicas por melhor de usuários reais. Consideramos que a avaliação com usuários reais é uma forma confiável de realizar testes dos descritores, pois eles refletem o uso de sistemas de busca do mundo real.



## 1.2 Justificativa

A maioria dos sistemas de CBIR é baseada em extração de propriedades de imagens como cor, textura ou forma. Cor é a evidência mais comumente usada porque ela não possui um alto custo computacional para ser extraída. Ela também apresenta resultados aceitáveis em termos de precisão quando comparada a outras abordagens avançadas que usam textura ou forma.

A evidência de cor pode ser extraída através de três principais abordagens: (1) abordagem global [40, 62]; (2) abordagem baseada em partições estáticas [50, 55] e (3) técnicas de partições usando segmentação [30, 53]. Cada uma destas técnicas possuem distintos compromissos entre a complexidade do algoritmo usado para extração das características, quantidade de espaços necessário para representar as características visuais e nível de eficácia.

Quando desenvolvemos um descritor, é importante definir previamente quais as evidências a serem utilizadas e qual o tipo de técnica usada para a extração das propriedades da imagem. Como o objetivo deste trabalho é voltado para busca em grandes coleções de imagens. É necessário o uso de características e métodos que possuam boa acurácia e um bom desempenho, de forma que, seja viável sua utilização em grandes bases. Dessa maneira, este trabalho foca as suas pesquisas na extração de propriedades de cor, pois o processo para sua extração possui um baixo custo computacional. A fim de extrair informações sobre como as diferentes cores estão distribuídas pela imagem e aumentar a eficácia do método proposto, analisamos a características de cor usando métodos de partições estáticas, pois eles possuem alto desempenho e boa eficácia.

## 1.3 Objetivos

O objetivo desta dissertação consiste em definir e desenvolver um método de busca de imagens com boa eficácia visando grandes coleções de imagens heterogêneas. Para tanto, utilizamos evidência de cor que possui baixo custo computacional e técnicas de particionamento a fim de aumentar o nível de precisão do descritor proposto.

A principal contribuição deste trabalho, é o desenvolvimento de um descritor simples que pode ser viável em grandes bases de imagens. Adicionalmente, propomos uma variação do descritor proposto com o objetivo de incluir características de forma a fim de aumentar sua acurácia.

Com o propósito de encontrar um melhor esquema de particionamento, este trabalho propõe um estudo prévio de extração de cor usando análise local a partir de segmentação e três tipos de esquemas de particionamento estático. Para mostrar a superioridade do descritor proposto,

testes foram realizados em bases de diferentes tamanhos, comparando-o com diferentes descritores comumente usados na literatura.

## 1.4 Organização do trabalho

Esta dissertação é organizada da seguinte forma, o próximo Capítulo apresenta a fundamentação teórica e os conceitos importantes para o entendimento do trabalho, além de incluir os trabalhos relacionados citados na literatura, destacando principalmente aqueles que utilizam abordagem baseada em partições. No Capítulo 3, o descritor proposto neste trabalho é apresentado em detalhes, assim como os esquemas de partições usados e o estudo prévio realizado a fim de encontrar a melhor forma de decompor a imagem. Na sequência, o capítulo 4 apresenta os experimentos realizados e discute os resultados obtidos. Por fim, o Capítulo 5 apresenta as conclusões e os trabalhos futuros.

## Capítulo 2

# Fundamentação Teórica e Trabalhos Relacionados

A recuperação de conteúdo multimídia é uma área que está sendo bastante explorada e que possui muitas aplicações em domínios específicos e em sistemas de domínio aberto. Para realizar a busca de imagens, os sistemas de busca usam como agente principal os descritores de imagens, eles são os responsáveis por reconhecer o conteúdo visual de uma determinada imagem, para que possam encontrar imagens semelhantes. O reconhecimento visual pode ser feito usando propriedades como cor, textura ou forma, através de três abordagens que serão detalhadas na próxima seção.

Este trabalho explora a recuperação de imagens baseada em conteúdo, fazendo parte da área de content-based image retrieval (CBIR). Sendo assim, neste capítulo são apresentados os conceitos relacionados aos descritores de imagens e suas características, sendo esses necessários ao entendimento desta dissertação.

### 2.1 Descritor de imagens

O principal componente em um sistema de busca de imagens é o descritor de imagens. Ele é responsável por fazer o reconhecimento do conteúdo visual de uma determinada imagem e compará-lo ao conteúdo visual existente em todas as imagens pertencentes a uma determinada base para que possa retornar imagens semelhantes, e possui dois componentes principais: (1) o extrator de características e (2) a função de distância. O componente (1) é responsável por extrair um conjunto de características de uma determinada imagem (cor, textura e forma, por exemplo),

enquanto que o componente (2) diz respeito ao cálculo de similaridade entre duas imagens tendo como base características das mesmas [59]. Mensurar o quanto duas imagens são semelhantes pode ser feito através de funções de distância como a L2 também conhecida como distância Euclidiana, L1 ou *city block*, dLog [54] e outras.

O funcionamento do processo de busca através do descritor de imagens é realizado da seguinte forma:

- o descritor extrai as características visuais, como cor textura ou forma, de uma imagem passada como consulta por um usuário. Essas propriedades visuais são processadas por meio de técnicas e algoritmos de processamento de imagem e são armazenadas em um vetor de características também chamado de assinatura da imagem.
  - o descritor de imagem realiza uma busca por vetores de características semelhantes ao vetor da imagem de consulta em uma determinada base, onde os vetores de características correspondentes a todas as imagens pertencentes à base foram previamente calculados.
  - inerente ao descritor está a função de similaridade, responsável por calcular o quanto um vetor de características de uma imagem A é similar ao vetor da imagem B. Com essa informação, é possível realizar uma ordenação das imagens em ordem de similaridade, ou seja, as imagens mais semelhantes à imagem de consulta, segundo o descritor, são retornadas para o usuário.
- A Figura 2.1 ilustra processo realizado pelo descritor para calcular a similaridade entre duas imagens.

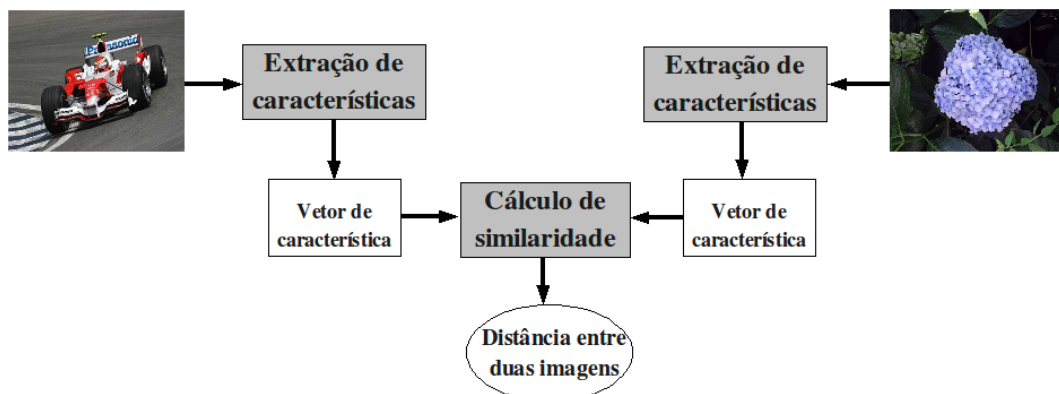


Figura 2.1: Processo realizado por um descritor de imagem.

## 2.2 Características de imagens

O descritor realiza a interpretação do conteúdo visual de uma imagem, extraindo evidências de cor, textura ou forma. Cada característica descreve a imagem segundo algum tipo de propriedade e possui desempenho diferenciado no seu processamento. O que determina o uso de cada uma delas ou a combinação de evidências, é o tipo de aplicação. Em geral, a cor é a mais usada para realizar busca de imagens devido o seu processo de extração possuir um baixo custo computacional.

A cor é a propriedade visual dominante em uma imagem, ela é comumente usada em sistemas de CBIR, pois ela é a primeira propriedade reconhecida pela visão humana. A cor pode ser representada por um histograma de cor [14], ou outras técnicas como *Color Moments* [30]. Como a extração de cor pode ser realizada por um processo simples, por exemplo histograma de cor, é possível desenvolver descritores bastante eficientes usando características de cor, entretanto, a eficácia pode ser comprometida, como por exemplo, no caso de uma imagem de consulta, possuir uma flor vermelha, serão consideradas relevantes, de acordo com o descritor, apenas imagens que possuam a cor vermelha, independente de possuírem flor. Além disso os descritores de cor são bastante sensíveis a variação de luz ambiente, contraste e brilho, pois a variação dessas três propriedades em uma mesma imagem pode facilmente confundir o descritor de cor, fazendo com que ele interprete de forma diferenciada duas imagens com a mesma cena, mas que possuam variação de brilho diversificada. A Figura 2.2 mostra duas imagens com cenas idênticas, sendo que elas possuem a variação de brilho como única diferença.

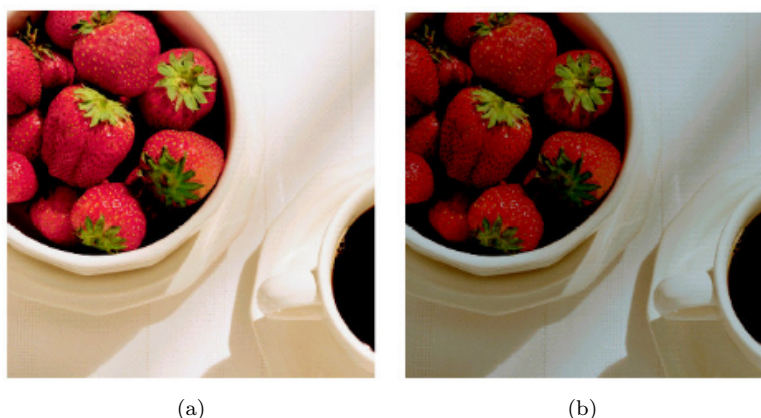


Figura 2.2: Imagem (a) com tons de cores mais claros e imagem (b) com tons escuros.

Verificamos que as duas imagens da Figura 2.2, com exceção da variação de brilho, são exatamente iguais, entretanto, para o descritor que extrai apenas propriedades de cor elas podem ser consideradas diferentes. Por esse motivo, descritores que usam apenas evidência de cor tendem a se

confundir mais facilmente em grandes bases de imagens pertencentes às mais diversas categorias.

Um descritor de imagem também pode extrair propriedades de forma com o objetivo de superar as desvantagens inerentes dos descritores que utilizam apenas cor, ou seja, fazendo com que o descritor possa identificar as formas dos objetos que compõem a imagem, minimizando a confusão causada pela variação de luz, brilho e contraste dos descritores de cor. A evidência de forma, pode ser extraída através da detecção de borda que consiste de uma técnica de processamento de imagem para determinar pontos de uma imagem onde ocorre mudança súbita da intensidade luminosa, preservando assim importantes propriedades estruturais dos objetos presentes na imagem. Após a detecção de borda, é possível obter o contorno dos objetos presentes na figura, entretanto esse processo pode ser prejudicado por bordas falsas criadas por sombras dos objetos, sobreposição, oclusão e textura dos diferentes objetos contidos na imagem. Isso pode ser amenizado através de um processamento prévio na imagem. Por outro lado, o processo completo para a obtenção do contorno do objeto presente na imagem é muito custoso, tornando-o inviável empregá-lo em sistemas de busca em coleções heterogêneas. Por esse motivo, é mais comum usar evidências de forma em domínios específicos como imagens médicas [41, 28], onde não há muita exigência com relação ao desempenho, mas sim com sua eficácia.

Uma imagem também pode ser descrita por sua propriedade de textura, que pode ser conceituada como a propriedade que contém informações sobre a distribuição espacial dos níveis de cinza de uma região da imagem [38, 60]. A extração de textura de uma imagem, assim como a propriedade de forma, exige um custo computacional maior que a extração de propriedade de cor. A característica de textura é usada em classificação de imagens, como por exemplo, aplicações de sensoriamento remoto [15], pois através da extração de textura da imagem é possível obter padrões e objetos homogêneos, obtendo como resultado final uma imagem com todos os pixels classificados como pertencentes ou não a uma determinada região.

Somente a extração das características da imagem não é capaz de descrever como essas propriedades estão distribuídas pela imagem. Por exemplo, um descritor de cor pode extrair as cores dominante na imagem de forma geral, mas não será capaz de distinguir como elas estão distribuídas por suas diferentes regiões. Nesse caso são necessários métodos que tenham a habilidade de realizar uma análise local na imagem. Sendo assim, existem três abordagens principais para extrair as propriedades visuais de uma imagem: abordagem global, abordagem baseada em partições estáticas e abordagem usando segmentação, as três serão detalhadas na próxima seção.

### 2.2.1 Abordagem global

Em abordagens globais, o conteúdo visual da imagem é analisado em sua totalidade, sem considerar partições ou regiões. A evidência de cor é extraída da imagem globalmente. Essa abordagem é bastante simples, pois analisa a imagem em apenas uma única iteração. Ela possui como vantagens, algoritmos de extração de características rápidos e a representação das características é extremamente compacta, dessa forma, o vetor de características necessita de pouco espaço para ser armazenado. No entanto, essa abordagem não possibilita ao descritor descrever mais precisamente as relações espaciais entre os objetos presentes na imagem, comprometendo assim a eficácia do descritor.

Essa abordagem é bastante empregada na literatura, principalmente, devido a sua simplicidade e bom desempenho [44, 34].

### 2.2.2 Abordagem baseada em segmentação

Como alternativa aos poucos detalhes sobre o conteúdo visual da imagem oferecidos pela abordagem global, a imagem pode ser melhor detalhada através da segmentação.

Segmentação é o processo de dividir uma imagem em múltiplas regiões ou objetos a fim de facilitar sua análise. Essa abordagem é muito usada para localizar objetos e formas em imagens, o resultado é um conjunto de regiões que possuem características simples. Desse modo, cada um dos pixels em uma determinada região é similar de acordo com alguma característica ou propriedades como cor, textura ou intensidade. Em [52] os métodos de segmentação são divididos em três categorias como bordas, regiões ou no conhecimento global a respeito da imagem. Em [8] são considerados apenas dois tipos de segmentos principais, bordas e regiões que podem ser separadas por diferentes critérios. Em segmentação baseada em regiões a imagem é dividida em regiões que possuem características similares [26]. A Figura 2.3 mostra uma imagem dividida em segmentação por região.

Na segmentação representada na Figura 2.3, o processo é iniciado com um conjunto de pixels sementes, aos quais vão sendo adicionados mais pixels que obedecem um critério estabelecido, que pode ser a cor do pixel de uma determinada região de acordo com um limiar, tal método é conhecido como segmentação por crescimento de região. Após esse processo, a imagem é segmentada de acordo com suas diferentes regiões, por exemplo, podemos ter interesse em separar uma área da imagem que represente o rio, floresta dentre outros. A partir dessa separação, é possível descrever separadamente o que está presente em diferentes regiões, isso nos permite uma maior flexibilidade

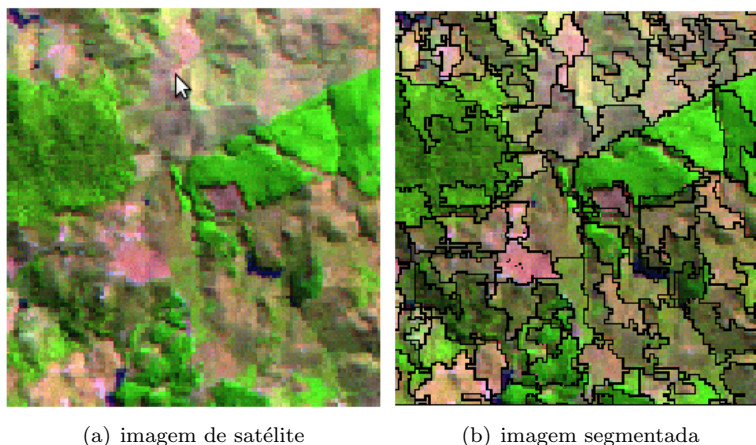


Figura 2.3: Exemplo de segmentação por região.

e mais detalhes ao descrever o conteúdo visual da imagem.

A imagem também pode ser segmentada através de detecção de borda. Uma borda é um contorno entre objetos ou entre um objeto e o fundo, indicando o limite entre objetos sobrepostos. Os algoritmos de segmentação para detecção de borda são baseados no processo de localização e realce dos pixels de borda, aumentando o contraste entre a borda e o fundo. A operação de segmentação baseada em bordas é realizada através do processo de limiarização que define um valor ou uma função para separar elementos de interesse na imagem. Frequentemente, a imagem original é convertida em imagem binária pelo processo de limiarização, atribuindo-se 0 aos pixels com tons de cinza menores que um determinado limiar e 1 aos maiores. Esse método é um dos mais simples no processo de segmentação baseada em detecção de borda.

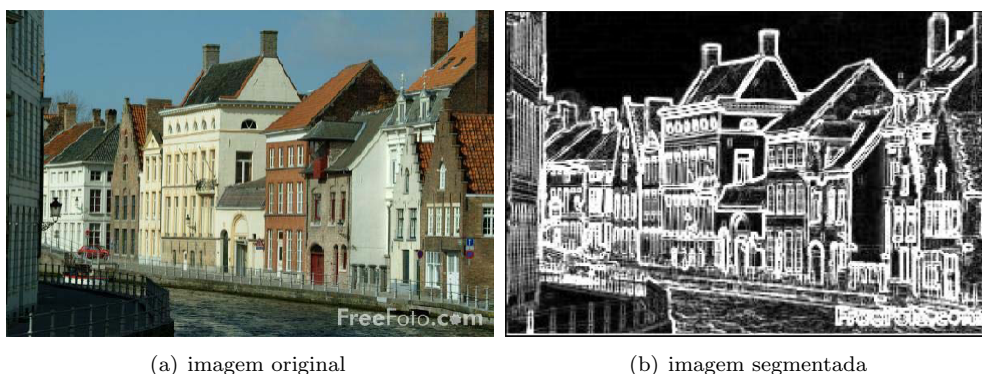


Figura 2.4: Exemplo de segmentação baseada em contorno.

A segmentação baseada em detecção de borda, como mostra a Figura 2.4, divide a imagem com base em alterações bruscas de intensidade que permite descrever a imagem através de seu conteúdo semântico. Dessa forma, o descritor pode descrever os objetos presentes na imagem em alto nível.



Essa abordagem é usada em aplicações que requerem alta eficácia, no entanto, ela pode ser inviável em grandes repositórios de imagens, onde a exigência por desempenho é imprescindível. A única forma de aplicá-la em sistemas de busca em grandes coleções, é simplificando o processo de segmentação, assim, ganha-se desempenho, mas perde-se eficácia.

### 2.2.3 Abordagem baseada em partições fixas

Uma maneira mais eficiente de dividir a imagem e introduzir algum tipo de informação espacial sobre o relacionamento dos objetos que a compõem, é baseada em partições fixas, ou seja, decompor a imagem de acordo com um esquema pré-definido como mostra a Figura 2.5.

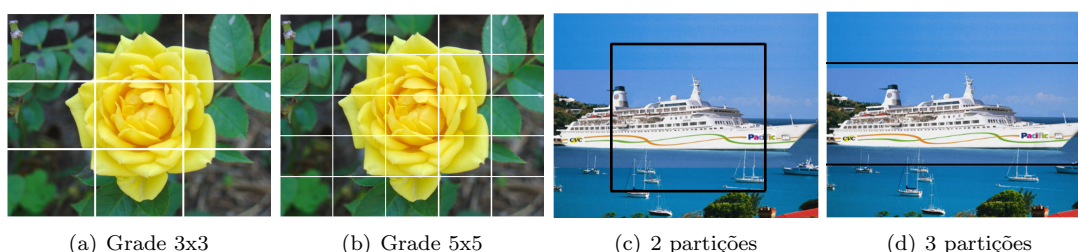


Figura 2.5: Exemplos de particionamento estático

A Figura 2.5 mostra vários tipos de particionamento fixo, eles diferem no tamanho das partições, na quantidade e na forma de particionamento. Após o processo de divisão, cada parte da imagem é descrita individualmente, com o objetivo de extrair informações locais, aumentando assim o nível de precisão do descritor. Em relação a abordagem global, essa abordagem é muito mais precisa, pois introduz importantes informações espaciais sobre o conteúdo da imagem, entretanto, a quantidade de espaço requerido para o armazenamento do vetor de características cresce de acordo com o número de partições. O desempenho no cálculo de similaridade entre dois vetores também pode ser afetado devido ao número de partições que pode ocasionar um aumento no tamanho do vetor de características. Dessa forma, o ideal é que se mantenha reduzida o número de partições, balanceando assim a relação o custo/benefício do uso dessa abordagem.

## 2.3 Características das bases de imagens

Para realização de testes de descritores, se faz necessário o uso de bases de imagens. Algumas bases de imagens como *Common Color Dataset (CCD)*, *Wang Database*<sup>1</sup> [4] e *Caltech256*<sup>2</sup> [3]

<sup>1</sup><http://www.ee.cityu.edu.hk/~mirror/> acessado em 04/2011

<sup>2</sup>[http://www.vision.caltech.edu/Image\\_Datasets/Caltech256/](http://www.vision.caltech.edu/Image_Datasets/Caltech256/) acessado em 04/2011

encontram-se disponíveis publicamente. Geralmente elas possuem um conjunto de imagens que estão relacionadas de acordo com alguma propriedade presente na imagem. Por exemplo, para avaliação de descritores que extraem características de textura, em geral, utiliza-se base em que as imagens são ricas em texturas. Para busca de qualquer tipo de imagem, geralmente, são usadas coleções de imagens heterogêneas. Em aplicações de domínios específicos como reconhecimento facial, são usadas bases contendo várias imagens de faces de tamanhos variados.

Na literatura, existem bases onde as imagens são inicialmente selecionadas e classificadas, montando assim um *benchmark* de avaliação, onde para cada imagem usada como consulta, o conjunto de respostas relevantes é previamente definido, assim, as medidas de avaliações são extraídas automaticamente, facilitando o processo de avaliação de descritores. Na Tabela 2.1 são apresentados alguns *benchmarks* de diferentes domínios comumente usados para a avaliação de descritores.

Nome da Base	Total de Imagens	Descrição
Wang	1.000	imagens heterogêneas coloridas divididas em 19 categorias
CCD	5.466	imagens heterogêneas coloridas
IRMA 10000	10.000	imagens de radiografia para aplicações médicas
Caltech-101	9144	imagens coloridas, contendo objetos divididos em 101 categorias
Caltech-256	30.607	imagens coloridas, contendo objetos divididos em 256 categorias
Brodatz	111	base de imagens de textura
UCID	1.338	imagens heterogêneas
Corel5K	5.000	imagens coloridas divididas em 50 categorias
Scene Dataset	4.485	imagens em tons de cinza divididas em 15 categorias

Tabela 2.1: Bases comumente usadas para avaliação de descritores de imagens.

As bases mostradas na Tabela 2.1 são comumente usadas em buscas de imagens, algumas com domínio específico como é o caso da IRMA 10000 [25] usada em aplicações médicas e a Brodatz<sup>3</sup> [2] que contém imagens ricas em textura para avaliações de descritores de textura. Como as bases Wang [12], CCD [13], Corel5K [17] e UCID [49] são formadas por imagens coloridas e heterogêneas, estas são normalmente usadas para avaliar descritores de cor que tenham o propósito de serem empregados em sistemas de busca para grandes repositórios na Web, ou seja, descritores com o objetivo de reconhecer imagens de qualquer domínio. A coleção Scene Dataset [29] também é aplicada em busca, entretanto, ela contém imagens em tons de cinza divididas em 15 categorias, sendo a maioria de cenas naturais como montanhas, florestas, construções, escritórios e outros, assim, os descritores que usam característica de forma podem ser avaliados nessa base a fim de tentar reconhecer os padrões de forma para essas diferentes cenas.

Outro ponto a ser observado, é o tamanho das bases, geralmente, menores que 10 mil imagens, são bases pequenas. Sendo assim, nesse tipo de ambiente o descritor avaliado tende a atingir altos

<sup>3</sup><http://www.cs.ucr.edu/~bcampana/texture/> acessado em 04/2011

níveis de precisão, entretanto, quando o descritor é usado em grandes coleções heterogêneas, o que comumente acontece é uma queda considerável em seu nível de precisão, pois em um ambiente maior e mais diversificado o método pode se confundir mais facilmente, aumentando sensivelmente a quantidade de imagens irrelevantes retornadas pelos sistemas de buscas.

Atualmente, estão sendo desenvolvidos métodos de busca [47, 63, 22] para serem empregados e avaliados em grandes coleções de imagens, geralmente, essas bases são coletadas do Flickr e possuem mais de um milhão de imagens. Esses métodos tem como objetivo realizar buscas em repositórios da Web, entretanto a precisão deles ainda é baixa para que eles possam ser empregados em ambientes reais, por esse motivo ainda são necessárias muitas pesquisas nessa área a fim de encontrar descritores boa precisão em grandes coleções, de forma que satisfaçam as exigências e necessidades dos usuários.

## 2.4 Trabalhos Relacionados

Na literatura existe uma ampla quantidade de descritores que se baseiam em diferentes linhas de desenvolvimento. Por exemplo, os mais simples são os descritores semelhantes aos propostos em [57, 58, 53] que extraem características de cor utilizando técnicas simples como os histogramas de cor global ou local. Outra linha de desenvolvimento é baseada em *bag of visual word* [47, 63, 22], métodos inspirados em métodos de recuperação de documentos textuais que, além de extrair as características visuais, nos permite indexar a base de imagens para melhorar o desempenho, entretanto tais métodos ainda estão em fase de desenvolvimento e por enquanto não alcançaram uma boa acurácia para que possam ser usados no ambiente Web.

Neste capítulo são apresentados alguns trabalhos relacionados, inclusive os baselines usados neste trabalho, focando principalmente em descritores que utilizam análise local para a extração de características. Consideramos alguns descritores de imagens existentes na literatura de acordo com alguns critérios como acurácia e tamanho da base utilizada para a validação. A seguir, é realizada uma breve descrição dos seguintes descritores: Color Layout Descriptor (CLD), Scalable Color Descriptor (SCD) e Edge Histograma Descriptor (EHD), os três pertencem ao padrão MPEG-7 [37] e foram usados como baselines, também incluímos os descritores Fuzzy Color Texture Histogram (FCTH) e Color and Edge Directivity Descriptor (CEDD) como baselines, pois eles se mostraram superiores aos três descritores pertencentes ao MPEG-7 [11], além destes, também descrevemos nesta seção outros descritores comumente usados na literatura, tais como, Cell/Color Histograms [55], Color Blocks [31] e Color Bitmap [35].

### MPEG-7: color layout descriptor - CLD

O color layout descriptor ou CLD [37] foi desenvolvido para capturar a distribuição espacial de cor de uma região arbitrária da imagem. No primeiro passo, a imagem é dividida em um grid de 64 blocos, em seguida a cor dominante de cada bloco é extraída, geralmente, através da média de cores ou algum outro método. Na sequência, cada o componente do espaço de cor YCrCb é transformado usando a transformada do coseno (DCT), gerando assim três conjuntos de coeficientes. Finalmente, é associada um peso para cada coeficiente, produzindo assim, um vetor de características com informações a respeito das cores predominante em cada bloco. O histograma gerado é bastante compacto, proporcionando assim uma alta eficiência no momento de realizar a busca. Para o cálculo de similaridade entre duas imagens, considere dois conjuntos de coeficientes extraídos, sendo um extraído da imagem A e outro da imagem B:  $A = \{DY, DCb, DCr\}$  e  $B = \{DY', DCb', DCr'\}$ . A distância entre as duas imagens será computada como:  $D(A, B) = \sqrt{\sum_i w_{yi}(DY - DY')^2} + \sqrt{\sum_i w_{bi}(DCb_i - DCb'_i)^2} + \sqrt{\sum_i w_{ri}(DCr_i - DCr'_i)^2}$ .

Aqui,  $(DY_i, DCb_i, DCr_i)$  representa o  $i^{th}$  coeficiente DCT dos respectivos componentes de cor. Os pesos são atribuídos apropriadamente, sendo que um componente com frequência mais baixa, recebe um peso maior.

### MPEG-7: scalable color descriptor - SCD

O scalable color descriptor ou SCD [37] usa o espaço de HSV para gerar um histograma de cor. Esse descritor utiliza quantização fixa e uniforme nesse espaço de cor. Os histogramas são codificados usando a transformada de Haar que facilita uma representação é escalável em termos de números de elementos e números de bits usado na representação do histograma. A precisão no momento da busca aumenta de acordo com a quantidade de bits usado na representação, entretanto, experimentos realizados em [37] demonstraram que bons resultados foram obtidos utilizando 64 bits na representação. O SCD usa, como medida de similaridade, a métrica L1 ou distância de Hamming para a comparação dos histogramas.

### MPEG-7: edge histogram descriptor - EHD

O edge histogram descriptor ou EHD [37] captura a distribuição de espacial de 5 tipos de arestas, sendo 4 direcionais e uma não direcional, de forma semelhante ao CLD, o EHD extrai características locais da imagem, porém, particionando-a em blocos de  $2^n \times 2^n$ , usualmente 16 pixels cada. As aresta são agrupadas em 5 categorias: vertical, horizontal, 45° diagonal, 135° diagonal, e

sem direção. Inicialmente, a imagem é dividida em blocos, e, para cada, bloco é aplicado um ou mais filtros para a detecção das arestas, sendo que para cada pixel de aresta é definido qual o tipo de aresta. Um exemplo de subimagem e as possíveis arestas é ilustrado na Figura 2.6.

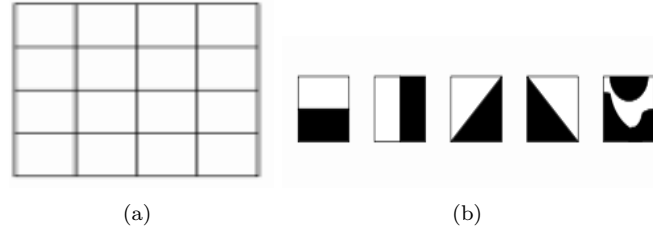


Figura 2.6: (a) subimagem e (b) direções das 5 arestas

A acurácia do EHD pode ser melhorada, combinando-o com descritores de cor.

#### Color and edge directivity descriptor - CEDD

O CEDD [9] inclui informações de textura, forma e cor. Nesse descritor, podemos considerar a imagem de maneira global e a imagem dividida em blocos para extrair informações locais. Para a extração de informações de característica de forma, aplica-se o descritor EHD com o objetivo de construir um histograma de de 6 elementos, sendo que 5 elementos correspondem aos tipos de arestas encontrados na imagem e outro correspondente a arestas que não se encaixam em nenhum dos 5 tipos. Entretanto, o EHD é aplicado de maneira diferente do descritor original, sendo que a magnitude das arestas é normalizada da seguinte forma:  $m_v = \frac{m_v}{MAX}$ ,  $m_h = \frac{m_h}{MAX}$ ,  $m_{d-45} = \frac{m_{d-45}}{MAX}$ ,  $m_{d-135} = \frac{m_{d-135}}{MAX}$ ,  $m_{nd} = \frac{m_{nd}}{MAX}$ , onde  $MAX = \max\{m_v, m_h, m_{d-45}, m_{d-135}, m_{nd}\}$  significa a magnitude máxima das arestas.

Então, dado um limiar, uma aresta pode se classificada em uma das 5 arestas previamente definidas, caso contrário, ela irá pertencer ao último elemento que corresponde a nenhum tipo de aresta.

Para a característica de cor, cada bloco é processado no espaço de cor HSV de acordo com os tipos de arestas previamente definidos. O primeiro passo é mapear cada aresta do bloco em um conjunto de 10 cores previamente definidas para gerar o histograma: preta, cinza, branca, vermelha, laranja, amarela, verde, ciano, azul e magenta, essas cores são mapeadas de acordo com o descritor Haar Wavelet usando as 20 regras do método fuzzy-linking [9]. No segundo passo, esse histograma é expandido para um histograma de cor de 24 elementos usando *Coordinate Logic Filters* (CLF) para as arestas verticais em todos os três canais HSV: o componente H é dividido em 8 áreas: vermelha para laranja, laranja, amarela, verde, cyan, azul, magenta e azul para vermelha;

o componente S é dividido em duas regiões fuzzy definindo os tons de cores baseadas no tom branco; e por fim, o canal V é dividido em três regiões: uma que define quando o pixel será preto e as outras para definir quando a cor terá tons de cinza. Baseando-se nessas divisões, um conjunto de 4 regras fuzzy são aplicadas, transformando as 10 cores iniciais em um histograma de 24 cores que englobam as cores preta, cinza, branca, vermelha escura, vermelha, vermelha clara, laranja escura, laranja, laranja clara, amarela escura, amarela, amarela clara, verde escura, verde, verde clara, ciano escura, ciano, ciano clara, azul escura, azul, azul clara, magenta escura, magenta e magenta clara. A informação de cor processada para cada tipo de aresta é igual a um histograma de  $6 \times 44 = 144$  elementos.

Nos experimentos realizados em [11], esse método mostrou-se superior aos descritores do MPEG-7, por esse motivo ele foi escolhido como baseline neste trabalho.

#### **Fuzzy color and texture histogram - FCTH**

O descritor Fuzzy Color and Texture Histogram (FCTH) [10], semelhante ao CEDD, que captura características de textura, forma e cor. Entretanto, para as informações de forma ele não usa os cinco filtros do EHD, mas sim as altas frequências da transformada de Haar Wavelet, desse modo, as características de forma e textura são representadas em um histograma de 8 elementos. O processo restante para a extração de cor é semelhante ao descritor CEDD e leva em consideração cada uma das características de forma, como a característica de cor é representada em um histograma de 24 elementos, o vetor de característica do FCTH, considerando cor, forma e textura, resulta em  $8 \times 24 = 192$  elementos. Semelhante ao descritor CEDD, o FCTH também mostrou-se superior aos descritores do MPEG-7 de acordo com os testes realizados em [11], sendo por esse motivo, escolhido como baseline

#### **Border/interior pixel classification - BIC**

O método BIC [54] é um descritor baseado na abordagem global que apresentou bons resultados em relação a eficiência e eficácia [46]. Esse método mostrou-se superior a outros descritores existentes na literatura [46] mesmo quando empregado em grandes coleções heterogêneas. BIC usa o espaço de cor RGB quantizado em 64 cores e classifica os pixels como borda ou interior. Um pixel é considerado como borda se pelo menos um de seus 4 vizinhos (pixel superior, inferior, esquerdo e direito) possuir um valor de cor diferente, caso contrário, o pixel é considerado de interior. Assim, dois histogramas globais são criados, um considerando pixels de interior e outro de borda. Essa

técnica de classificação nos dá a idéia de como os pixels estão distribuídos pela imagem como um todo. BIC também possui uma rápida função de distância denominada de  $dLog$ . A função  $dLog$  calcula a diferença entre o logaritmo dos elementos presentes em um histograma a fim de reduzir o efeito negativo introduzido quando poucos elementos possuem alta frequência e outros possuem baixa frequência. Esses valores com alta frequência tendem a dominar a diferença entre dois histogramas e, em geral, tais valores pertencem ao plano de fundo da imagem, o qual engloba a maior parte da imagem, mas não determinam a sua semântica na maioria dos casos. A função  $dLog$  minimiza essa distorção e é definida como:

$$dLog(a, b) = \sum_{i=0}^{i < M} |f(a[i]) - f(b[i])| \quad (2.1)$$

$$f(x) = \begin{cases} 0, & \text{se } x = 0 \\ 1, & \text{se } 0 < x \leq 1 \\ \lceil \log_2 x + 1 \rceil, & \text{caso contrário} \end{cases} \quad (2.2)$$

Na equação anterior,  $a$  e  $b$  são dois histogramas com  $M$  elementos, onde cada elemento significa uma cor. O valor  $a[i]$  representa o  $i^{th}$  componente do histograma  $a$  e  $b[i]$  representa o  $i^{th}$  componente do histograma  $b$ . Usamos a função  $dLog$  porque ela mostrou-se ser eficiente e mais eficaz que outras métricas de distância existentes na literatura, como L1, L2 ou distância de Hamming [54, 46]. Ela também nos permite gerar histogramas bastante compactos, pois os seus elementos são normalizados em um intervalo de  $[0, 9]$  o que possibilita minimizar o efeito negativo introduzido por cores que possuam altas frequências e que, na maioria das imagens, pertencem ao seu plano de fundo, agregando pouco informação semântica a respeito da imagem. A normalização permite que cada elemento do histograma seja armazenado em apenas 4 bits, permitindo que os vetores de características ocupem pouco espaço para armazenamento.

### Cell/Color Histograms

A abordagem Cell/Color Histograms [55] é baseada em um grid de células a fim de extrair a distribuição de cor de cada célula individualmente, mas diferente das abordagens que usam os histogramas de cor convencionais, o descritor Cell/Color Histograms gera um histograma das cores realmente presentes na imagem, representando como as cores estão distribuídas entre as células. Esse descritor baseia-se na idéia de que o número de cores realmente presente na imagem é fre-

quentemente baixo. Esse descritor foi proposto usando uma base de avaliação com 20 mil imagens coletadas do CD corel stock photos. Descrevemos esse descritor como trabalho relacionado devido ao seu simples esquema de particionamento, e por ele realizar uma análise local da distribuição de cores, processo semelhante a abordagem proposta neste trabalho.

### **Color Bitmap**

O descritor Color Bitmap [35] é uma técnica simples que analisa a informação de cor globalmente e por meio de particionamento fixo. O método calcula a média e o desvio padrão dos valores da imagem para cada canal RGB separadamente de forma global, em seguida, a imagem é particionada em células, e a média e o desvio padrão de cada canal em cada célula são calculados separadamente. Esse descritor foi proposto em [35] e avaliado em em três bases, uma contendo 800 imagens, outra com 470 imagens, e outra com 10 mil imagens, sua técnica particiona a imagem em regiões fixas a fim de aumentar a acurácia do método, objetivo semelhante ao descritor proposto nesta dissertação.

### **Color Blocks**

O descritor Color Blocks [31] trabalha com análise de cor através de informações locais, dividindo a imagem em blocos, sendo que cada bloco recebe pesos diferentes de acordo com a sua localização. A técnica do color blocks é bastante semelhante aos descritores citados anteriormente e a várias outras técnicas existentes na literatura que usam particionamento estático para a extração de características da imagem. O primeiro passo do algoritmo é dividir a imagem em diferentes sub-blocos, em seguida, a fim de recuperar as imagens que tenham o conteúdo da consulta localizado em diferentes áreas, cada sub-bloco da consulta é comparado com cada um dos blocos da imagem que está sendo comparada para realizar o casamento do histograma. A intenção da técnica é recuperar imagens que possuam o mesmo objeto da imagem de consulta localizado em diferentes áreas. Semelhante aos outros descritores citados neste trabalho, o Color Blocks também foi avaliado em uma base pequena, e os experimentos realizados em [31] mostram que ele foi superior ao método que emprega um histograma simples de cor.

Nesta dissertação também usamos informações de cor, sendo extraídas através de análise local a fim de aumentar a precisão do método proposto, adicionalmente realizamos testes com características presentes na borda dos objetos que compõem a imagem. Ao observarmos os trabalhos relacionados citados nesta seção, verificamos que a maioria possui uma característica em comum



Descritor	Evidência	Base Usada nos Experimentos	Particionamento	Função de Similaridade
MPEG-7: CLD	cor	CCD	fixo	Distância Euclidiana
MPEG-7: SCD	cor	CCD	global	City Block
MPEG-7: EHD	aresta	CCD	fixo	City Block
CEDD	cor	Wang, mais 15 mil imagens	fixo	Tanimoto Coeficiente
FCTH	cor	Wang, mais 15 mil imagens	fixo	Tanimoto Coeficiente
BIC	cor	20 mil imagens do CD Corel	global	dLog
Cell histograms	cor	20 mil imagens do CD Corel	fixo	função própria
Color Bitmap	cor	bases de 800, 470 e 10.000 imagens	fixo	Hamming e Euclidiana
Color Blocks	cor	base pequena	fixo	Matriz de similaridade

Tabela 2.2: Resumo dos trabalhos relacionados.

que é análise local através de particionamento estático, pretende-se com isto aumentar a precisão dos métodos sem causar grande impacto no desempenho. A Tabela 2.2 mostra um resumo dos trabalhos relacionados e suas características.

Esta dissertação concentra suas pesquisas em busca de imagens em ambientes grandes e heterogêneos e possui como base o descritor BIC, baseando-se no fato de que quando o BIC foi comparado com outros descritores comumente citados na literatura, ele apresentou melhores resultados em termos de eficiência e eficácia [46]. O uso do BIC permite um bom desempenho na extração das características e compacta representação do vetor de características. Entretanto, como um descritor global, o vetor de características não representa a concentração de pixels em regiões específicas da imagem, dessa forma, o descritor tende a se confundir mais facilmente quando duas imagens possuírem cores semelhantes, mas os objetos contidos em cada uma das imagens pertencerem a categorias distintas.

Baseando-se nessa observação, nossa proposta é definir um descritor baseado em partições que explora a distribuição dos pixels localmente. Para tanto, primeiramente estudamos uma técnica de segmentação simples para decompor a imagem em regiões, e em seguida classificar os pixels de cada região como borda ou interior, dessa forma, a imagem seria melhor descrita por suas informações locais. Entretanto, a segmentação possui um custo maior associado, por esse motivo, também estudamos particionar a imagem em partições fixas usando esquemas estratégicos a fim de aumentar a precisão do descritor sem afetar consideravelmente o seu desempenho.

## Capítulo 3

# Local Color Pixel Classification - um Descritor Baseado em Características Locais

A principal contribuição deste trabalho, apresentada neste capítulo, consiste em um descritor de imagens para busca, simples e eficaz quando aplicado em grandes bases de imagens heterogêneas. Essa proposta, chamada de Local Color Pixel Classification (LCPC), extrai características de cor, decompondo a imagem em um esquema de partições pré-definido e descrevendo cada partição individualmente a fim de introduzir informações espaciais sobre o conteúdo da imagem. Nessa proposta foram realizados estudos de decomposição da imagem através do uso de segmentação e vários esquemas de partições estáticas até chegarmos a melhor configuração em termos de eficácia para compor a abordagem proposta, mantendo o desempenho em níveis viáveis quando aplicados em grandes bases de imagens.

Uma contribuição adicional, apresentada também neste capítulo, inclui uma variação do descritor proposto combinando-o com características de forma a fim de retornar imagens que possuam cores diferentes das cores presentes na imagem de consulta.

Para mostrar a eficácia e a superioridade do descritor proposto, foram conduzidos experimentos a fim de compará-lo com BIC e três descritores do padrão MPEG-7 comumente descritos na literatura [13, 51, 37]. Também incluímos os descritores Color and Edge Directivity Descriptor (CEDD) e Fuzzy Color and Texture Histogram (FCTH) em nossos experimentos porque eles superaram em

eficácia os descritores do MPEG-7] [11].

Os experimentos foram realizados utilizando duas pequenas e categorizadas bases de imagens, e uma grande coleção com mais de 100 mil imagens heterogêneas coletadas da Web. Nossa suposição é que os descritores apresentam altos níveis de precisão e diferenças insignificantes entre eles em pequenas bases de imagens, enquanto que em grandes bases os níveis de precisão tendem a ser menores e as diferenças de precisão entre os descritores são consideravelmente maiores.

Na próxima seção descrevemos o desenvolvimento inicial do trabalho a fim de encontrar a melhor técnica de particionamento (segmentação ou fixa) para o descritor proposto. Nas próximas seções são descritas os esquemas de partições estáticas e o funcionamento do método, também descrevemos a inclusão e combinação de características de borda no descritor proposto.

### 3.1 Segmentação versus Particionamento Fixo

A fim de encontrar o melhor método de decomposição da imagem para a extração das características locais de cor, utilizamos primeiramente a técnica de segmentação por crescimento de região. A segmentação apresenta maior flexibilidade ao descrever o conteúdo da imagem, permitindo que atributos dos objetos sejam derivados, auxiliando no entendimento da imagem, assim, é possível obter um delineamento próximo ao que poderia ser gerado manualmente por um intérprete humano. Informações semânticas necessárias para a interpretação das imagens estão em grande parte representadas em objetos significativos e em suas relações mútuas e não em pixels individuais. Sendo assim, este trabalho concentrou, primeiramente, suas pesquisas no particionamento da imagem baseando-se em segmentação, pois mesmo que os descritores que usam segmentação sejam mais custosos, a perda no desempenho poderia ser justificada pelo alto nível de precisão obtido através do uso dessa abordagem.

O uso de segmentação exige um alto poder computacional, por esse motivo, procuramos a mais simples possível a fim de viabilizá-la em nosso método. Para tanto, utilizamos o algoritmo de segmentação por crescimento de região por ele ser um algoritmo simples sem muitas iterações e com um bom desempenho [5].

Como alternativa ao grande custo agregado às abordagens baseadas em segmentação, também exploramos as abordagens de particionamento fixo com base em esquemas estrategicamente pré-definidos. O particionamento fixo da imagem não permite separar suas regiões de forma precisa, e em muitos casos, partes do objeto principal pode se encontrar em diferentes partições. Por outro lado, descritores que usam os esquemas de particionamento estático são mais eficientes e podem

ser facilmente empregados, principalmente, em sistemas que exigem um alto desempenho. As duas formas de decomposição da imagem serão detalhadas nas seções a seguir.

### 3.1.1 Abordagem baseada em segmentação por crescimento de região

Os algoritmos de segmentação por crescimento de região partem do princípio de que partes semelhantes das imagens devem pertencer ao mesmo objeto e unimos estas em uma mesma região. Esse processo inicia considerando vários pixels de partes estratégicas da imagem como sendo uma região e é repetido iterativamente, unindo-se regiões na imagem até que se atinja um ponto de parada, que geralmente é baseado em algum critério de semelhança entre regiões. Existem muitas extensões e variações dessa técnica, entretanto buscamos pela mais simples a fim de não afetar o desempenho da abordagem proposta. Sendo assim, utilizamos a técnica de agrupar pixels que possuam propriedades similares partindo de pontos estratégicos da imagem. Os pontos foram pré-definidos a fim de agilizar o processo de segmentação, tal esquema pode ser visualizado na Figura 3.1.



Figura 3.1: Pontos sementes para segmentação.

O primeiro passo do algoritmo de segmentação era definir os pontos sementes e, a partir deles, crescer as regiões anexando a cada ponto semente outros pixels que possuam propriedades similares. Para a definição dos pontos sementes, a imagem era dividida em 9 partes iguais e os pontos considerados sementes são os pixels que se localizam no centro de cada partição, por exemplo, a Figura 3.1 ilustra esse passo do algoritmo, onde os pontos sementes estão representados no centro de cada divisão da imagem e estão marcados com a cor vermelha.

Os pixels sementes podem ser agrupados por proximidade de valor escalar ou desvio padrão. Com valor escalar, é feita uma normalização de todos os pixels e agrupam-se pixels com valores próximos, formando assim uma região de interesse. Na Figura 3.2, os pixels com valores não distantes de 5 são agrupados em três regiões.

50	51	50	102
51	49	50	102
240	240	102	102
241	240	103	103

50	51	50	102
51	49	50	102
240	240	102	102
241	240	103	103

Figura 3.2: Agrupamento de pixels em regiões.

Para o método proposto fixamos duas regiões (foreground/background), com o objetivo de separar o objeto, apresentado no centro da imagem, dos objetos que compõem o fundo da imagem, ver Figura 3.3. Caso o algoritmo encontrasse mais de duas regiões, as regiões mais próximas, de acordo com um limiar, ou seja, duas regiões com pixels possuindo valores próximos são agrupadas a fim de fixarmos a quantidade de segmentos da imagem para apenas duas regiões. Após esse processo, o resultado é uma imagem com todos os pixels classificados como pertencente a uma das duas regiões. Em imagens em que os cores do objeto central estão em contraste com as cores pertencentes ao fundo da imagem, os pixels eram classificados como pertencentes à região do centro ou à região correspondente ao fundo da imagem. Em seguida, os pixels de cada região foram classificados como borda ou interior (mesmo esquema de classificação do BIC), dessa forma, foram extraídos dois histogramas de cada região da imagem, um representando os pixels de borda, e outro de pixels de interior.

A Figura 3.3 mostra exemplos de decomposição usando segmentação por crescimentos de região. Como podemos observar, nessa técnica, nem todas as imagens podem ser bem segmentadas, pois é difícil obter um limiar de segmentação adequado para qualquer tipo de imagem. A segmentação pode ser mais robusta e flexível no momento de decompor a imagem, entretanto, a quantidade de segmentos varia de acordo com o tipo de imagem, por exemplo, imagens com grande quantidade de borda e grande quantidade de objetos tendem a possuir mais segmentos enquanto que imagens com texturas mais homogêneas possuem menos segmentos, ou seja, manter o controle da quantidade de regiões de cada região tornou-se um problema e, para solucioná-lo agrupamos os segmentos em apenas duas regiões, entretanto nos deparamos com um outro problema mostrado na Figura 3.3,

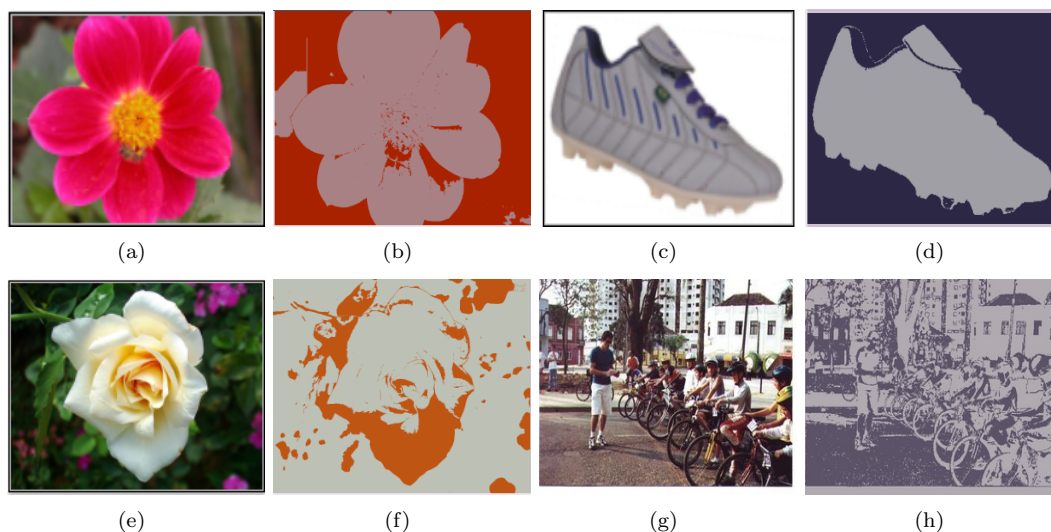


Figura 3.3: As Figuras *a*, *b*, *c* e *d* são exemplos de segmentações bem sucedidas. As Figuras *e* e *f* mostram um exemplo mal sucedido

onde os segmentos do plano de fundo e centro da imagem misturavam-se devido ao limiar que não era adequado para aquela imagem em questão como são os exemplos das imagens (f) e (h) da Figura 3.3. Procurar por métodos que pudessem resolver o problema poderiam afetar muito o desempenho dos sistemas de busca em grandes coleções heterogêneas, por esses motivos, essa abordagem foi abandonada em favor da abordagem baseada em particionamento fixo. Sendo assim, realizamos estudos comparando algumas formas de particionamento fixo para encontrar o melhor esquema em termos de precisão.

### 3.1.2 Abordagem baseada em partições estratégicas

Nos desenvolvimentos iniciais desse trabalho, procuramos a melhor forma de particionar a imagem para compor o descritor proposto. Para tanto, realizamos experimentos com esquemas de 2, 3 e 5 partições, com base em observações das disposições dos objetos presentes na imagem. Vários trabalhos na literatura [10, 9] usam diferentes formas de partições fixas, alguns particionam a imagem em grid [39, 31], por exemplo 3x3 ou 4x4, e outros usam esquemas estratégicos pré-definidos [56] para dividir a imagem. Com o objetivo de dividir a figura de forma estratégica, consideramos que o objeto de interesse encontra-se, geralmente, em seu centro, dessa forma, separamos primeiramente a imagem em plano de fundo e centro (esquema com duas partições). Para a configuração com 3 partições, o plano de fundo foi dividido em mais duas partes, e para a configuração com 5 partições, o plano de fundo foi dividido mais ainda em 4 partes. Diferentemente dos esquemas que dividem a imagem simplesmente em vários blocos, as formas de divisão estratégicas usadas neste trabalho

tentam preservar o objeto presente no centro da imagem para que ele possa receber um peso maior em relação às outras partições da imagem, pois consideramos que é essa partição que agrega muito mais informações semânticas a respeito do conteúdo da imagem. A partição de centro, nos 3 esquemas, corresponde a 50% do tamanho total da imagem, e as partições do fundo foram divididas em tamanhos iguais sem sobreposição. Ver esquema na Figura 3.4, que mostra o tamanho para cada um dos três esquemas de partições considerando uma imagem com tamanho de  $100 \times 100$  pixels.

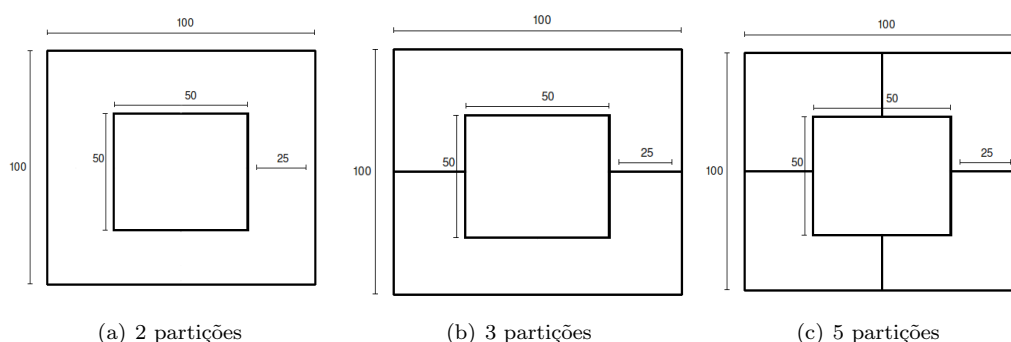


Figura 3.4: Tamanho dos três esquemas de partições.

Após o processo de divisão, os pixels de cada partição foram classificados como borda ou interior, usando o mesmo esquema de classificação do descritor BIC, gerando assim dois histogramas para cada partição, um com as informações de borda e outro de interior. Realizamos um estudo comparativo entre os três esquemas de particionamento a fim de encontrar o melhor deles em termos de eficácia para incluí-lo na abordagem proposta.

Outro parâmetro considerado nessa proposta foi o esquema de quantização de cor. A fim de escolher o melhor esquema de quantização para cada tipo de particionamento da imagem, testamos configurações de 64 e 128 cores. De fato, a configuração de 64 cores é largamente adotada na prática [54], entretanto realizamos esse experimento para avaliarmos se aumentando o número de cores de 64 para 128, o impacto na eficácia do descritor proposto seria positivo, em outras palavras, se haveria ganho nos níveis de precisão que compensasse as perdas em desempenho.

A principal desvantagem dessa técnica é a quantidade de espaço necessário para armazenar o vetor de características da imagem, pois ele tende a aumentar de acordo com o número de partições. Além do mais, isso afeta diretamente o desempenho do método, em particular, o tempo de processamento para calcular a similaridade entre a imagem de consulta e as imagens pertencentes à base.

As próximas seções descrevem em detalhes os esquemas de partições utilizados em testes a fim

de escolher o melhor dentre eles para compor o método proposto. Na próxima seção descrevemos o estudo comparativo dos três esquemas de particionamento, demonstrando detalhadamente os experimentos que nos levaram a concluir que o esquema com 5 partições foi a melhor escolha para fazer parte do descritor LCPC.

### 3.1.2.1 Esquema 1 - Partições de fundo e centro da imagem

Pensando em termos de centro/fundo da imagem, podemos dividi-la em duas partições, uma considerando pixels de centro e outra considerando pixels que compõem o plano de fundo. Isso nos permite saber quais cores predominam no centro e no fundo da imagem. Em geral, o centro define a semântica da imagem e, como tal, é mais importante ao determinar a similaridade entre imagens. Também foi considerado, nesse esquema de particionamento, que o objeto principal da imagem está na sua parte central. Isso foi feito com base em observações de várias imagens disponíveis na Web, onde a maioria dos objetos presentes na imagem localizam-se em seu centro, pois o usuário comum tende a centralizar os objetos de interesse (primeiro plano) no centro da imagem.

Esse esquema é baseado na idéia de que a maioria das imagens possuem os objetos de interesse na sua partição central. Sendo assim, dividimos a imagem em duas partições a fim de encontrar a concentração de diferentes cores pixels em cada uma das divisões, resultando assim, em uma melhor descrição dos objetos que compõem a imagem. Nesse esquema com duas divisões, a partição do centro corresponde a 50% do total de pixels que compõem a imagem, e as bordas laterais e superiores possuem tamanhos iguais, a configuração do tamanho pode ser visualizada na Figura 3.4, onde é ilustrada uma imagem com dimensões de 100x100 pixels, nesse caso o tamanho da partição central possui dimensões 50x50 e as bordas laterais, inferiores e superiores consistem tamanhos iguais correspondentes a 25 pixels.

Vamos considerar um exemplo em que o plano de fundo representa mais de 60% do conteúdo da imagem, mesmo que o plano de fundo seja grande, ele possui pouca informação sobre o significado do que está sendo representado na imagem. Por outro lado, o plano central consiste de poucos pixels, contudo, esta é a parte mais importante para definir o conteúdo da figura. Um exemplo bastante claro pode ser visualizado na Figura 3.5, onde a maioria dos pixels do plano de fundo consiste em tons da cor verde, em contraste a isso, no centro está presente uma pequena porção de pixels de diferentes cores, representando o objeto principal da figura. É exatamente essa menor porção de pixels que determina a semântica da imagem. Por essa razão, atribuímos um peso para a partição do plano central maior do que para a partição do fundo. Isso é levado em consideração



no momento do cálculo de similaridade.



Figura 3.5: Imagem dividida em duas partições.

Após o passo de divisão, é realizada a classificação dos pixels de cada partição como borda ou interior. Gerando dois histogramas para cada partição, um com pixels de borda e outro com pixels de interior. O tamanho dos histogramas depende da quantização usada. Para a quantização de 64 cores, o tamanho total do vetor de características para cada partição da imagem será de 128 elementos, sendo que 64 serão pixels de borda e os outros 64 serão de interior. Da mesma forma, para a quantização de 128, teremos dois histogramas, cada um com 128 elementos, totalizando um vetor de características de 256 elementos para cada partição. A Figura 3.6 mostra a imagem dividida em duas partições após o processamento de classificação dos pixels.

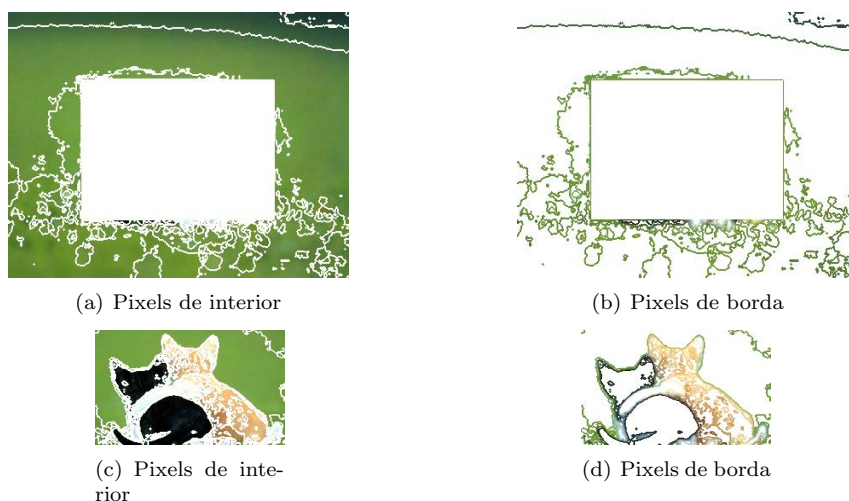


Figura 3.6: Esquema 1: duas partições com pixels classificados como borda ou interior.

Na Figura 3.6, as imagens (a) e (c) correspondem aos pixels de interior, onde os pixels que não possuem cor branca são considerados para gerar os histogramas de interior. Nas imagens (b) e (d) os pixels coloridos serão usados para gerar os histogramas de borda.

### 3.1.2.2 Esquema 2 - Uma partição de centro com duas partições de fundo da imagem

Podemos encontrar em grandes bases de imagens heterogêneas, situações típicas de imagens em que as cores do plano de fundo variam de acordo com os diferentes objetos que o compõem. Sendo assim, o plano de fundo pode ser dividido com o objetivo de representar esses diferentes conceitos.



Figure 3.7: Imagem dividida em 3 partições.

Um exemplo específico dessa situação pode ser visualizado em imagens onde, o topo do plano de fundo contém tons de cores azuis representando o céu, enquanto que a parte inferior é constituída, em sua maioria, de tons de cores verdes representando o gramado do solo. Em geral, várias imagens possuem a parte superior da imagem representando algum conceito do mundo real, diferente do que está sendo representado no centro e na parte inferior da imagem. Para esses casos, o conteúdo da imagem pode ser melhor descrito, dividindo-a em três partições: uma correspondente ao centro da imagem (semelhante ao esquema com duas partições), a segunda correspondente ao topo do plano de fundo e a terceira correspondendo a parte inferior do plano de fundo. Dessa forma, estamos separando três informações espaciais importantes sobre o conteúdo visual da imagem, o que nos permite analisá-las individualmente. Essa ideia é ilustrada na Figura 3.7

Semelhante ao esquema com duas partições, os pixels de cada parte são classificados como borda ou interior, gerando um total de 6 histogramas, dois (borda e interior) para cada uma das três partições. Nesse esquema também foi utilizado quantização de 64 e 128 cores, e também mantivemos o maior peso para a partição do centro, por esta ser a que agrega mais informação a respeito do conteúdo semântico da imagem.

A Figura 3.8 ilustra a classificação dos pixels após o passo de divisão da imagem, onde temos na imagem (a) os pixels que não possuem a cor branca representando os pixels de interior, e na imagem (b) os pixels de borda, sendo que ambas as imagens constituem o plano de fundo da imagem. A mesma situação se repete para partição que representa o centro da imagem. Desse modo, temos mais detalhes sobre as informações espaciais dos pixels que compõem o plano de fundo, espera-se

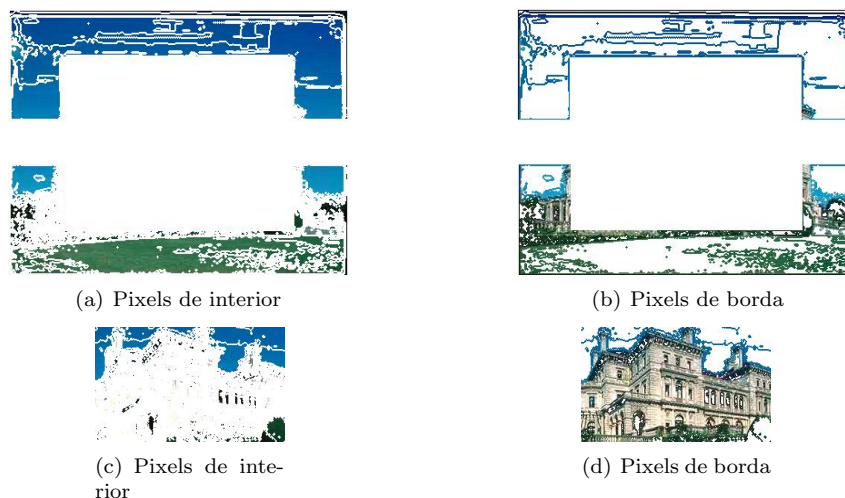


Figura 3.8: Esquema 2: três partições com pixels classificados como borda ou interior.

que o nível de precisão do método aumente consideravelmente.

### 3.1.2.3 Esquema 3 - Uma partição de centro com quatro partições de fundo da imagem

Uma outra forma de particionar a imagem é dividindo-a em cinco partições, o objetivo aqui é capturar o que está presente no centro da imagem e também representar os diversos objetos que compõem o seu plano de fundo, em muitos casos, este é bastante heterogêneo sendo constituído de diferentes cores em diferentes regiões. Por exemplo, essas diversas cores podem representar objetos em segundo plano como árvores, ruas, céu, pessoas, construções e assim por diante. Um exemplo pode ser verificado na Figura 3.9.

A estratégia aqui é dividir a imagem em cinco partições: uma partição central correspondendo ao objeto de interesse, e outras quatro partições correspondente a quatro diferentes regiões do plano de fundo denominadas de parte superior esquerda, parte superior direita, parte inferior esquerda e parte inferior direita.

Como mencionado nos parágrafos anteriores, na Figura 3.9 existe um objeto no centro que determina a semântica da figura, na parte superior esquerda contém tons de cores rosas correspondentes a flores rosas, a parte superior direita contém tons de cores azuis representando o céu, as partes inferiores esquerda e direita são constituídas de tons verdes representando árvores, uma cena típica do mundo real.

A intuição é que, com a utilização desse esquema mais detalhado, a imagem será melhor descrita, aumentando ainda mais a acurácia do método, diferente das partições nos esquemas 1 e 2. No



Figura 3.9: Imagem dividida em cinco partições.

entanto, é natural que o tempo de processamento para comparar duas imagens seja ainda maior que nas configurações anteriores devido ao número maior de partições. Por esse motivo, utilizamos apenas a quantização de 64 cores, assim, o vetor de características permaneceu com tamanho viável para realizarmos busca de imagens em grandes bases heterogêneas.

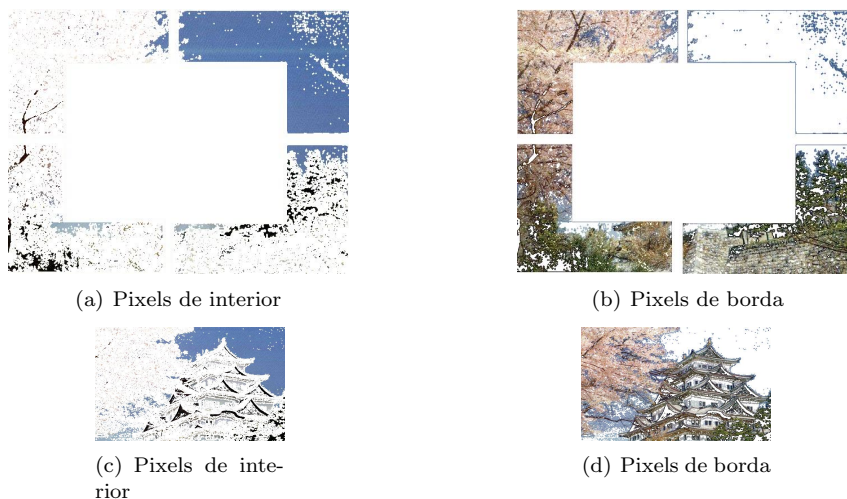


Figura 3.10: Esquema 3: cinco partições com pixels classificados como borda ou interior.

A Figura 3.10 mostra a classificação dos pixels como borda ou interior após a divisão da imagem em cinco partes, sendo que para as imagens (a) e (c) os pixels que não possuem a cor branca representam pixels de interior e, para as imagens (b) e (d) segue o mesmo esquema, os pixels que não possuem cor branca são os pixels de borda. Nessa mesma figura, verificamos que a imagem é rica em textura, gerando uma quantidade muito grande de pixels de borda. Para esses tipos de imagens o esquema 3 é mais adequado, pois o plano de fundo é bastante diversificado o que ocasiona uma grande quantidade de informações em cada partição, e essas informações são melhores analisadas através de análise local.

### 3.1.2.4 Representação das características visuais

A representação dos valores dos histogramas segue o mesmo esquema do descritor BIC. Através da função de distância  $dLog$ , é possível armazenar os resultados normalizados da função  $f(x)$  (Equação 2). A vantagem dessa forma de representar é que a comparação dos histogramas se torna muito simples, sendo assim, eles podem ser armazenados na metade do espaço necessário para a representação original.

Em descritores baseados nas abordagens globais, apenas um histograma é gerado representando toda a imagem, por exemplo se o método utiliza o esquema de quantização de 64 cores, gerando dois histogramas (borda e interior), o tamanho total do vetor de características será de apenas 128 elementos. Em abordagens baseadas em partições, o total de elementos do histograma depende da quantidade de partições e do esquema de quantização usado. Nesse trabalho, utilizamos quantização de 64 e 128 cores, no esforço de aumentarmos o nível de precisão do método proposto.

O número total de elementos do vetor de características pode ser calculado através da seguinte equação:  $2*Q*P$ , onde  $Q$  é a quantização usada,  $P$  é o número de partições, e a constante de valor 2 está associada aos dois histogramas gerados para cada partição (borda e interior). Cada valor do histograma de características ocupa 4 bits para armazenamento, assim, o tamanho, em bit, necessário para armazenar os histogramas será o número total de elementos que ele possui vezes quatro bits. A Tabela 3.1 mostra a quantidade de bits necessários para cada um dos histogramas nos três esquemas de partições (2, 3 e 5).

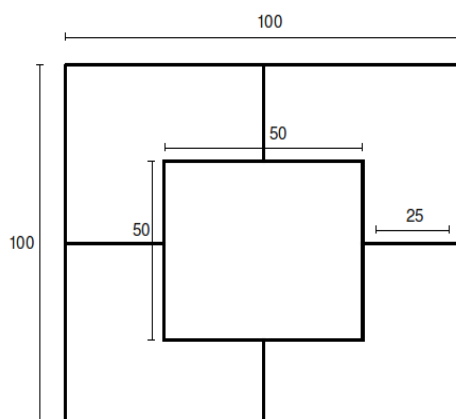
Esquema	Armazenamento em bits
2 Partições, 64 cores	1024
2 Partições, 128 cores	2048
3 Partições, 64 cores	1536
3 Partições, 128 cores	3072
5 Partições, 64 cores	2560

Tabela 3.1: Espaço necessário para o armazenamento do vetor de características.

Na Tabela 3.1, podemos verificar que o tamanho do vetor de características cresce de acordo com a quantidade de partições e a quantização usada. Por essa razão, na abordagem com 5 partições, somente o esquema de quantização de 64 cores foi usado. Se o esquema com 128 cores fosse adotado, o tamanho do vetor de características seria de 1280 elementos, o que, certamente, afetaria bastante o desempenho do método, em particular, o tempo para o processamento do cálculo de similaridade.

## 3.2 Local Color Pixel Classification - LCPC

Propomos neste trabalho, um descritor de imagem, denominado de Local Color Pixel Classification (LCPC), para realizar busca de imagens em grandes bases heterogêneas. O descritor extrai evidência de cor, baseando-se em uma análise local com partições fixas. Com o propósito de incluir informações espaciais sobre o conteúdo visual da imagem sem afetar o desempenho da abordagem proposta, decompomos a imagem estrategicamente em partições pré-definidas, descritas na Seção 3. O esquema de particionamento escolhido para fazer parte da proposta de análise local do descritor, foi o de 5 partições pois este apresentou melhor eficácia que os outros esquemas. Sendo que a partição de centro da imagem corresponde a 50% do seu tamanho e o fundo é dividido em quatro partes iguais, por exemplo, considerando a Figura 3.11 que representa uma imagem de dimensões 100x100 pixels particionada em cinco partes, o tamanho da partição do centro possui dimensões de 50x50 e a largura das bordas possui tamanho de 25 pixels para esta imagem em questão.



(a) 5 partições

Figura 3.11: Esquema de partições usado pelo LCPC.

Nossa principal contribuição é um método simples, mas muito efetivo para realizar busca de imagens em grandes coleções. Partimos do pressuposto que muitos descritores existentes na literatura funcionam muito bem em pequenas bases com imagens de categorias pré selecionadas. Mas quando estes são executados no contexto de grandes coleções de imagens heterogêneas, os seus níveis de precisão caem consideravelmente [46].

Adicionalmente, realizamos experimentos com uma extensão do descritor proposto a fim de incluir características de forma, combinando as evidências de cor e forma dos objetos presentes na imagem. O método LCPC e sua extensão serão detalhados nas seções a seguir.

### 3.2.1 LCPC com características de cor

O descritor LCPC extrai evidências de cor usando o esquema de particionamento com 5 partições a fim de incorporar informações espaciais sobre o conteúdo visual da imagem. Esse esquema é simples, mas muito efetivo em grandes coleções de imagens.

O descritor, primeiramente, particiona a imagem em 5 partes, sendo uma partição de centro e quatro partições do plano de fundo. Após esse processo de divisão, o espaço de cores é quantizado em  $4 \times 4 \times 4 = 64$  cores e os pixels de cada partição são classificados como borda ou interior, o mesmo esquema de classificação do descritor BIC. Um pixel  $p$  é considerado como borda se pelo menos um dos pixels de sua 4-vizinhança possuir uma cor diferente, um pixel será considerado de interior quando todos os 4 vizinhos possuírem a mesma cor. Dois histogramas de cor são criados para cada uma das 5 divisões da imagem: um com os pixels de borda e outro com pixels de interior, dessa forma, são obtidos 5 histogramas de 128 elementos para cada partição, totalizando assim um vetor de características com 640 elementos. A classificação dos pixels através da análise local nos dá a idéia de como eles estão distribuídos em cada uma das partições e também nos dá a noção da quantidade de bordas em cada uma. Assim, é possível ser mais preciso no momento de computar a similaridade entre duas imagens.

A similaridade entre duas imagens corresponde à similaridade entre cada uma das partições. Assim, assumimos que o centro da imagem, geralmente, define a semântica da imagem. Desse modo, é atribuído um peso para cada partição da imagem, sendo que para a partição central é atribuído um peso maior. Em nossos estudos preliminares a melhor configuração de peso foi 0.7 para a partição central e 0.3 igualmente dividido entre as outras partições restantes. Quando calculamos a similaridade entre duas imagens, foi realizada uma combinação linear de pesos. Isso nos garante que a partição do centro irá ter um maior impacto no resultado final de similaridade.

A distância de similaridade entre duas imagens que nos permite comparar dois histogramas é dada pela função  $dLog$  proposta em [54] e definida na Seção 2.

### 3.2.2 LCPC com características de cor e aresta

Descritores baseados em apenas característica de cor tendem a se confundir facilmente ao realizar uma busca de imagem em grandes bases, principalmente quando as cores do plano de fundo e centro são similares, pois não conseguem capturar os objetos presentes na figura. Outra desvantagem é que o conjunto-reposta provido por esses tipos de descritores, são constituídos de imagens que possuem sempre a mesma cor. Por exemplo, dada uma imagem de consulta que contém

uma flor vermelha, os descritores de cor retornarão como resposta apenas imagens que possuam flores vermelhas ou imagens de cores vermelhas. Note entretanto que imagens que contenham flores de outras cores provavelmente seriam relevantes para o usuário. Pensando nisso, propomos uma combinação das informações de cor e borda dos objetos a fim de diversificar as cores do conjunto-reposta e aumentar acurácia do método. Para esse propósito, adotamos o descritor de detecção de arestas baseado em [51] com algumas modificações.

Extrair todos os detalhes de forma que uma imagem possui é um processo computacionalmente complexo, e portanto, demanda muito mais tempo do que um usuário esperaria pelo resultado de uma busca online. Por essa razão, a técnica para detecção de borda será simplificada, mas de maneira que continue descrevendo significativamente as bordas dos objetos que estão sendo representados na imagem.

A detecção de aresta é essencial em uma operação de identificação de mudanças locais significativas nos tons de uma imagem. Essas mudanças podem ser descritas através de derivadas. Como a imagem depende de duas coordenadas espaciais, as bordas podem ser expressas através de derivadas parciais, o cálculo dessas derivadas pode ser aproximado através do gradiente da imagem. O gradiente é um vetor cuja direção indica os locais nos quais os níveis de cinza sofrem maior variação, a Figura 3.12 ilustra esse conceito.

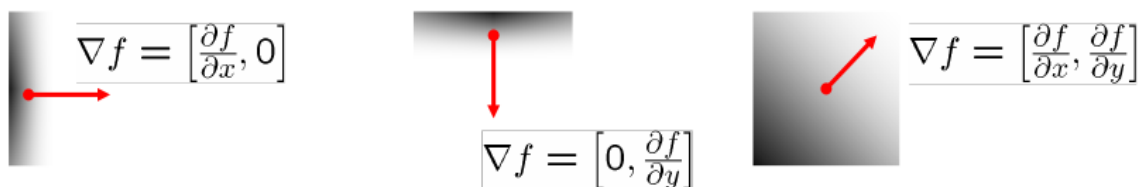


Figura 3.12: Detecção de arestas através do gradiente.

Através do cálculo do gradiente, é possível computar a direção de cada um dos pixels de aresta. Utilizar apenas o LCPC para computar as bordas e, em seguida, calcular o histograma de aresta não iria nos possibilitar calcular o gradiente para calcular as direções das arestas. Para tanto, empregamos o operador de Sobel responsável por calcular o gradiente. Operador de Sobel [27] que consiste de duas máscaras de tamanho 3x3, uma responsável por calcular o gradiente na horizontal e outra na vertical. A magnitude do gradiente  $|G|$  é então calculada pela soma das duas máscaras, resultando em  $|G| = |G_x| + |G_y|$ . As duas máscaras do operador Sobel são mostradas na Figura 3.13.

Após computar as bordas da imagem, a direção de cada um dos pixels de caresta é calculada usando o cálculo do gradiente através da equação:  $\theta(x, y) = \text{invtan}(G_y/G_x)$ . Com isso serão



-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Figura 3.13: Operador Sobel para computar os gradientes na horizontal e vertical.

geradas uma grande quantidade de direção dos pixels de aresta o que ocasionaria um histograma com muitos elementos. Para evitar esse problema, a direção das arestas é discretizada para 4 direções: vertical, horizontal, 45 e 135 graus. Esses valores são representados como um histograma de 4 elementos. Computar o histograma logo após a detecção das arestas gera a inserção de muitos ruídos, ou seja, muitos pixels que não são bordas seriam considerados como tal. Por esse motivo, antes de realizar o cálculo do histograma, a imagem é binarizada para realçar as bordas e diminuir a quantidade de ruídos. Todo o processo de detecção de arestas é realizado em cada uma das 5 partições usada pelo LCPC, gerando um histograma de 4 elementos para cada.

As informações de cor e forma são extraídas para cada uma das partições simultaneamente. Após o particionamento da imagem, analisam-se os 4 vizinhos para classificar o pixel como borda ou interior, caso o pixel seja considerado de borda, o operador de Sobel é então executado para calcular o gradiente. Dessa forma, não é necessário realizar a convolução duas vezes na imagem, uma para classificar os pixels como borda e interior, e a outra para computar o gradiente. Assim, o processo de extrair p de cor e aresta é otimizado. A Figura 3.14 mostra um exemplo de imagem após a detecção das bordas.

A função de distância usada para calcular a similaridade entre dois histogramas de aresta foi a Chi quadrado, pois, ao contrário da função dLog, não queremos atenuar as diferenças, por exemplo, entre duas imagens A e B, sendo que a imagem A possui grande quantidade de arestas e a imagem B possui uma textura mais homogênea. Nesse caso, é necessário que a grande quantidade de arestas presentes na imagem A fique evidente para diferenciá-la da imagem B. Desse modo, podemos definir a função Chi quadrado como:

$$d(h, g) = \sum_{i=0}^{i < n} (h[i] - g[i])^2 / (h[i] + g[i]) \quad (3.1)$$

Na equação 3.1,  $h$  e  $g$  são dois histogramas e  $n$  representa o tamanho de cada um deles.

Para unir as duas evidências (cor e ângulos de direção de pixels de borda), combinamos as duas



Figura 3.14: Imagens originais e após a aplicação das máscaras de Sobel.

informações usando uma média ponderada das distâncias de similaridades de cor e borda. Em outras palavras, um peso  $p$  é aplicado na distância Chi quadrado  $DQ$  que representa o distância dos histogramas de ângulos. O peso para o LCPC (evidência de cor) é igual a  $1-p$  aplicado na distância  $d\text{Log } DL$ . Desse modo, teremos a distância final calculada como:  $D = p \times DQ + (1 - p) \times DL$ . Na seção de experimentos discutiremos detalhadamente como chegamos a um valor para  $p$ .

## Capítulo 4

# Experimentos

Neste capítulo discutiremos em detalhes os experimentos executados. Primeiramente, realizamos um estudo comparativo entre os esquemas de particionamento descritos no capítulo 3, a fim de encontrar o melhor esquema em termos de eficácia para incluir no descritor proposto. Após a definição do melhor esquema de particionamento, a superioridade do descritor LCPC, em termos de eficácia, é comparada com alguns descritores comumente citados na literatura

O principal objetivo dos experimentos práticos é comparar o comportamento de todos os descritores em bases pequenas, e com categorias de imagens pré-selecionadas, bem como compará-los em grandes coleções heterogêneas sem nenhum tipo de divisão em classes. Com esse propósito, as avaliações dos descritores, em bases sem categorização, foram realizadas utilizando avaliadores reais, simulando mais próximo da realidade qual seria o comportamento dos descritores caso eles fossem inseridos em um ambiente de busca real. Para comprovar a superioridade do método proposto utilizamos algumas medidas de avaliações comumente usadas em busca de imagens, e testes estatísticos para demonstrar se os resultados obtidos são realmente significativos.

Neste capítulo também são apresentados detalhes de todo o projeto de experimento, como as medidas de avaliações, bases de imagens usadas e ambientes onde foram executados os experimentos. A seguir são apresentados todos esses detalhes utilizados para comparar descritores e os resultados obtidos.

### 4.1 Métricas de Avaliação Utilizadas

Para comparar e realizar um estudo sobre o comportamento de descritores em bases de imagens, é necessário o uso de métricas que permitam avaliá-los segundo alguns critérios pré-estabelecidos.

Adotamos nesta dissertação, métricas que possam avaliar os descritores em termos de eficácia através das medidas como, curvas de precisão/revocação, média de precisão e outras que serão descritas a seguir.

#### 4.1.1 Curvas de Precisão X Revocação

Para demonstrar a aplicabilidade deste trabalho, experimentos práticos foram conduzidos. A avaliação é medida em termos de eficácia que tem por objetivo mensurar a capacidade do descritor em recuperar imagens relevantes. A interpretação do que é uma imagem relevante pode variar dependendo da intenção do usuário que realizou a busca. Por isso, o processo de avaliação de eficácia é bastante complicado em alguns casos. Existem várias medidas para avaliar a eficácia de um método de busca, e uma forma de avaliação, comumente usada na literatura, é através das curvas de Precisão x Revocação.

Precisão é expressa como a razão entre o número de documentos relevantes recuperados e o número total de documentos recuperados, sendo necessário avaliar a relevância dos documentos recuperados. Esse é um parâmetro fundamental para a avaliação de sistemas de busca. Para avaliar a precisão é importante que o sistema informe o número total de documentos recuperados. A Revocação se refere à relação entre o número de documentos relevantes recuperados e o número total de documentos relevantes existentes na base que está sendo avaliada. Para mensurar a Revocação é necessário conhecer o número total de documentos relevantes contidos no sistema. Em outras palavras, considere  $N$  o conjunto de resposta ideal,  $|N|$  é o número de documentos deste conjunto e  $R$  o conjunto-resposta recuperado pelo método de busca. Então, podemos definir que  $Revocação = (|N \cap R|)/|N|$  e  $Precisão = (|N \cap R|)/|R|$ .

Uma curva de Precisão x Revocação indica a variação dos valores de Precisão para diferentes valores de Revocação. Em um gráfico deste tipo, quanto mais alta uma curva estiver, mais eficaz é um descritor.

#### 4.1.2 Precisão na N-ésima posição do ranking - P@N

Uma forma simples e rápida de avaliar a eficácia é utilizando base de imagens previamente classificadas, neste caso, a avaliação parte do pressuposto de que a base de imagens é previamente classificadas, ou seja, o conjunto-reposta para cada consulta é previamente conhecido. Desta forma, a avaliação de eficácia pode ser feita automaticamente, nos experimentos conduzidos em duas bases específicas a avaliação de eficácia foi feita desta maneira.

Em bases onde não existe uma classificação prévia para a obtenção automática de medidas de eficácia. Uma maneira de realizar avaliação é por meio de usuários reais, esse tipo de processo tem a vantagem de conseguir medições que refletem a opinião de usuários potenciais do sistema. Para estes casos, a eficácia é mensurada através da medida P@N que indica a fração de imagens relevantes retornadas sobre um total N de imagens retornadas.

Nos experimentos realizados nesta dissertação, utilizamos as medidas P@10, P@20 e P@30, ou seja, foram avaliadas as 10, 20 e 30 primeiras imagens do topo do ranking de imagens retornadas, respectivamente. Considerando o ambiente da Web, apenas uma pequena parcela de resposta será exibida aos usuários durante uma consulta. Por esse motivo, é importante que as respostas relevantes estejam nas primeiras posições do conjunto-resposta, daí a importância de utilizar essa medida.

### 4.1.3 Mean Average Precision - MAP

Alguns autores acreditam que a eficácia dos sistemas de busca pode ser expressa por um único valor [54]. Sendo assim, é possível comparar dois sistemas e dizer qual o melhor deles em escalas de valores relativos ou absolutos.

Uma métrica que mede o sistema em apenas um único valor é a medida de avaliação Mean Average Precision (MAP).

MAP é uma medida da eficácia média de um sistema para  $N_q$  consultas, isto é, calcula-se a média dos valores de precisão média obtidas para cada consulta  $q$ .

$$MAP = \frac{\sum_{q=1}^{N_q} (PAVG_q)}{N_q} \quad (4.1)$$

Na equação 4.1,  $PAVG_q$  é a precisão média para uma consulta  $q$  e  $N_q$  é o número de consultas utilizadas.

### 4.1.4 Método de Pooling

Quando a base de imagens utilizada para avaliação é grande e não possui classificação prévia, não é possível a obtenção automática das medidas quantitativas de eficácia, como MAP e Precisão x Revocação. Uma forma de avaliação é através de usuários reais que avaliam somente as  $N$  respostas do topo (geralmente 30 respostas), pois é inviável para usuário classificar a base inteira. Nesse caso, podemos calcular medidas mais simples como P@N considerando apenas as  $N$  imagens do

topo do ranking das imagens retornadas. Entretanto, é possível calcular as medidas que necessitam do conhecimento do conjunto-resposta para uma dada consulta por meio do Método de Pooling. Esse método é bastante empregado na coleção baseada na Web da TREC [21] [20].

O Método de Pooling determina o conjunto de relevantes para um conjunto  $Q$  de imagens de consulta da seguinte forma: para cada consulta, executamos todos os descritores que estão sendo comparados. As  $N$  imagens do topo recuperadas por cada método são colocadas em um conjunto único, dessa forma, não é possível saber quais descritores recuperaram quais imagens. Em seguida, as imagens do conjunto são avaliadas por usuários reais como relevante ou irrelevante em relação a imagem de consulta. Como resultado, teremos um conjunto de imagens classificadas como relevantes que irão compor o conjunto de imagens relevantes para referência. Desta maneira, podemos comparar o conjunto de referência com o conjunto-resposta de cada descritor para extrair as medidas de eficácia em que é necessário saber previamente o conjunto de imagens relevantes para uma determinada consulta.

Através desse método, é possível calcular medidas como curvas de Precisão x Revocação, e MAP em bases que não são classificadas previamente. É importante ressaltar que este método não garante que todas as imagens relevantes, para uma determinada consulta, serão encontradas na base, por esta razão, dizemos que o método calcula apenas a revocação relativa. Este processo evita o imenso trabalho de se classificar uma grande coleção, como por exemplo, a base usada neste trabalho que possui mais de 100 mil imagens.

#### 4.1.5 Teste de validação Wilcoxon Matched-Pairs Signed-Ranks

Com o propósito de garantir que os resultados do método proposto são realmente significativamente superiores em relação aos outros descritores usados como baseline, utilizamos o teste estatístico Wilcoxon Matched-Pairs Signed-Ranks que é comumente usado para validar esse tipo de resultado [61]. Este é um teste bastante similar ao T-test [6]. Com a diferença de que o T-test é usado sempre que houver certeza de que a distribuição de probabilidades dos valores dados é normal.

O teste Wilcoxon é aplicado para determinar se a diferença entre as médias, de dois grupos de dados, é realmente significativa ou se é devido simplesmente ao acaso. Para aplicar o teste, é necessário atender alguns requisitos: (1) o teste deve ser utilizado quando os dados tiverem algo em comum; (2) a distribuição dos dados não deve ser uma distribuição normal; e (3) os avaliadores devem ser selecionados aleatoriamente. Os experimentos realizados nesta dissertação atendem a

todos esses requisitos.

## 4.2 Base de Imagens

Para realização dos experimentos práticos dos descritores, foram usadas três diferentes bases de imagens. Entre elas estão duas bases pequenas e classificadas, e uma base com mais de 100 mil imagens não classificadas. As três bases são heterogêneas e apropriadas para realização de busca de imagens utilizando evidências de cor e forma. Todas as bases são descritas nas próximas seções.

A intenção de usar duas bases pequenas é para confirmar a nossa hipótese em que descritores comportam-se bem nesse tipo de ambiente, alcançando níveis de precisão, na maioria das vezes, superior a 80%. Entretanto, em bases grandes e heterogêneas o comportamento é outro e as diferenças em termos de precisão são acentuadas, sendo que a maioria dos descritores tendem a alcançar um nível de precisão inferior a 40%.

### 4.2.1 Base Wang

A base Wang [12] consiste de um conjunto de 1000 imagens pertencentes à base Corel Stock Photo, todas as figuras foram selecionadas manualmente para formar um conjunto com 10 categorias, sendo que cada categoria contém 100 imagens. A base Wang pode ser considerada como uma base em potencial para busca de imagens, onde o usuário possui uma imagem de consulta de uma determinada categoria, tendo como objetivo buscar imagens semelhantes que pertençam a mesma categoria. As 10 classes são usadas para estimar a relevância do conjunto-reposta retornado por uma consulta, por exemplo, dada uma imagem de consulta pertencente a base e a uma das 10 categorias, é assumido que as 99 imagens restantes são relevantes para esta consulta e todas as outras são irrelevantes. A Figura 4.1 mostra 10 de imagens pertencentes à base Wang que foram utilizadas como consulta. Todas elas foram selecionadas de forma aleatória, pois consideramos que qualquer imagem pertencente a uma categoria poderia ser usada como consulta.

Essa base foi usada em outros trabalhos [11] [37], e por esse motivo ela é usada em nossos experimentos, permitindo uma comparação direta com outros métodos de busca de imagens.

### 4.2.2 Base Common Color Dataset - CCD

A base Common Color Dataset (CCD) consiste em um total de 5466 imagens. Para esta base, existe um conjunto de 50 imagens de consulta, onde para cada consulta é determinado um conjunto

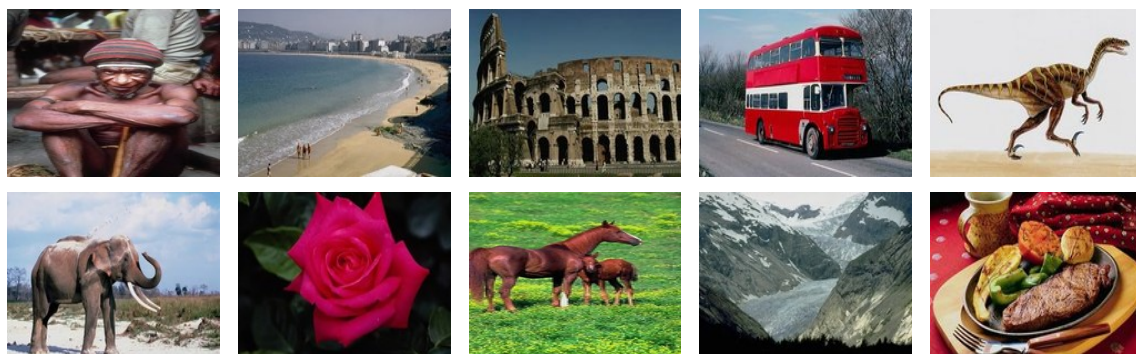


Figura 4.1: Imagens usadas como consulta para experimentos na base Wang.

de imagens relevantes, que será usado para o cálculo de medidas de eficácia. A base CCD é o banco de imagens onde foram realizados os testes para os descritores do MPEG-7, ela contém figuras que se originaram de programas de televisão como noticiários e esportes. Esta base é designada para avaliar descritores de cor. Dois exemplos de consulta são ilustrados na Figura 4.2, onde a primeira imagem, de cada exemplo, foi usada como consulta e as outras são exemplos de respostas relevantes.

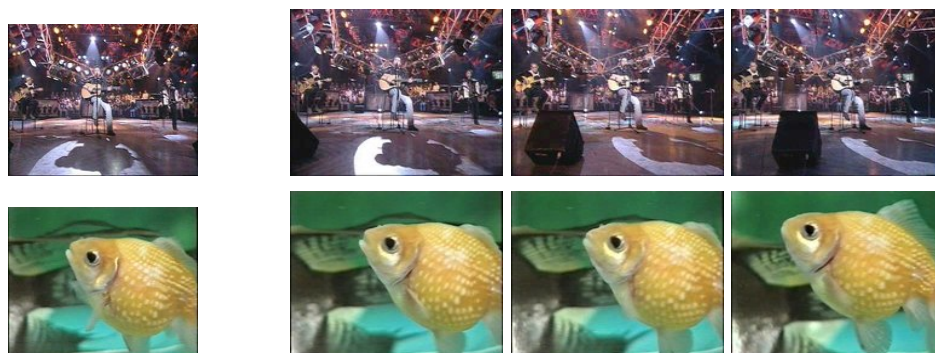


Figura 4.2: Dois exemplos de consulta na base CCD e 3 respostas para cada consulta.

Nesta base os descritores avaliados tendem a apresentar altos valores de precisão, pois ela é bastante comportada e as imagens relevantes para uma determinada consulta são bastante semelhantes à imagem de consulta, elas diferem-se em pequenas variações de ângulos das imagens ou diferentes imagens tiradas da mesma cena, sendo que o brilho e contraste das imagens continuam os mesmos. Isso pode ser observado no exemplo da Figura 4.2. Apesar desse comportamento, essa base foi incluída porque os descritores do MPEG-7 usados como baselines nesta dissertação, usam essa base em experimentos realizados em [37, 11]. Para as avaliações dos descritores, usamos um conjunto de 50 imagens de consultas previamente definidas de acordo com o *ground truth* pertencente a essa base e disponível no site dos descritores do MPEG-7 <sup>1</sup>.

<sup>1</sup><http://www.ee.cityu.edu.hk/~mirror/> acessada em 05/2011



### 4.2.3 Base Yahoo

A base de imagens Yahoo foi coletada pelo grupo de Banco de Dados e Recuperação de Informação (BDRI) da Universidade Federal do Amazonas (UFAM), de forma recursiva a partir de endereços de sites contidos no diretório Yahoo. Ela é composta por 103.378 imagens pertencentes a domínios diversos, onde foram excluídos ícones, banners e imagens com tamanho inferior a 320x200. As páginas HTML também foram armazenadas para manter a referência entre cada documento HTML e as imagens neles contidas, permitindo também trabalhos de busca e classificação de imagens baseados em características visuais. Dessa maneira, a base possui um conteúdo bastante heterogêneo e representativo do conteúdo da Web e não possui nenhum tipo de classificação. Essa base foi usada em outros trabalhos como [46].

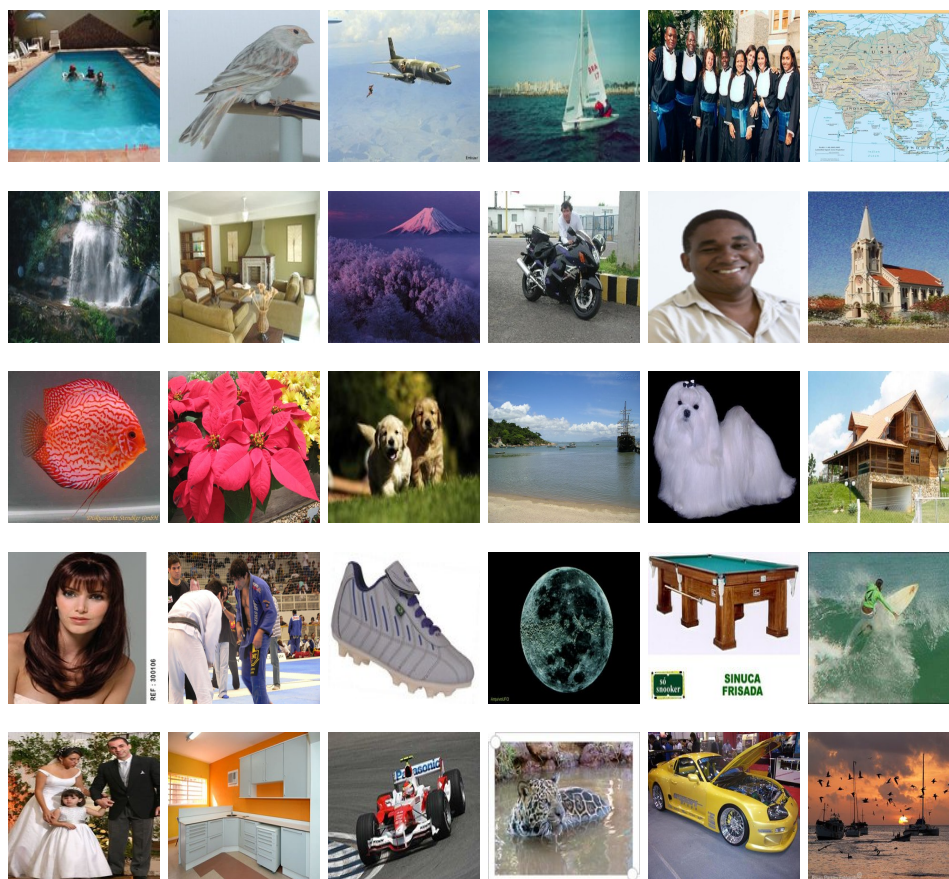


Figura 4.3: 30 imagens de consulta usadas na base Yahoo.

Esta base foi usada para avaliar o conjunto de descritores que alcançaram altos níveis de precisão nas bases menores (CCD e Base Wang), através de avaliações de usuários reais. Para esta base, foram utilizadas 30 imagens de consulta que se enquadram em diferentes categorias como esporte,

lazer, flores, animais, entre outras, para cada imagem de consulta foi retornado um conjunto-resposta com 30 imagens. A Figura 4.3 apresenta as 30 imagens usadas como consulta.

Para facilitar o trabalho de avaliação, foi usada a ferramenta [45] que integra todas as etapas necessárias para a realização de experimentos práticos de testes comparativos entre descritores, e também usamos a ferramenta [36], pois ela contém as implementações dos descritores usados como baseline.

## 4.3 Resultados

Inicialmente, apresentamos os resultados dos testes com os diferentes esquemas de particionamento a fim de definir o melhor deles para compor o descritor LCPC. Em seguida apresentamos a comparação dos resultados de eficácia entre o LCPC e os baselines em cada uma das bases usadas, bem como, o teste de validação dos resultados.

### 4.3.1 Estudo comparativo entre os três esquemas de particionamento

Neste experimento foram realizados testes comparativos usando os três esquemas de particionamento (2, 3 e 5 partições) descritos na Seção 3, variando o esquema de quantização entre 64 e 128 cores, a fim de encontrar o melhor esquema de quantização e de particionamento em termos de eficácia e eficiência. Para tanto, realizamos os experimentos nas base do Yahoo usando 50 imagens de consulta e um total de 25 avaliadores reais, sendo 5 avaliadores para cada descritor avaliado.

A quantização foi variada apenas para os particionamentos com 2 e 3 partições, para o esquema com 5 partições foi mantida a quantização de 64 cores, dessa forma, foram avaliados um total de 5 diferentes configurações para compor o descritor LCPC: (1) LCPC com 2 partições e 64 cores (LCPC2Partições64); (2) LCPC com 2 partições e 128 cores (LCPC2Partições128); (3) LCPC com 3 partições e 64 cores (LCPC3Partições64); (4) LCPC com 3 partições e 128 cores (LCPC3Partições128); (5) LCPC com 5 partições e 64 cores (LCPC5Partições64). Todas estas 5 configurações foram comparadas com o descritor BIC original, a fim de nos certificar de que todas as configurações seriam superiores a ele em termos de eficácia. Escolhemos o BIC porque dentre vários descritores avaliados na base do Yahoo, ele mostrou-se superior aos outros [46].

As métricas de avaliações usadas foram P@10, P@20, P@30, ou seja, os usuários avaliaram quais imagens eram relevantes ou irrelevantes dentre as 10, 20 e 30 primeiras imagens do topo do ranking de imagens retornadas como resposta. Utilizamos essas métricas, por elas serem mais

simples e práticas para realizar avaliações em grandes coleções de imagens heterogêneas, sem que seja necessário a classificação prévia das imagens presentes na base.

#### 4.3.1.1 Eficácia

De acordo com os resultados obtidos, todas as cinco configurações avaliadas alcançaram níveis de precisão maiores que o BIC. Os resultados das 5 avaliações nas três medidas de precisão (P@10, P@20 e P@30) para os 6 descritores comparados podem ser verificados nas Tabelas 4.1, 4.2 e 4.3, bem como a média geral das precisões e o cálculo de variância entre cada uma das 5 avaliações.

	Descritor	Avaliação1	Avaliação2	Avaliação3	Avaliação4	Avaliação5	Média	Variância
1	BIC	0,37	0,42	0,29	0,34	0,46	<b>0,38</b>	0,00443
2	LCPC2Partições64	0,50	0,30	0,40	0,51	0,47	<b>0,44</b>	0,00763
3	LCPC2Partições128	0,52	0,49	0,51	0,5	0,46	<b>0,50</b>	0,00053
4	LCPC3Partições64	0,52	0,43	0,58	0,51	0,53	<b>0,51</b>	0,00293
5	LCPC3Partições128	0,53	0,46	0,47	0,52	0,56	<b>0,51</b>	0,00177
6	LCPC5Partições64	0,57	0,52	0,56	0,55	0,57	<b>0,55</b>	0,00043

Tabela 4.1:  $P@10$  para 5 avaliações de 50 consultas.

	Descritor	Avaliação1	Avaliação2	Avaliação3	Avaliação4	Avaliação5	Média	Variância
1	BIC	0,26	0,30	0,19	0,23	0,37	<b>0,27</b>	0,00475
2	LCPC2Partições64	0,39	0,19	0,28	0,41	0,35	<b>0,33</b>	0,00808
3	LCPC2Partições128	0,39	0,36	0,39	0,37	0,41	<b>0,39</b>	0,00038
4	LCPC3Partições64	0,41	0,32	0,44	0,37	0,41	<b>0,39</b>	0,00215
5	LCPC3Partições128	0,41	0,31	0,34	0,40	0,42	<b>0,38</b>	0,00233
6	LCPC5Partições64	0,45	0,39	0,43	0,44	0,46	<b>0,43</b>	0,00073

Tabela 4.2:  $P@20$  para 5 avaliações de 50 consultas.

	Descritor	Avaliação1	Avaliação2	Avaliação3	Avaliação4	Avaliação5	Média	Variância
1	BIC	0,21	0,25	0,14	0,18	0,31	<b>0,22</b>	0,00427
2	LCPC2Partições64	0,33	0,14	0,23	0,35	0,28	<b>0,27</b>	0,00713
3	LCPC2Partições128	0,32	0,30	0,33	0,32	0,27	<b>0,31</b>	0,00057
4	LCPC3Partições64	0,33	0,25	0,37	0,30	0,33	<b>0,32</b>	0,00198
5	LCPC3Partições128	0,33	0,25	0,28	0,33	0,34	<b>0,31</b>	0,00153
6	LCPC5Partições64	0,38	0,32	0,36	0,37	0,39	<b>0,36</b>	0,00073

Tabela 4.3:  $P@30$  para 5 avaliações de 50 consultas.

Como podemos perceber nas Tabelas 4.1, 4.2 e 4.3, as médias das avaliações dos 5 avaliadores possuem pouca variância, provando que, apesar do pequeno número de avaliadores por descritor, os resultados podem ser considerados confiáveis. Em geral, a precisão P@10 (Tabela 4.1) para cada descritor, com exceção do BIC, alcançou níveis acima de 40%, mais especificamente, os métodos com 3 partições ultrapassaram os 50% de precisão, e o esquema com 5 partições alcançou 55,4% de precisão, esse último valor pode ser considerado um alto nível de precisão, pois os experimentos foram realizados em uma grande coleção de imagens. Nesse tipo de ambiente os descritores baseados em cor tendem a alcançar baixos níveis de precisão devido à heterogeneidade das imagens [46]. Em

P@20, a precisão obtida pelas configurações avaliadas ultrapassou os 30%, e o LCPC5Partições64 obteve um nível 43,3%, todos os valores permaneceram maiores que a precisão de 27,7% obtida pelo BIC. O mesmo comportamento foi observado em P@30, onde os esquemas com partições também superaram os valores de precisão obtidos pelo BIC.

Em relação aos três esquemas de partições avaliados, a configuração com 5 partições foi a que apresentou melhores resultados em termos de eficácia, pois como podemos observar nas Tabelas 4.1, 4.2 e 4.3, o esquema com 5 partições alcançou 55%, 43% e 36% em P@10, P@20 e P@30 respectivamente, contra os valores do segundo melhor esquema (LCPC3Partições64) em termos de eficácia que obteve valores de precisão de 51%, 39% e 32% em P@10, P@20 e P@30 respectivamente. Outro ponto importante a ser observado é que a variação na quantização não causou grande impacto nos valores de precisão obtidos pelos esquemas com dois e três partições. Por exemplo, os valores de precisão obtidos pelos esquemas LCPC3Partições64 e LCPC3Partições128 foram iguais em P@10 e P@20, havendo uma pequena diferença em P@30 onde o LCPC3Partições64 obteve 32% de precisão e o LCPC3Partições128 alcançou 31%.

#### 4.3.1.2 Eficiência

A eficiência dos descritores foi avaliada através de duas medidas: (1) o tempo de processamento necessário para extrair as características de uma imagem e (2) o tempo necessário para calcular a similaridade entre duas imagens (imagem de consulta e imagem da base). Essas medidas foram calculadas para as 50 imagens de consulta para cada descritor, incluindo o descritor BIC original. As Tabelas 4.4 e 4.5 mostram os resultados.

	Descritor	Média	Desvio padrão
1	BIC	28,1	43,1
2	LCPC2Partições64	21,4	33,9
3	LCPC2Partições128	21,6	34,2
4	LCPC3Partições64	21,2	33,4
5	LCPC3Partições128	21,7	34,3
6	LCPC5Partições64	21,7	34,3

Tabela 4.4: Média do tempo de extração de característica em *ms*.

Na Tabela 4.4 observamos uma pequena variação entre o tempo de extração de característica dos descritores. Isso pode ser explicado pelo fato de que apesar dos esquemas de particionamento dividirem a imagem, o número de pixels a ser analisado pelo algoritmo de extração de característica permanece o mesmo, ou seja, isso independe do número de partições da imagem. Além disso, o particionamento estático é bastante rápido, o que não prejudica o desempenho do algoritmo no

	Descritor	Média	Desvio padrão
1	BIC	45,6	4,7
2	LCPC2Part64	63,6	9,0
3	LCPC2Partições128	105,0	9,8
4	LCPC3Partições64	85,6	6,3
5	LCPC3Partições128	151,0	1,4
6	LCPC5Partições64	126,0	8,8

Tabela 4.5: Média do tempo de cálculo de similaridade em  $\mu s$ .

momento da extração de características.

A Tabela 4.5 apresenta o tempo do cálculo de similaridade entre duas imagens. Esse tempo pode aumentar de acordo com o tamanho do vetor de características, que por sua vez, aumenta de acordo com o número de partições e a quantização usada. Como esperado, o descritor LCPC3Partições128 que possui o maior vetor de característica com  $2 \times 128 \times 3(\text{particoes}) = 768$  elementos, é o que apresenta maior tempo de cálculo de similaridade com  $151\mu s$ . Entretanto, é importante observar que apesar do descritor LCPC5Partições64 possuir um tempo de similaridade de  $126\mu s$  e o BIC apresentar  $45,6\mu s$ , o LCPC5Partições64 apresenta maior valor de precisão. Além disso, a perda no desempenho pode ser superada através do uso de técnicas de paralelismo no cálculo de similaridade ou indexação dos vetores de características. Sendo dessa forma, possível empregá-lo em sistemas de busca em grandes coleções de imagens.

Apesar desse pequeno estudo relacionado a eficiência dos descritores, não é a intenção deste trabalho focar em questões de eficiência, mas sim na eficácia dos descritores. Esta breve demonstração de eficiência dos descritores foi apenas para mostrar que os tempos de extração e de cálculo de similaridade podem ser viáveis para busca de imagens em grandes coleções.

#### 4.3.1.3 Validação dos Resultados

Para comparar as diferenças entre as médias de avaliações de cada imagem de consulta e verificar se os resultados são realmente significativos, aplicamos o teste estatístico Wilcoxon Matched-Pairs Signed-Ranks. Para aplicá-lo, a média das 50 consultas foram comparadas da seguinte forma: a média de precisão em P@10 da primeira consulta do BIC foi comparada com a média da primeira consulta de cada um dos outros descritores, e o mesmo foi feito para a segunda, terceira, quarta consulta e assim por diante. O mesmo processo foi repetido para as medidas P@20 e P@30. As Tabelas 4.6, 4.7 e 4.8 apresentam os níveis de significância dos resultados de todos os descritores.

A Tabela 4.6 compara, através da linha 1 e coluna 2 ou linha 2 e coluna 1, os métodos LCPC2Partições64 e BIC, como a média do LCPC2Partições64 foi maior que a do BIC em P@10,

Descritor		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
BIC	<b>1</b>	0	100	100	100	100	100
LCPC2Partições64	<b>2</b>	100	0	98	100	87	100
LCPC2Partições128	<b>3</b>	100	98	0	73	27	99
LCPC3Partições64	<b>4</b>	100	100	73	0	57	99
LCPC3Partições128	<b>5</b>	100	87	27	57	0	95
LCPC5Partições64	<b>6</b>	100	100	99	99	95	0

Tabela 4.6: Nível de significância em  $P@10$ .

Descritor		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
BIC	<b>1</b>	0	100	100	100	100	100
LCPC2Partições64	<b>2</b>	100	0	90	100	84	100
LCPC2Partições128	<b>3</b>	100	90	0	66	32	99
LCPC3Partições64	<b>4</b>	100	100	66	0	65	100
LCPC3Partições128	<b>5</b>	100	84	32	65	0	98
LCPC5Partições64	<b>6</b>	100	100	99	100	98	0

Tabela 4.7: Nível de significância em  $P@20$ .

Descritor		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
BIC	<b>1</b>	0	100	100	100	100	100
LCPC2Partições64	<b>2</b>	100	0	93	99	62	100
LCPC2Partições128	<b>3</b>	100	93	0	60	73	99
LCPC3Partições64	<b>4</b>	100	99	60	0	77	100
LCPC3Partições128	<b>5</b>	100	62	73	77	0	99
LCPC5Partições64	<b>6</b>	100	100	99	100	99	0

Tabela 4.8: Nível de significância em  $P@30$ .

então ele é superior com nível de confiança de 100%. Comparando o BIC (linha 1) com todos os outros descritores em todas essas 3 tabelas, percebemos que todos os métodos são superiores ao BIC nas três medidas ( $P@10$ ,  $P@20$  e  $P@30$ ) com nível de confiança de 100%, pois os valores de médias de todas as abordagens são maiores que as médias do BIC. Entre os esquemas de particionamento avaliados, verificamos que o melhor deles foi o LCPC5Partições64, pois ao compará-lo com todos os outros nos três níveis de precisão, ele mostrou-se significativamente superior e melhor que todas as outras configurações de particionamento com nível de confiança acima de 95% (Tabela 4.6: linha 6 e coluna 5). Para comparar as configurações LCPC3Partições64 e LCPC3Partições128 em  $P@10$ , primeiramente analisamos a linha 4 e coluna 5 da Tabela 4.6, então podemos observar que um método não é significativamente superior ao outro, pois o nível de confiança foi de apenas 57%.

Um outro ponto importante a ser observado, é que a variação de quantização não causou grande impacto no nível de precisão dos esquemas avaliados, por esse motivo, para compor o método LCPC a quantização de 64 cores foi mantida e, entre os esquemas de particionamento, escolhemos a configuração com 5 partição, pois essa foi significativamente superior a todas as outras.

### 4.3.2 Comparação entre o LCPC e outros descritores

Esta seção mostra os resultados comparativos entre o LCPC, a extensão do LCPC com características de cor e forma (LCPC+Aresta) e os descritores do MPEG-7 (SCD, EHD e CLD), FCTH e CEDD nas bases Wang e CCD. Na base Yahoo, a acurácia do LCPC e LCPC+Aresta foi comparada apenas com os descritores FCTH e CEDD, pois eles superaram os descritores do MPEG-7 em termos de eficácia nas bases Wang e CCD. Além disso, na base Wang, foi realizado um teste empírico de peso para a combinação de cor e forma, a fim de descobrir qual a melhor valor de peso para as duas características, ou seja, qual a característica que tem maior impacto na acurácia do descritor LCPC+Aresta. A seguir são descritos os detalhes dos resultados dos experimento e dos testes de validação.

#### 4.3.2.1 Teste de atribuição de peso para combinação de características de cor e forma no descritor LCPC

A base Wang foi usada para o teste de atribuição de peso da combinação de características de cor e forma para encontrar a melhor configuração em termos de precisão. Para tanto, usamos um total de 10 consultas, sendo que para cada consulta existem na base 99 relevantes que pertencem a mesma categoria. Dado que a base é classificada foi possível a extração automática de métricas de avaliações sem a necessidade de usuários reais. Nesse experimento usamos as curvas de Precisão x Revocação para encontrar a melhor configuração de peso. Dessa forma, é possível saber qual característica tem maior impacto no cálculo dos níveis de precisão do descritor LCPC. Após a definição da melhor atribuição de peso, esta configuração foi usada para testes de comparação com os outros descritores avaliados. A Figura 4.4 mostra o resultado do teste de peso para a combinação das duas evidências.

Como podemos observar na Figura 4.4, a melhor configuração de peso para a combinação de cor e forma foi de 10% para a evidência de forma e 90% para a cor, de acordo com a figura, a precisão cai a medida que aumentamos o peso para a evidência de forma, estes resultados podem ser modificados explorando melhor o descritor de aresta, pois existem muitos parâmetros que podem ser estudados para fazer com que o nível de precisão do descritor aumente. Apesar do peso para a forma ser baixo, as diferenças dos resultados no conjunto-resposta para a maioria das consultas são interessantes, pois, em geral, os descritores de cor retornam no conjunto-resposta apenas imagens que possuem a mesma cor da imagem usada como consulta. Por exemplo, se a consulta possui um carro amarelo, as imagens retornadas serão apenas de carros amarelos ou figuras que possuam a

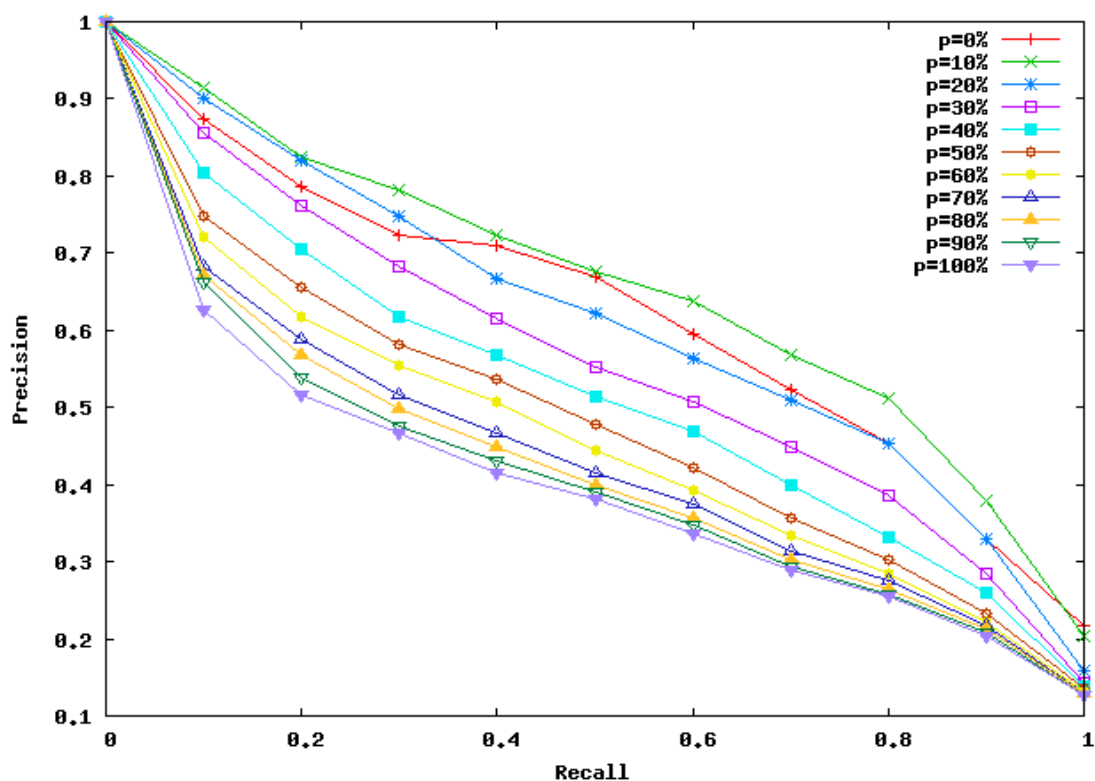


Figura 4.4: Curvas de Precision x Recall para o teste de peso na base Wang.

cor amarela, diferentemente dos descritores que usam evidência de forma que são capazes, neste caso, de retornar algumas imagens de carros de outras cores.

Após a definição da configuração de peso mais adequada para a combinação do descritor LCPC+Aresta, este foi comparado, em termos de eficácia, com o LCPC e todos os outros descritores usados como baseline.

#### 4.3.2.2 Resultados na base Wang

Para a comparação de todos os descritores na base Wang, usamos as mesmas 10 consultas usadas no teste de atribuição de peso, também considerando que para cada consulta existem 99 imagens relevantes. Essa classificação nos possibilitou calcular automaticamente as métricas de avaliação MAP e curvas de Precisão x Revocação. Para esse experimento também foi realizado o teste de validação

##### Eficácia

A maioria dos descritores avaliados na base Wang tiveram um comportamento semelhante, ou seja, os níveis de precisão ficaram bem próximos, isto pode ser melhor ilustrado na Figura 4.5 que



mostra as curvas de Precision x Recall de todos os descritores.

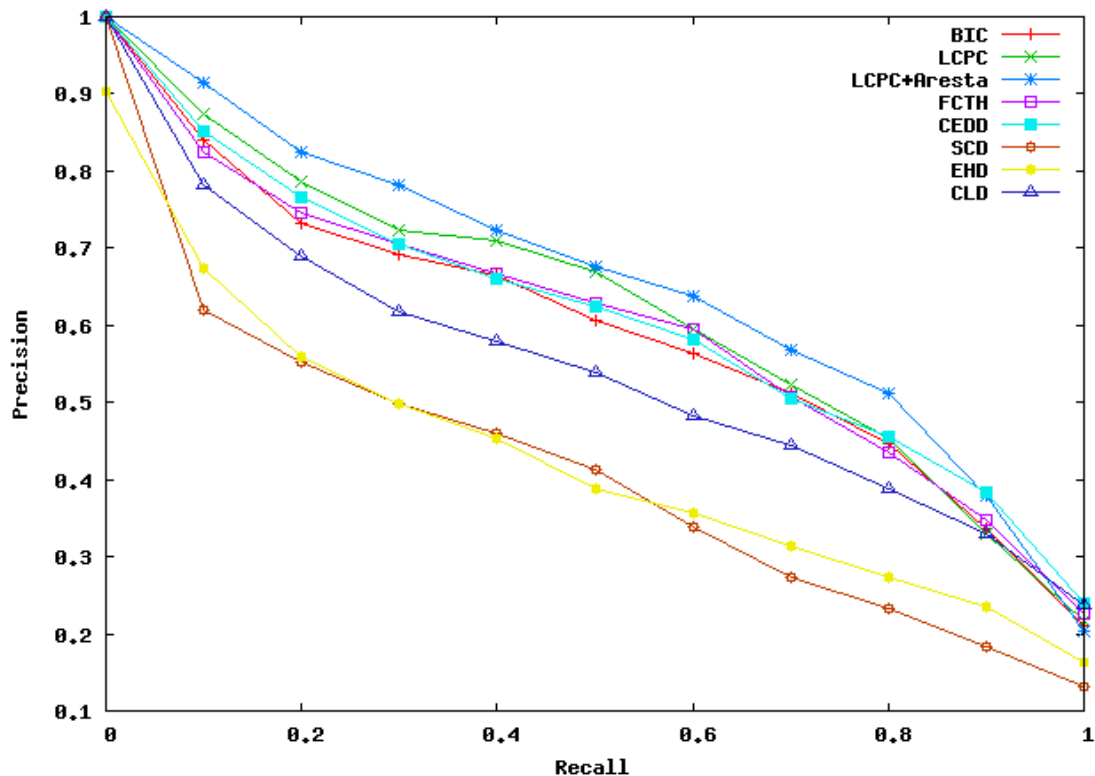


Figura 4.5: Curvas de Precision x Recall para os descritores na base Wang.

A Figura 4.5 mostra que, apesar de os descritores LCPC e LCPC+Aresta alcançarem os maiores níveis de precisão, ainda sim, suas curvas estão bem próximas dos descritores FCTH e CEDD. Por esse motivo, é importante analisarmos o teste de validação para descobrir se essas pequenas diferenças são realmente significativas. Em relação aos descritores do MPEG-7, os métodos LCPC e LCPC+Aresta apresentaram curvas bem maiores.

Os valores de MAP exibidos na Tabela 4.9 para todos os descritores, também apresentam o mesmo comportamento, os descritores do MPEG-7 tiveram MAP bem inferior aos descritores LCPC, LCPC+Aresta, FCTH e CEDD, enquanto que estes últimos permaneceram com valores bem próximos

Descritor	LCPC+Aresta	LCPC	BIC	CEDD	FCTH	CLD	EHD	SCD
MAP	0,67	0,62	0,59	0,59	0,58	0,51	0,39	0,38

Tabela 4.9: Valores de MAP para os descritores na base Wang.

### Validação

Para termos certeza de os valores de precisão entre os descritores são realmente significativos, aplicamos o teste de validação comparando a média de MAP da consulta 1 de um descritor à média da consulta 1 de outro descritor, o mesmo processo foi executado para a consulta 2 e assim por diante até a décima consulta. Os níveis de confiança entre os métodos podem ser visualizados na Tabela 4.10.

Descritor		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>
LCPC	<b>1</b>	0	75	81	30	44	92	99	99
LCPC+Aresta	<b>2</b>	75	0	95	81	72	97	100	99
BIC	<b>3</b>	81	95	0	0	0	84	99	99
CEDD	<b>4</b>	30	81	0	0	15	68	100	97
FCTH	<b>5</b>	44	72	0	15	0	81	99	97
CLD	<b>6</b>	92	97	84	68	81	0	97	77
EHD	<b>7</b>	99	100	99	100	99	97	0	0
SCD	<b>8</b>	99	99	99	97	97	77	0	0

Tabela 4.10: Validação dos descritores na base Wang.

De acordo com a Tabela 4.10, o descritor LCPC+Aresta é superior ao BIC (linha 2 e coluna 3) com nível de significância de 95% e, também é significativamente superior aos descritores do MPEG-7. Entretanto ele não se apresenta de forma superior em relação aos descritores FCTH e CEDD, sendo o nível de confiança de 72% para o FCTH e 81% para o CEDD. O descritor LCPC não se mostrou superior ao FCTH e nem ao CEDD, mas, de acordo com a tabela, ele é superior aos descritores do MPEG-7. Esse comportamento é esperado, uma vez que a base é pequena, então os níveis de precisão dos descritores tendem a ser altos e bem próximos nesse tipo de ambiente.

#### 4.3.2.3 Resultados na base CCD

Para os experimentos na base CCD, foram usadas um total de 50 consultas. Devido ao fato dessa base ser classificada, é possível a extração automática das curvas de Precisão x Revocação e MAP. Nessa base, todos os descritores foram avaliados. Os detalhes sobre os resultados de eficácia e validação são descritos a seguir.

##### Eficácia

A base CCD é pequena contendo exatamente 5.466 imagens. Dessa forma, os valores de precisão, para a maioria das abordagens estudadas, foram bem altos (a cima de 70%) como podemos verificar na Figura 4.6.

De forma semelhante ao comportamento apresentado na base Wang, a Figura 4.6 mostra que os descritores do MPEG-7 (SCD, EHD e CLD) permaneceram com valores de precisão menores que os demais métodos. Quanto às curvas dos descritores LCPC, LCPC+Aresta e BIC, elas apresentam

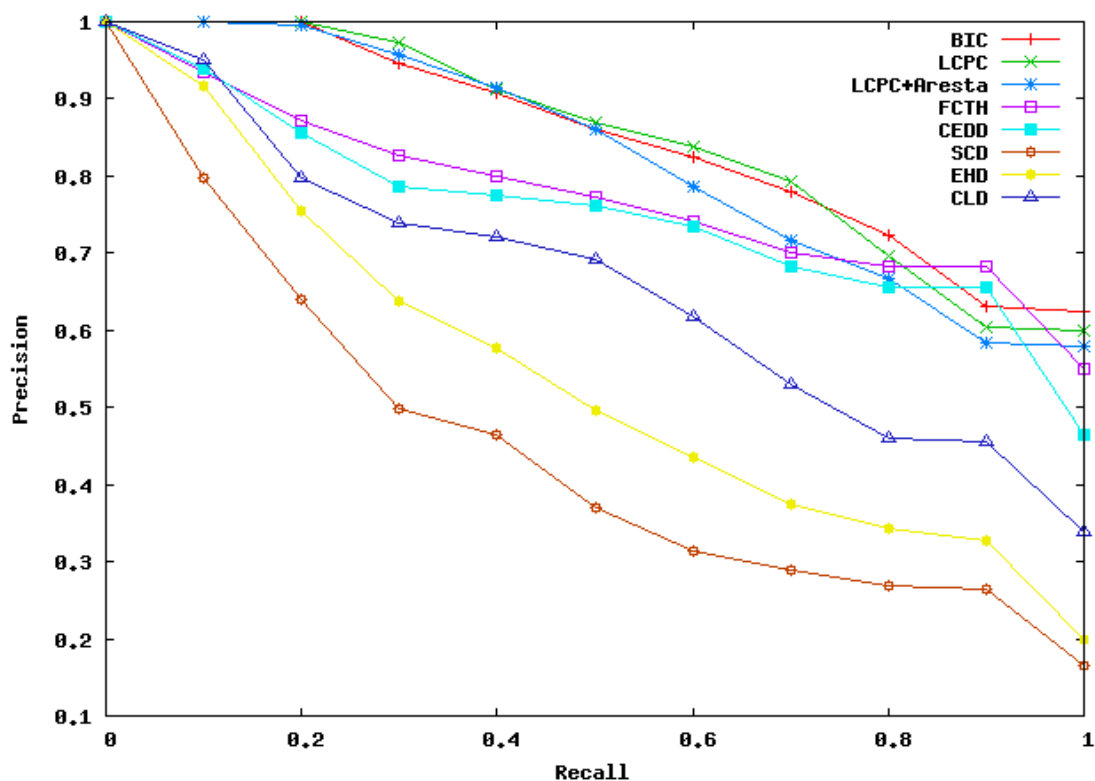


Figura 4.6: Curvas de Precisão x Revocação para os descritores na base CCD.

os melhores resultados entre todas as técnicas avaliadas, inclusive sendo superiores aos descritores.

Para confirmar o comportamento apresentado nas curvas de Precisão x Revocação, calculamos os valores de MAP, apresentados na Tabela 4.11, onde é comprovado que os descritores LCPC, LCPC+Aresta e BIC alcançam altos valores de precisão, enquanto que as técnicas FCTH e CEDD se apresentam melhores que os métodos do MPEG-7, mas não superam os três primeiros descritores.

Descritor	LCPC	BIC	LCPC+Aresta	FCTH	CEDD	CLD	EHD	SCD
MAP	0,84	0,84	0,82	0,71	0,67	0,56	0,42	0,35

Tabela 4.11: Valores de MAP para os descritores na base CCD.

### Validação

Para validar os resultados obtidos na base CCD e saber se os descritores LCPC, LCPC+Aresta são realmente superiores aos métodos FCTH, CEDD e descritores do MPEG-7, aplicamos o teste estatístico. A Tabela 4.12 mostra estes resultados.

Pela Tabela 4.10, percebemos que os métodos LCPC, LCPC+Aresta e BIC não são significativamente diferentes. Esse resultado era esperado, uma vez que, na Figura 4.6, que mostra as curvas

Descritor		<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>
LCPC	<b>1</b>	0	84	76	100	99	100	100	100
LCPC+Aresta	<b>2</b>	84	0	15	100	98	100	100	100
BIC	<b>3</b>	76	15	0	100	100	100	100	100
CEDD	<b>4</b>	100	100	100	0	70	98	100	100
FCTH	<b>5</b>	99	98	100	70	0	99	100	100
CLD	<b>6</b>	100	100	100	98	99	0	100	100
EHD	<b>7</b>	100	100	100	100	100	100	0	65
SCD	<b>8</b>	100	100	100	100	100	100	65	0

Tabela 4.12: Validação dos descritores na base CCD.

de Precision x Recall, é possível observar que as curvas dos três métodos estão muito próximas. Por outro lado, as três abordagens são superiores aos outros métodos com 100% de confiança, com exceção apenas do FCTH, em que o método LCPC é superior a ele com 99% (linha 5 e coluna 1) e a superioridade do LCPC+Aresta em relação ao FCTH é de 98% de confiança (linha 2 e coluna 5).

#### 4.3.2.4 Resultados na base Yahoo

De acordo com os resultados obtidos na base Wang e CCD, os descritores com melhor eficácia foram: LCPC, LCPC+Aresta, BIC, CEDD e FCTH. Por esse motivo, estes foram os escolhidos para serem avaliados na base Yahoo. Essa base é grande e bastante heterogênea, e por essa razão, nesse tipo de ambiente, a precisão dos descritores cai consideravelmente, pois os métodos tendem a se confundir mais facilmente, principalmente, os que baseiam-se apenas em evidência de cor.

Para realizar os experimentos, usamos um total de 30 consultas de diferentes categorias e um conjunto-resposta contendo 30 imagens para cada consulta. Como essa base não possui classificação prévia, contamos com a ajuda de 18 avaliadores, sendo 6 para cada descritor. Foram avaliados 5 descritores, devido ao número reduzido de avaliadores, alguns deles avaliaram mais de um descritor para completar o mínimo de 6 avaliações para cada método.

Como métricas de avaliações, aplicamos primeiramente o método de pooling para extrair, através das avaliações dos usuários, o conjunto de relevante, dessa forma, foi possível calcular as medidas de MAP e curvas de Precisão x Revocação mostradas na Figura 4.7. Em seguida aplicamos o teste de validação para confirmação dos resultados.

#### Eficácia

Percebemos pela Figura 4.7 que a curva de Precisão x Revocação do método LCPC foi superior aos demais métodos, e a segunda maior curva foi a combinação das evidências de cor e forma, o método LCPC+Aresta, logo depois aparece a curva do BIC muito próxima do método

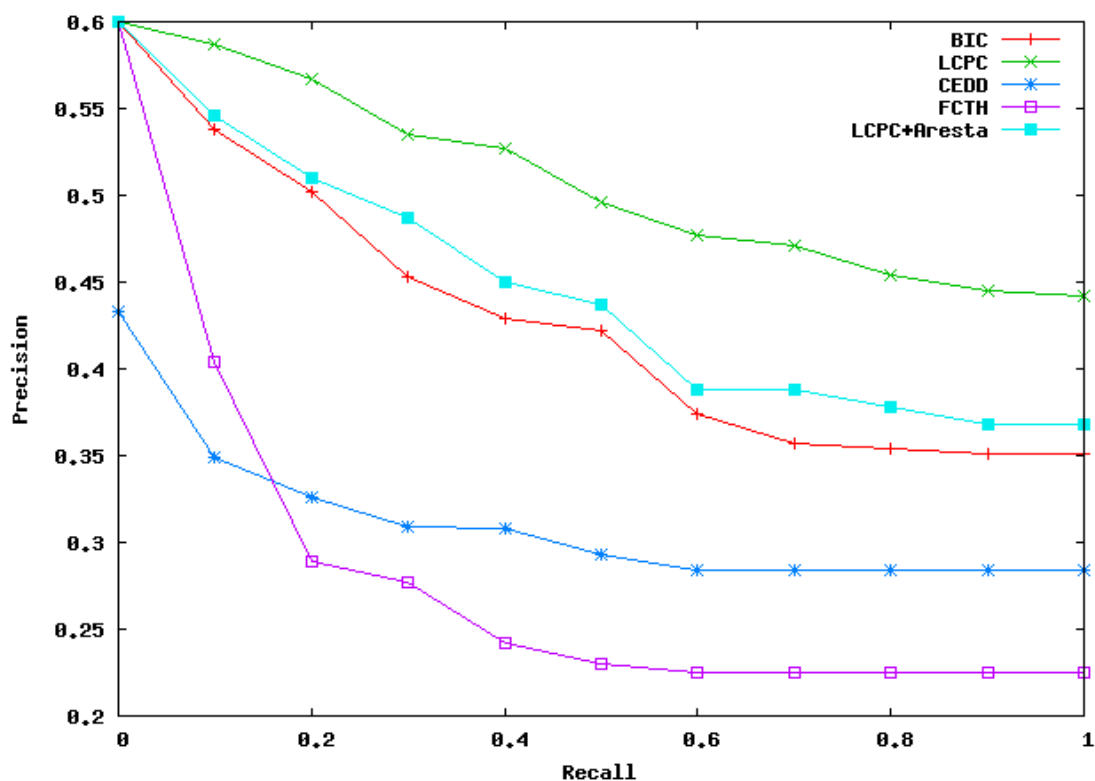


Figura 4.7: Curvas de Precisão x Revocação para os descritores na base Yahoo.

LCPC+Aresta. Um outro ponto a ser observado é que enquanto o nível de precisão nas bases menores (Wang e CCD) foi bastante alto, na base Yahoo os valores de precisão estão abaixo dos 60%, isso acontece devido ao tamanho da base e à sua heterogeneidade.

Vale ressaltar que apesar do descritor LCPC+Aresta não ter sido superior ao LCPC, ele é capaz de retornar imagens relevantes que possuam cores diferentes da imagem de consulta. Essa situação é ilustrada nas Figuras 4.8 e 4.9 que mostram os resultados retornados pelo LCPC e LCPC+Aresta, onde a imagem usada como consulta é um carro amarelo, nesta situação o método que extrai apenas evidência de cor (LCPC), retorna apenas imagens que possuam a cor amarela, enquanto que a combinação das evidências de cor e forma (LCPC+Aresta) é capaz de inserir no conjunto-resposta imagens relevantes que possuam outras cores.

Nas Figuras 4.8 e 4.9, a primeira imagem mostrada como resposta também foi usada como consulta. Essa mesma consulta foi utilizada para os descritores LCPC e LCPC+Aresta, entretanto podemos observar que nas respostas retornadas pelo LCPC+Aresta existem imagens relevantes constituídas de outras cores. Poder ter esta diversidade no conjunto-resposta é um dos grandes benefícios de incluir propriedades de forma para realizar busca de imagens.

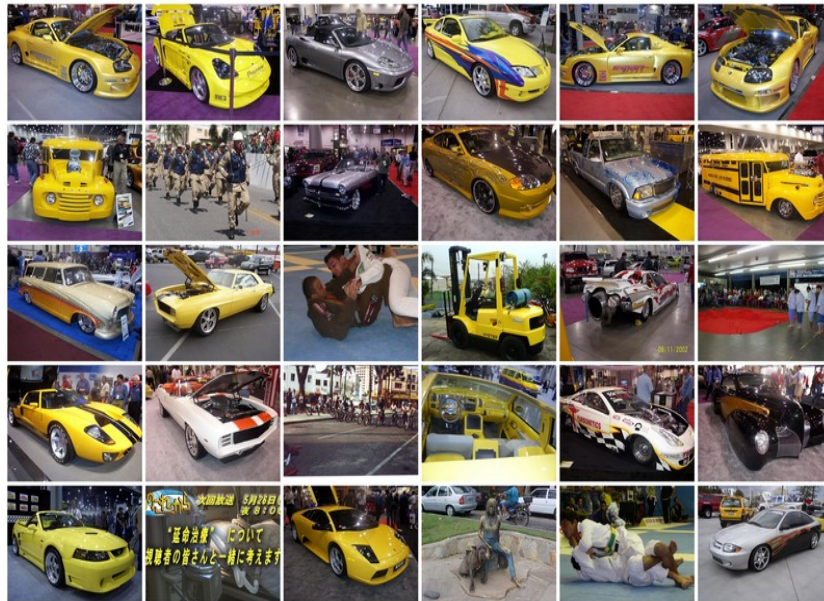


Figura 4.8: 30 imagens de resposta do descritor LCPC+Aresta na base Yahoo.



Figura 4.9: 30 imagens de resposta do descritor LCPC na base Yahoo.

Em geral, o descritor LCPC+Aresta foi superior em imagens onde os objetos existentes eram constituídos de uma maior quantidade de retas como casas, prédios e em situações em que as cores dos objetos presentes no centro da imagem eram semelhantes às cores presentes em seu plano de fundo. Para algumas categorias e alguns de tipos de imagens o método LCPC+Aresta foi superior, como por exemplo, as imagens mostradas na Figura 4.10, entretanto o LCPC+Aresta não superou o LCPC, que no geral foi o melhor entre todos os descritores avaliados neste trabalho.



Figura 4.10: Exemplos de consultas em que o LCPC+Aresta retornou os melhores resultados.

Para analisar melhor os resultados de eficácia obtidos pelos descritores na base Yahoo, calculamos as medidas de MAP que são mostradas na Tabela 4.13. Nesta mesma tabela também são mostradas as porcentagens de ganho do método LCPC em relação aos demais descritores.

Descritor	MAP	Ganhos do LCPC
LCPC	0,41	-
LCPC+Aresta	0,37	+10,81%
BIC	0,27	+51,85%
CEDD	0,21	+95,24%
FCTH	0,20	+105,00%

Tabela 4.13: MAP e porcentagem de ganho dos descritores na base Yahoo.

Como podemos observar na Tabela 4.13, os ganhos do método LCPC são consideravelmente maiores que os demais descritores. O LCPC também foi superior ao BIC que se mostrou superior a vários outros descritores como mostrado em [46]. As diferenças entre LCPC e BIC são evidentes na maioria das consultas, ou seja, a quantidade de respostas relevantes retornadas pelo LCPC é bem maior que a quantidade retornada pelo BIC. Para exemplificar, as Figuras 4.11 e 4.12, mostram as respostas retornadas pelos descritores LCPC e BIC quando usamos como consulta uma imagem da categoria flor.

Ao observarmos as Figuras 4.11 e 4.12, verificamos que a quantidade de respostas relevantes providas pelo LCPC é consideravelmente maior que a do BIC. A maior eficácia do método LCPC deve-se a análise local realizada no momento da extração das características, quando analisamos localmente a imagem, é possível ter maior precisão sobre o seu conteúdo, dessa forma, aumentamos a acurácia do método.

### Validação

Para confirmação dos resultados obtidos pelo LCPC na base Yahoo, aplicamos o método estatístico. Através dele podemos comprovar a superioridade do descritor LCPC, os resultados podem ser verificados na Tabela 4.14.

A Tabela 4.14 mostra que o LCPC é superior aos descritores BIC, CEDD e FCTH com nível de 100% de confiança, em outras palavras, os ganhos obtidos pelo LCPC são realmente significa-





Figura 4.11: 30 imagens de resposta do descritor LCPC na base Yahoo.

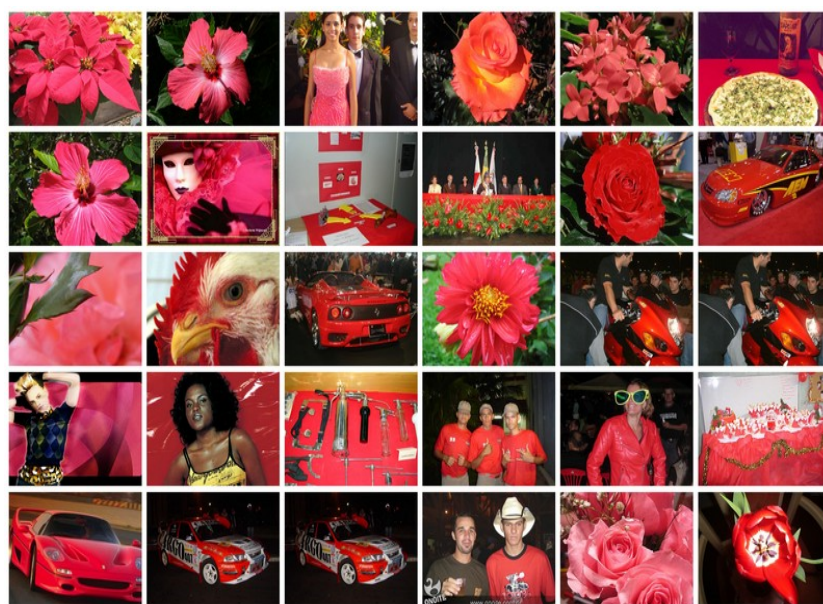


Figura 4.12: 30 imagens de resposta do descritor BIC na base Yahoo

Descritor		1	2	3	3	4
LCPC	1	0	94	100	100	100
LCPC+Aresta	2	94	0	70	99	99
BIC	3	100	70	0	89	99
CEDD	4	100	99	89	0	14
FCTH	5	100	99	89	14	0

Tabela 4.14: Validação dos descritores na base Yahoo.



---

tivos. Em relação o LCPC+Aresta, o nível de confiança entre eles é de 94%, que também pode ser considerado significativo, ou seja, o LCPC também é superior ao LCPC+Aresta. Apesar da combinação das características de cor e forma não ter sido melhor que o descritor de cor, o descritor LCPC+Aresta possui resultados interessantes que ainda podem ser melhorados em trabalhos futuros.

## Capítulo 5

# Conclusões e Trabalhos futuros

Com a popularidade dos dispositivos de aquisição de imagens e conteúdo multimídia em geral, é crescente a necessidade de sistemas que possam buscar e processar este conteúdo de forma eficiente e eficaz com o objetivo de atender aos diferentes interesses de diversos usuários.

Este trabalho propôs de um descritor para busca de imagens, a fim de encontrar um método com capacidade de realizar buscas de imagens pertencentes a qualquer domínio e em repositórios de qualquer tamanho.

Com o objetivo de desenvolver um descritor com alta acurácia, analisamos a evidência de cor localmente, particionando a imagem em partições estratégicas. Para não afetar o desempenho do método, foi realizado um particionamento pré-definido, tentando manter um equilíbrio entre eficácia e eficiência. Dessa forma, realizamos um estudo para encontrar o melhor esquema de particionamento a ser inserido no método proposto. Com o intuito de empregar abordagem proposta em bases de qualquer tamanho, os experimentos foram executados em bases pequenas, comumente usadas na literatura, e em base grande constituída de diversas categorias, representando fielmente os repositórios de imagens encontrados na Web.

Em ambientes grandes e heterogêneos, os descritores que usam apenas evidências de cor, tendem a se confundir mais facilmente, ocasionando assim, uma queda na qualidade das respostas providas por ele. Pensando assim, a abordagem proposta foi combinada com características de forma, entretanto a combinação de cor e forma não superou os ótimos resultados obtidos pela abordagem proposta LCPC, por outro lado, a combinação através do LCPC+Aresta retornou resultados interessantes, diversificando as cores das imagens relevantes no conjunto-resposta.

Para comprovar os ótimos resultados obtidos pelo LCPC, ele foi comparado a outros descritores

comumente citados na literatura, sendo assim, três métodos do padrão MPEG-7 (SCD, CLD e EHD) foram incluídos na avaliação. Também incluímos as técnicas CEDD e FCTH, pois de acordo com os experimentos realizados em [46], eles superam os métodos do MPEG-7 em termos de eficácia. Realizamos os testes comparativos e comprovamos que o método LCPC foi superior aos demais, principalmente, na base contendo mais de 100 mil imagens. Como era esperado, nas bases menores, os descritores alcançaram altos níveis de precisão, com valores entre 70% e 100% em 30% de revocação. Com exceção dos descritores do MPEG-7, as curvas dos descritores tendem a permanecer muito próximas, isso acontece devido a base ser pequena e bem comportada. Entretanto, quando os descritores são empregados em ambientes grandes e heterogêneos, o nível de precisão tende a baixar consideravelmente e as diferenças entre os métodos são bem mais visíveis. Por exemplo, na base de 100 mil imagens, os descritores LCPC e LCPC+Forma alcançaram valores de precisão entre 60% e 30% em até 30% de revocação, enquanto que o FCTH e CEDD alcançaram precisão entre 30% e 20% em até 30% de revocação.

Através do cálculo de MAP, confirmamos as diferenças entre os descritores avaliados na base do Yahoo sendo que o LCPC alcançou uma média de MAP de 0,41, o BIC obteve 0,27, FCTH alcançou 0,21 e o CEDD obteve 0,20, mostrando que existe uma grande diferença entre o LCPC e os demais métodos. Os testes estatísticos também comprovaram a superioridade do LCPC com nível de 100% de confiança. Todos esses valores comprovam nosso pressuposto de que em bases pequenas os descritores alcançam altos níveis de precisão, enquanto que em ambientes grandes e heterogêneos os níveis de precisão tendem a ser baixos, entretanto mesmo que o LCPC não tenha alcançado altos níveis de precisão, as médias obtidas por ele foram superiores a vários descritores comumente citados na literatura.

## 5.1 Trabalhos futuros

Apesar dos bons resultados adquiridos com o desenvolvimento da abordagem proposta, ainda existem muitos pontos que podem ser explorados para melhorar a eficácia dos descritores de busca, de forma que eles possam ter um bom equilíbrio entre eficiência e eficácia. Alguns pontos são citados a seguir:

- Estudo sobre forma de particionar a imagem:

O foco deste trabalho foi em encontrar a melhor forma de particionar uma imagem, mas sim no desenvolvimento de um descritor. É de extrema importância o estudo profundo da

melhor forma de particionar a imagem. Outros esquemas fixos podem ser avaliados, como por exemplo, imagens divididas em grade 2x2, 3x3 e assim por diante. Também seria interessante explorar outros algoritmos de segmentação para decompor a imagem, neste trabalho, fizemos uma tentativa de decompor a imagem com apenas um algoritmo de segmentação, entretanto, existem vários que podem ser estudados [7, 32, 43]. Um outro estudo interessante, seria a inclusão de técnicas de aprendizagem para mudar o esquema de particionamento de acordo com o tipo da imagem.

- Explorar outros parâmetros do descritor de forma:

O método de extração de características de forma que foi utilizado neste trabalho possui um grande potencial, uma vez que, seus resultados foram interessantes e superou descritores como FCTH e CEDD. Desse modo, é interessante, como trabalho futuro, explorar melhor seus parâmetros como aumentar a quantidade de arestas usadas, explorar melhor o limiar de binarização, definir o tamanho mínimo para as arestas em termos de quantidade de pixels, incluir a relação quantidade de pixels de aresta pela quantidade total de pixels na imagem, testar outras funções de distância e etc. Realizando esses testes o método pode obter maior acurácia, o que não foi possível fazer nesta dissertação, pois a cada modificação de um parâmetro é necessário realizar novos testes de avaliação, um processo que demanda bastante tempo. Mas pelos resultados obtidos nos experimentos, acredita-se que explorando melhor o descritor, podemos alcançar resultados promissores.

- Pesquisar técnicas de combinação das evidências:

Neste trabalho, foi aplicada apenas uma maneira de combinar as características de cor e forma. Desse modo, cabe investigar outras técnicas que possam melhorar a eficácia dos descritores como, por exemplo, Programação Genética [18]. Uma outra possível investigação, seria incluir na combinação, outras evidências como textura.

- Estudo comparativo em relação ao desempenho dos descritores:

Nesta dissertação foram realizados vários testes relacionados a eficácia dos descritores, dessa forma, é necessário avaliá-los com relação ao desempenho em grandes bases de imagens, pois quanto maior a base, maior é tempo necessário para realizar uma busca, ou seja, com o estudo de desempenho será possível saber se os descritores são capazes de serem empregados em ambientes reais como os grandes repositórios da Web.

- Pesquisar formas de indexar os vetores de características do método proposto:

Uma forma simples de aumentar o desempenho do método LCPC, seria paralelizar a extração de seus vetores de características, pois são extraídos um histograma para cada uma das cinco partições, com a inserção de paralelização tanto no momento de extrair histogramas quanto no momento do cálculo de similaridade é possível aumentar o desempenho do método e conseqüentemente diminuir o tempo de espera que um usuário teria ao realizar a busca. Além da paralelização, um outro estudo promissor seria explorar técnicas de indexação dos histogramas, semelhante à indexação realizada em buscas textuais, com a indexação seria possível aumentar consideravelmente o tamanho da base de imagens sem afetar o desempenho do método.

# Referências Bibliográficas

- [1] Flickr blog. geotagging - one day later. [http://blog.flickr.com/flickrblog/2006/08/geotagging/\\_one/\\_html](http://blog.flickr.com/flickrblog/2006/08/geotagging/_one/_html), August 2006.
- [2] Brodatz. <http://www.cs.ucr.edu/~bcampana/texture/>, April 2011.
- [3] Caltech 256. [http://www.vision.caltech.edu/Image\\_Datasets/Caltech256/](http://www.vision.caltech.edu/Image_Datasets/Caltech256/), April 2011.
- [4] Mpeg-7 image retrieval refinement based on relevance feedback. <http://www.ee.cityu.edu.hk/~mirror/>, April 2011.
- [5] R. Adams and L. Bischof. Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(6):641–647, jun 1994.
- [6] T. W. Anderson and J. D. Finn. *The New Statistical Analysis of Data*. Springer, 1996.
- [7] S. Avula, J. Tang, and S. Acton. Image retrieval using segmentation. In *Systems and Information Engineering Design Symposium, 2003 IEEE*, pages 289–294, april 2003.
- [8] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice Hall, 1982.
- [9] 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, ICVS'08*, pages 312–322, Berlin, Heidelberg, 2008. Springer-Verlag.
- [10] S. A. Chatzichristofis and Y. S. Boutalis. FCTH: Fuzzy Color and Texture Histogram - A Low Level Feature for Accurate Image Retrieval. pages 191–196, 2008.
- [11] 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*, 24, 2010.

- [12] Y. Chen, J. Z. Wang, and D. Geman. Image categorization by learning and reasoning with regions. *Journal of Machine Learning Research*, 5:913–939, 2004.
- [13] L. Cieplinski. Mpeg-7 color descriptors and their applications. In *Proceedings of the 9th International Conference on Computer Analysis of Images and Patterns, CAIP '01*, pages 11–20, London, UK, 2001. Springer-Verlag.
- [14] A. Del Bimbo. *Visual information retrieval*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.
- [15] J. dos Santos, C. Ferreira, R. da S. Torres, M. Gonçalves, and R. Lamparelli. A relevance feedback method based on genetic programming for classification of remote sensing images. *Information Sciences*, In Press, Corrected Proof:–, 2010.
- [16] M. Douze, H. Jégou, H. Sandhawalia, L. Amsaleg, and C. Schmid. Evaluation of gist descriptors for web-scale image search. In *Proceeding of the ACM International Conference on Image and Video Retrieval, CIVR '09*, pages 19:1–19:8, New York, NY, USA, 2009. ACM.
- [17] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Proceedings of the 7th European Conference on Computer Vision-Part IV, ECCV 02*, pages 97–112, London, UK, UK, 2002. Springer-Verlag.
- [18] C. D. Ferreira, R. da S. Torres, M. A. Gonçalves, and W. Fan. Image retrieval with relevance feedback based on genetic programming. In *Proceedings of the 23rd Brazilian symposium on Databases, SBBD '08*, pages 120–134, Porto Alegre, Brazil, Brazil, 2008. Sociedade Brasileira de Computação.
- [19] T. Gevers and A. Smeulders. Pictoseek: combining color and shape invariant features for image retrieval. *Image Processing, IEEE Transactions on*, 9(1):102–119, jan 2000.
- [20] D. Hawking, N. Craswell, and P. Thistlewaite. Overview of trec-7 very large collection track. In *TREC*, pages 40–52, 1998.
- [21] D. Hawking, N. Craswell, P. Thistlewaite, and D. Harman. Results and challenges in web search evaluation. In *WWW*, pages 1321–1330, 1999.

- [22] E. Horster and R. Lienhart. Deep networks for image retrieval on large-scale databases. In *MM 08: Proceeding of the 16th ACM international conference on Multimedia*, pages 643–646, Vancouver, British Columbia, Canada, 2008. ACM.
- [23] E. Horster, R. Lienhart, and M. Slaney. Image retrieval on large-scale image databases. In *Proceedings of the 6th ACM international conference on Image and video retrieval, CIVR '07*, pages 17–24, New York, NY, USA, 2007. ACM.
- [24] Y. Huang, Q. Liu, S. Zhang, and D. Metaxas. Image retrieval via probabilistic hypergraph ranking. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3376–3383, june 2010.
- [25] O. C.-B. Image, T. M. Lehmann, M. O. Güld, C. Thies, B. Fischer, K. Spitzer, A. Daniel Keysers, A. Hermann Ney, B. Michael Kohlen, B. Henning Schubert, and B. Berthold B. Wein. The irma project: A state of the art report. In *In Korea-Germany Workshop on Advanced Medical Image*, pages 161–171, 2003.
- [26] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, 1989.
- [27] N. Kanopoulos, N. Vasanthavada, and R. L. Baker. Design of an image edge detection filter using the sobel operator. *IEEE Journal of Solid-state Circuits*, 23:358–367, 1988.
- [28] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas. Fast and effective retrieval of medical tumor shapes. *Knowledge and Data Engineering, IEEE Transactions on*, 10(6):889–904, nov/dec 1998.
- [29] F.-F. Li and P. Perona. A Bayesian Hierarchical Model for Learning Natural Scene Categories. In *CVPR 05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 05) - Volume 2*, volume 2, pages 524–531, Washington, DC, USA, 2005. IEEE Computer Society.
- [30] J. Li, J. Z. Wang, and G. Wiederhold. Irm: Integrated region matching for image retrieval. pages 147–156. In *Proc. of ACM Multimedia Intl. Conf.*, 2000.
- [31] X. Li. Image retrieval based on perceptive weighted color blocks. *Pattern Recogn. Lett.*, 24:1935–1941, August 2003.
- [32] Z. Li. *A stochastic segmentation method for interesting region detection and image retrieval*. PhD thesis, Charlotte, NC, USA, 2009. AAI3372075.



- [33] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International J. Comput. Vision*, 60:91–110, November 2004.
- [34] G. Lu. *Multimedia Database Management*. Artech House, 1999.
- [35] T.-C. Lu and C.-C. Chang. Color image retrieval technique based on color features and image bitmap. *Inf. Process. Manage.*, 43:461–472, March 2007.
- [36] M. Lux and S. A. Chatzichristofis. Lire: lucene image retrieval: an extensible java cbir library. In *MM*, pages 1085–1088, 2008.
- [37] B. Manjunath, J.-R. Ohm, V. Vasudevan, and A. Yamada. Color and texture descriptors. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):703–715, jun 2001.
- [38] D. Marceau, P. Howarth, J. Dubois, and D. Gratton. Evaluation of the grey-level co-occurrence matrix method for land-cover classification using spot imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 28(4):513–519, jul 1990.
- [39] K. H. T. B. Mathias Eitz and M. Alexa. A descriptor for large scale image retrieval based on sketched feature lines. In *Eurographics Symposium on Sketch-Based Interfaces and Modeling*, pages 29–38, 2009.
- [40] H. A. Moghaddam, T. T. Khajoie, A. H. Rouhi, and M. S. Tarzjan. Wavelet correlogram: A new approach for image indexing and retrieval. *Pattern Recognition*, 38(12):2506–2518, 2005.
- [41] J. Montagnat and H. Delingette. Volumetric medical images segmentation using shape constrained deformable models. In J. Troccaz, E. Grimson, and R. Mösges, editors, *CVRMed-MRCAS'97*, volume 1205 of *Lecture Notes in Computer Science*, pages 13–22. Springer Berlin / Heidelberg, 1997. 10.1007/BFb0029220.
- [42] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42:145–175, 2001.
- [43] M. Ozden and E. Polat. A color image segmentation approach for content-based image retrieval. *Pattern Recogn.*, 40:1318–1325, April 2007.
- [44] G. Paschos, I. Radev, and N. Prabakar. Image content-based retrieval using chromaticity moments. *IEEE Trans. on Knowl. and Data Eng.*, 15(5):1069–1072, 2003.

- [45] O. A. B. Penatti and R. da Silva Torres. Eva: an evaluation tool for comparing descriptors in content-based image retrieval tasks. pages 413–416. *Multimedia Information Retrieval*, 2010.
- [46] O. B. Penatti and R. da S. Torres. Color descriptors for web image retrieval: a comparative study. Campo Grande, MS, October 2008. XXI Brazilian Symposium on Computer Graphics and Image Processing.
- [47] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, june 2008.
- [48] B. G. Prasad, K. K. Biswas, and S. K. Gupta. Region-based image retrieval using integrated color, shape, and location index. *Comput. Vis. Image Underst.*, 94:193–233, April 2004.
- [49] G. Schaefer and M. Stich. Ucid - an uncompressed colour image database. In *In Storage and Retrieval Methods and Applications for Multimedia 2004, volume 5307 of Proceedings of SPIE*, pages 472–480, San Jose, USA, 2004.
- [50] N. Sebe, M. S. Lew, and D. P. Huijsmans. Multi-scale sub-image search. pages 79–82. *Proceedings of the seventh ACM international conference on Multimedia*, 1999.
- [51] T. Sikora. The MPEG-7 visual standard for content description - An overview. *IEEE Trans. Circuits and Systems for Video Technology*, 11(6):696–702, 2001.
- [52] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision*. Thomson, 1998.
- [53] R. O. Stehling, A. X. Falcão, and M. A. Nascimento. An adaptive and efficient clustering-based approach for content-based image retrieval in image databases. pages 356–365. In *Proc. of IDEAS Intl. Symposium*, 2001.
- [54] R. O. Stehling, M. A. Nascimento, and A. X. Falcaão. A compact and efficient image retrieval approach based on border/interior pixel classification. In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, pages 102–109, New York, NY, USA, 2002. ACM.
- [55] R. O. Stehling, M. A. Nascimento, and A. X. Falcão. Cell histograms versus color histograms for image representation and retrieval. pages 315–336. *Knowledge and Information Systems Intl. Journal*, 2003.

- [56] M. Stricker, A. Dimai, and E. Dimai. Color indexing with weak spatial constraints. In *in Proc. SPIE Storage and Retrieval for Image and Video Databases*, pages 29–40, 1996.
- [57] M. A. stricker and M. orengo. Similarity of color images. In W. niblack and R. C. jain, editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 2420, pages 381–392, 1995.
- [58] J. Sun, X. Zhang, J. Cui, and L. Zhou. Image retrieval based on color distribution entropy. *Pattern Recogn. Lett.*, 27:1122–1126, July 2006.
- [59] R. D. S. Torres and A. X. Falcão. Content-based image retrieval: Theory and applications. *Revista de Informática Teórica e Aplicada*, 13:161–185.
- [60] F. Ulaby, F. Kouyate, B. Brisco, and T. Williams. Textural information in sar images. *Geoscience and Remote Sensing, IEEE Transactions on*, GE-24(2):235–245, march 1986.
- [61] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics*, 1(6):80–83, 1945.
- [62] A. Williams and P. Yoon. Content-based image retrieval using joint correlograms. *Multimedia Tools Appl*, 34(2):239–248, 2007.
- [63] X. Zhang, Z. Li, L. Zhang, W.-Y. Ma, and H.-Y. Shum. Efficient indexing for large scale visual search. pages 1103–1110, Kyoto, sep 2009.