

de um ambiente de DWing de imagens, bem como um esquema estrela estendido genérico. As considerações finais são feitas na seção 3.6.

3.1 Introdução

Com relação especificamente ao armazenamento de imagens em ambientes de DWing, existem poucos trabalhos na literatura. Wong *et al.* (Wong et al., 2004) propõem um sistema de DW adaptado a dados multimídia que permite a realização de estudos clínicos e científicos sobre dados de pacientes. Arigon *et al.* (Arigon et al., 2007) propõem um modelo multidimensional capaz de lidar com dados multimídia representados por diferentes descritores. Chen *et al.* (Chen et al., 2008) propõem uma estrutura XML para organizar dados multimídia, especialmente vídeos, em um cubo de dados. Para isso, é descrito um esquema estrela particular de DW, bem como um algoritmo de consulta por abrangência. Jin *et al.* (Jin et al., 2010) descrevem um cubo visual e uma ferramenta OLAP multidimensional de coleções de imagens. Dessa forma, informações presentes em imagens *web*, indexadas por máquinas de busca ou fotos compartilhadas em redes sociais, podem ser analisadas estatisticamente.

Como pode ser observado, ainda não existe um consenso sobre a definição de um DW de imagens com relação ao esquema, dados e processamento de consultas OLAP. Neste trabalho, é utilizada a definição de DW de imagens proposta em (Annibal et al., 2010), a qual é descrita na Definição 1. O diferencial dessa definição com relação às demais é que ela é mais flexível, desde que considera a imagem como um dado intrínseco do DW.

Definição 1. *Um DW de imagens é projetado segundo um esquema estrela diferenciado que possui, além de tabelas de dimensão com dados convencionais, uma ou mais tabelas de dimensão específicas para a manipulação de dados sobre imagens.*

No trabalho de Annibal et. al (2010), é considerado que os dados de imagem que devem ser armazenados no DW são os vetores de características, bem como os dados necessários para o mecanismo de filtragem de imagens baseado na técnica Omni (seção 3.2.4). Por meio dessa abordagem, é possível a realização de consultas OLAP estendidas com predicados de similaridade de imagens.

3.2 Espaços Métricos e Consultas por Similaridade

Os dados que descrevem uma imagem são representados por vetores de características, que são *arrays* de dimensão n , onde n é o número de características obtidas. Para realizar

as consultas utilizando essas estruturas, é necessário definir uma função que meça o grau de semelhança entre esses vetores. Esse tipo de consulta, baseada em dados “parecidos” e não “iguais”, é conhecida como consulta por similaridade (Traina-Jr et al., 2007).

Relacionado a espaços métricos está o conceito de métodos de acesso métricos (MAMs). MAMs são índices ou técnicas de indexação projetados para atuarem como um caminho otimizado aos dados, evitando a análise exaustiva durante a recuperação dos elementos que satisfazem a uma consulta. Um método de acesso é definido por sua estrutura de dados, bem como por algoritmos de busca e de alteração. A pesquisa existente na literatura sobre MAM é bastante extensa, sendo que exemplos incluem a VP-tree (Yianilos, 1993); MVP-tree (Bozkaya and Ozsoyoglu, 1999); GNAT (Brin, 1995); M-tree e variações (Ciaccia et al., 1997; Skopal and Hoksza, 2007; Skopal and Lokoc, 2008; Zhou et al., 2003, 2005); Slim-tree e variações (Pola et al., 2009; Traina-Jr et al., 2002; Vieira et al., 2004); Onion-tree (Carélo et al., 2009, 2011) e Técnica Omni (Traina-Jr et al., 2007).

Na seção 3.2.1 são introduzidos os conceitos de espaços métricos e vetoriais. Na seção 3.2.2 são listadas as funções de distância mais utilizadas para realização de consultas por similaridade. Na seção 3.2.3 são explicadas algumas técnicas de extração de características de imagens. Na seção 3.2.4 é abordada a técnica Omni, o qual é utilizado na proposta desse projeto de mestrado.

3.2.1 Espaços Métricos e Vetoriais

Na literatura, pelo menos duas técnicas têm sido estudadas para a modelagem dos vetores de características, o modelo de espaço métrico e o modelo de espaço vetorial (Traina-Jr et al., 2007). Um espaço métrico é um par ordenado (\mathcal{U}, d) , onde \mathcal{U} é um conjunto de objetos e $d : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^+$ é uma métrica, isto é, uma medida de distância que obedece às propriedades de identidade ($\forall x \in \mathcal{U}, d(x,x) = 0$); simetria ($\forall x, y \in \mathcal{U}, d(x,y) = d(y,x)$); não-negatividade ($\forall x, y \in \mathcal{U}, d(x,y) \geq 0$); e desigualdade triangular ($\forall x, y, z \in \mathcal{U}, d(x,y) \leq d(x,z) + d(z,y)$) (Ciaccia and Patella, 2002).

Diferente do espaço métrico, que não exige que os objetos possuam tamanho fixo, os objetos pertencentes a um espaço vetorial devem possuir dimensão fixa. Dessa forma, um vetor de características representado nesse modelo pode ser visto como um ponto em um espaço n -dimensional, onde n é o tamanho do vetor (Traina-Jr et al., 2007).

Espaços métricos são adequados para representar certos tipos de objetos como impressões digitais, pois o número de características que podem ser extraídas (deltas, terminações, etc.) variam entre as pessoas, e, portanto, não seria possível mapear um objeto desse domínio em um ponto no espaço n -dimensional (Traina-Jr et al., 2007). Já espa-

ços vetoriais são adequados para representar objetos com número fixo de características, como um histograma de níveis de cinza extraído de uma imagem, cuja dimensionalidade é o número de variações da escala de cinza.

As técnicas e algoritmos voltados à realização de consultas por similaridade descritos nesta dissertação de mestrado focam em operações sobre o modelo de espaço métrico, pois ele é menos restritivo e, munido de funções de distância adequadas, ainda inclui o espaço vetorial (Traina-Jr et al., 2007).

Em espaços métricos, há basicamente dois tipos de consultas de interesse: *range query* e *K-nearest neighbor query* (Chávez et al., 1999), as quais são ilustradas na Figura 3.1. Assim, sendo $q \in \mathcal{U}$ o elemento de consulta e r um número real que indica um raio de tolerância a partir de q (centro de consulta), esses tipos de busca são definidos da seguinte forma:

- **Range query** $R(q, r)$: Consulta por abrangência, a qual retorna todos os elementos que estão dentro da distância de r a q , isto é, $\{u \in \mathcal{U} \mid d(q, u) \leq r\}$.
- **K-nearest neighbor query** $Knn(q)$: Busca pelos vizinhos mais próximos, a qual retorna os k elementos que estão mais perto de q em \mathcal{U} , isto é, um conjunto $A \subseteq \mathcal{U}$, tal que $|A| = k$ e $\forall u \in A, v \in \mathcal{U} - A, d(q, u) \leq d(q, v)$. Sendo que em casos de empate, qualquer conjunto de k elementos que satisfaça a condição é aceitável.

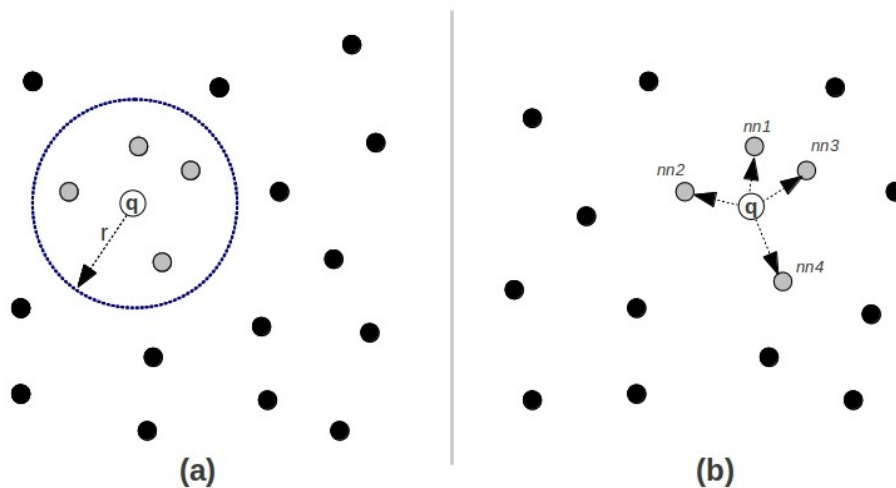


Figura 3.1: Exemplos de consultas em espaços métricos. (a) Consulta por abrangência e (b) consulta pelos 4 vizinhos mais próximos. Adaptado de (Traina-Jr et al., 2007).

3.2.2 Funções de Distância

As funções de distância expressam numericamente o grau de dissimilaridade entre dois objetos. Dessa forma, a similaridade é inversamente proporcional ao resultado dessa função, isto é, quanto menor a distância obtida, maior a similaridade. Apesar de existirem diversas funções de distância para vários tipos de aplicações presentes na literatura, as distâncias da família Minkowski são as mais utilizadas (Wilson and Martinez, 1997). Conhecidas também como norma L_p , essa família de distâncias é válida para espaços vetoriais, nos quais os objetos são definidos por meio de tuplas $\{x_1, x_2, \dots, x_n\}$, onde n é o tamanho do vetor (quantidade de características). E essa família de distâncias é definida pela Equação 3.2.1

$$L_p((x_1, \dots, x_n), (y_1, \dots, y_n)) = \sqrt[p]{\sum_{i=1}^n |x_i - y_i|^p} \quad (3.2.1)$$

Alguns casos particulares desse grupo de funções podem ser obtidos a partir de variações do parâmetro p . Dessa forma, para $p = 2$, tem-se a distância L_2 , conhecida como distância Euclidiana, e para $p = 1$, tem-se a distância L_1 , também chamada de *City Block* ou distância Manhattan. Ao se calcular o limite da função L_p quando p tende ao infinito, obtém-se a função L_∞ , chamada Infinity ou Chebychev. Neste projeto de mestrado utiliza-se a distância Euclidiana.

Cada função de distância da família Minkowski gera, ao redor de um elemento de consulta q , uma forma geométrica diferente formada pelos pontos equidistantes a um dado raio de abrangência r_q . Assim, em uma *range query*, diferentes resultados podem ser obtidos dependendo da função de distância adotada. Na Figura 3.2, é ilustrada a representação espacial das formas geométricas geradas pelas distâncias L_1 , L_2 e L_∞ no espaço bidimensional. Pela figura, é possível notar que, de acordo com a distância escolhida, os elementos contidos no raio de abrangência r_q podem mudar.

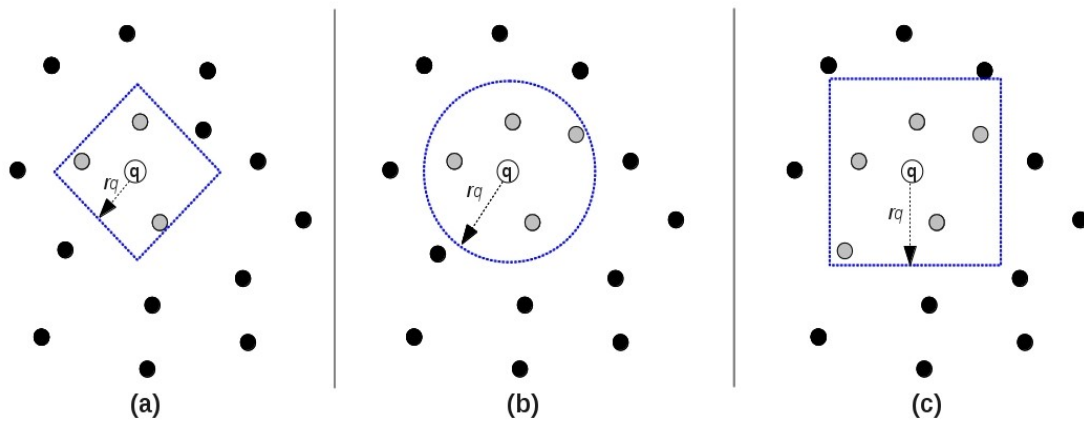


Figura 3.2: Exemplos das formas geométricas geradas pelas distâncias L_p , considerando um espaço bidimensional. (a) Distância *Manhattan* (L_1), (b) Distância Euclidiana (L_2) e (c) Distância *Chebychev* (L_∞).

3.2.3 Extração de Características

Os principais desafios na implementação de sistemas de recuperação de imagens por conteúdo (*content-based image retrieval*, CBIR) são encontrar as representações do conteúdo intrínseco das imagens mais apropriadas e definir uma forma de calcular a similaridade entre duas imagens baseada nessa representação abstrata (Datta et al., 2008).

Nas áreas de processamento de imagens e visão computacional são estudados possíveis descritores de imagens, os quais representam as imagens com base principalmente nas propriedades de cor, textura e forma. Os algoritmos que implementam esses descritores, ou seja, realizam o processo de extração de características, são chamados de extratores. A seguir, são exemplificados alguns descritores de imagens.

- **Cor:** Dentre as técnicas existentes, o histograma de cor é o mais simples e o mais utilizado. Ele representa a distribuição das cores ou níveis de cinza presentes em uma imagem. Por exemplo, para uma imagem em tons de cinza, um histograma h consiste de um vetor de tamanho 255, onde cada posição i representa um nível de cinza, em que a posição $h[i]$ armazena o número de vezes que a intensidade de *pixel* i ocorre na imagem. As vantagens do uso de histogramas são sua facilidade de implementação, rapidez e a possibilidade de extração de várias informações probabilísticas, as quais também podem ser utilizadas como características da imagem. As desvantagens são que histogramas clássicos não captam informações espaciais dos *pixels* e um mesmo histograma pode ser gerado para imagens diferentes, embora cada imagem gere apenas um histograma.

- **Textura:** Não há uma definição formal do conceito de textura, mas pode-se dizer que por essa propriedade é possível identificar padrões repetitivos sobre a superfície da imagem, medindo propriedades tais como aspereza, suavidade e regularidade (Gonzalez and Woods, 2006). Em processamento de imagens, um meio para descrever essas propriedades é a abordagem estatística, sendo que nessa categoria os descritores de Haralick (Haralick, 1979) obtêm ótimos resultados. Os descritores de Haralick são obtidos por meio de matrizes de coocorrência da imagem, as quais representam as relações entre a posição dos *pixels* e seus valores. Dessa forma, uma matriz de coocorrência G é composta por valores g_{ij} que indicam o número de vezes que *pixels* de intensidade i e j aparecem na imagem de acordo com uma posição relativa pré-determinada, isto é, a partir de uma direção, tipicamente $0^\circ, 45^\circ, 90^\circ, 135^\circ$, e da distância entre *pixels*. Na Tabela 3.1, são ilustrados os principais descritores de Haralick utilizados na literatura.

Descritor	Equação
Variância	$\sum_i \sum_j (i - j)^2 G(i, j)$
Entropia	$\sum_i \sum_j G(i, j) \log(G(i, j))$
Energia	$\sum_i \sum_j G^2(i, j)$
Homogeneidade	$\sum_i \sum_j \frac{G(i, j)}{(1 + i - j)}$
3º momento	$\sum_i \sum_j (i - j)^3 G(i, j)$
Inversão da variância	$\sum_i \sum_j \frac{G(i, j)}{(i - j)^2}$

Tabela 3.1: Principais descritores de Haralick (Haralick et al., 1973).

- **Forma:** Existem vários descritores de forma existentes na literatura, os quais são categorizados em descritores baseados em contorno e descritores baseados em região. Enquanto no primeiro apenas a borda dos objetos é explorada, no segundo todos os *pixels* dentro de uma região são levados em conta no descritor, sendo, portanto, mais confiáveis para formas que possuam contornos mais complexos (Amanatiadis et al., 2011; Kim and Kim, 2000).

3.2.4 Técnica Omni

Consultas por similaridade entre objetos complexos são geralmente muito caras, e portanto, a redução do número de comparações é uma questão essencial para esse tipo de aplicação. Assim, para tornar essas operações mais rápidas, MAMs têm sido utilizados de maneira muito promissora, pois proporcionam acesso eficiente aos dados, por meio da redução do espaço de busca e, por conseguinte, da diminuição dos cálculos de distância

necessários. Dessa forma, com o auxílio dos MAMs, as buscas são restringidas às porções do conjunto de dados nas quais os objetos armazenados têm maior probabilidade de serem similares a um elemento de consulta (Carélo et al., 2009).

A técnica Omni consiste de um mecanismo de filtragem baseado no uso de representantes globais do conjunto de dados (elementos representativos). Para cada consulta, esses elementos especificam uma região do espaço métrico na qual garantidamente os objetos mais similares ao objeto de consulta residem. Dessa forma, uma vez que a poda no espaço métrico foi realizada, os objetos pertencentes à região encontrada são considerados candidatos à resposta da consulta. Apesar do método garantir que não haverá falsos negativos, devido à propriedade da desigualdade triangular, ele não assegura a ausência de falsos positivos (Traina-Jr et al., 2007). Portanto, apenas a definição da região não é suficiente, sendo necessária uma segunda etapa de refinamento.

De maneira mais detalhada, a principal ideia da técnica Omni consiste em selecionar um conjunto de h elementos como representantes globais e armazenar, para os demais objetos da base de dados, as respectivas distâncias a cada um dos representantes escolhidos. O conjunto dos representantes globais é descrito formalmente na Definição 2.

Definição 2. *Dado um espaço métrico $\mathbb{M} = \{\mathbb{S}, d()\}$, o conjunto \mathcal{F} dos elementos representativos é definido por: $\mathcal{F} = \{f_1, f_2, \dots, f_h \mid f_g, f_h \in \mathbb{S}, f_g \neq f_j, \forall g \neq j\}$, onde cada f_g é um representante de \mathbb{S} e h é o número de representantes.*

As consultas por abrangência realizadas utilizando essa técnica passam por duas etapas: etapa de filtragem e refinamento. A primeira consiste na seleção de objetos candidatos à resposta com o auxílio dos representantes globais. Esse processo é realizado por meio da especificação de uma *mbOr* (*minimum-bounding-Omni-region*). Dado um conjunto de representantes $\mathcal{F} = \{f_1, f_2, \dots, f_h\}$, $\mathcal{F} \subset \mathbb{S}$, a *mbOr* de um elemento de consulta $s_q \in \mathbb{S}$ com raio r_q , é definida por:

$$mbOr_{\mathcal{F}}(s_q, r_q) = \bigcap_{g=1}^h I_g \quad (3.2.2)$$

onde I_g é um subconjunto de objetos $s_i \in S (S \subseteq \mathbb{S})$ que pertencem ao intervalo fechado $[I_g^{inf}, I_g^{sup}]$ formado pelo representante g , onde $I_g^{inf} = d(f_g, s_q) - r_q$, se $d(f_g, s_q) \geq r_q$ ou $I_g^{inf} = 0$, caso contrário; e $I_g^{sup} = d(f_g, s_q) + r_q$, sendo $d(a, b)$ a distância entre os objetos. Formalmente:

$$I_g = \{s_i \mid I_g^{inf} \leq d(f_g, s_i) \leq I_g^{sup}, \forall s_i \in \mathbb{S}\} \quad (3.2.3)$$

Na Figura 3.3 é ilustrada a definição de uma *mbOr* considerando um conjunto de dados para o qual dois elementos representativos foram selecionados, f_1 e f_2 . De acordo com a figura, dada uma imagem de consulta s_q e um raio de abrangência r_q , a distância de s_q até f_1 e f_2 é representada, respectivamente, por d_1 e d_2 . Como pode ser observado, cada elemento representativo gera um anel no espaço métrico que contorna o objeto de consulta. Dessa forma, a *mbOr* corresponde à intersecção dos anéis gerados por todos os elementos representativos, como ilustrado pela área hachurada.

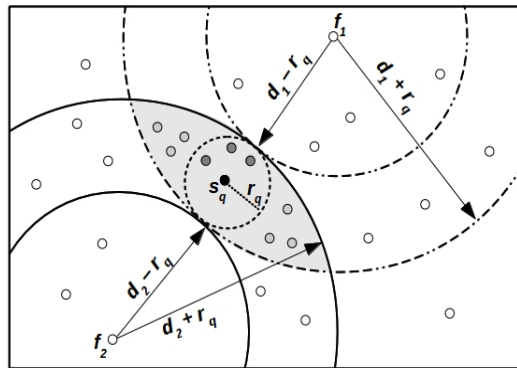


Figura 3.3: Definição de uma *mbOr* (área hachurada) para uma consulta ao elemento s_q , com raio r_q e dois elementos representativos, f_1 e f_2 . Adaptado de (Traina-Jr et al., 2007).

A etapa de refinamento consiste no cálculo efetivo das distâncias entre o elemento de consulta e os objetos contidos na *mbOr* para determinar a resposta da busca. Nota-se que a distância real entre os candidatos e o objeto de consulta é realizada apenas nessa etapa, quando a maioria dos objetos da base de dados já foi descartada (Traina-Jr et al., 2007).

Dentre os fatores que influenciam na delimitação de uma boa *mbOr* estão a quantidade de representantes globais e o posicionamento deles no espaço métrico. Assim, um conjunto de elementos representantes é considerado adequado quando ele reduz a *mbOr* o máximo possível, diminuindo a quantidade de distâncias a serem calculadas na etapa de refinamento.

Na proposta da técnica Omni, foi mostrado que o número h de representantes pode ser obtido de acordo com a dimensionalidade intrínseca do conjunto de dados, que consiste da quantidade mínima de atributos necessária para representar os objetos no conjunto. Uma maneira de encontrar uma aproximação para a dimensionalidade intrínseca é por meio da correlação da dimensão fractal D_2 (Belussi and Faloutsos, 1995; Traina-Jr et al., 2000). Ademais, foi mostrado que um valor adequado para h é o próximo inteiro que contém a dimensão intrínseca, ou seja, $\lceil D_2 \rceil + 1$.

Na Figura 3.4 é ilustrada a influência do posicionamento dos elementos representativos na eficiência da técnica Omni. No caso da Figura 3.4a, os três elementos representativos, f_1 , f_2 e f_3 , estão dispostos nas bordas do conjunto de dados, originando uma $mbOr$ muito próxima da região da consulta. Por outro lado, na Figura 3.4b os representantes globais estão muito próximos entre si, gerando uma $mbOr$ maior e, portanto, mais cálculos de distância devem ser realizados na etapa de refinamento.

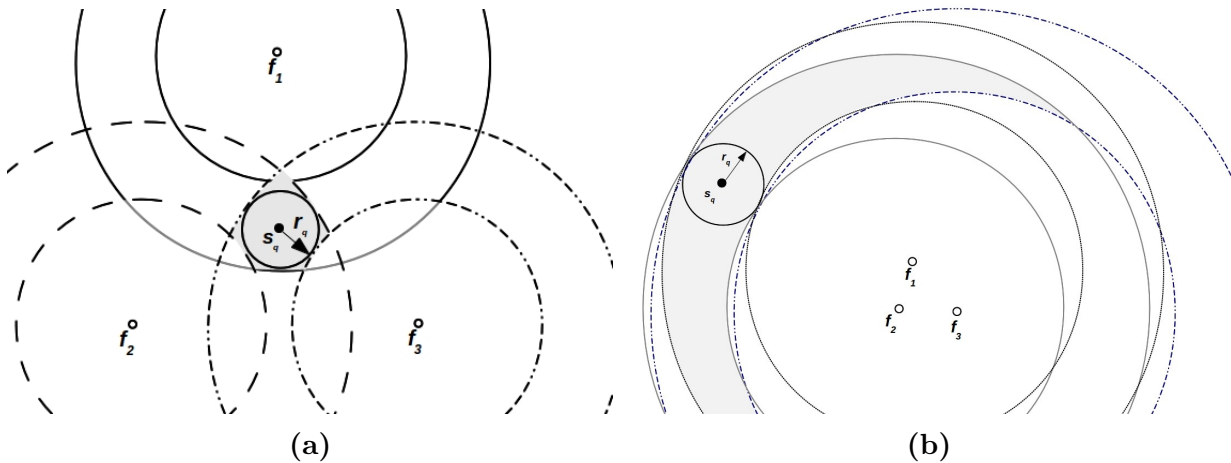


Figura 3.4: Influência do posicionamento dos representantes globais na determinação de uma $mbOr$ ótima. Na Figura 3.4a, os elementos estão dispostos nas bordas do conjunto de dados. Já na Figura 3.4b, os representantes estão muito próximos entre si. Adaptado de (Traina-Jr et al., 2007)

Como a redução máxima da $mbOr$ nem sempre pode ser atingida (Traina-Jr et al., 2007), uma heurística na determinação do melhor posicionamento dos representantes consiste em selecionar aqueles elementos do espaço que se encontram mais próximos das bordas do conjunto de dados, o que é feito pelo algoritmo *Hull of Foci* (HF) (Traina-Jr et al., 2007).

3.3 Camadas Perceptuais

Nas áreas de processamento de imagens e visão computacional são pesquisados diversos descritores que representam o conteúdo intrínseco das imagens baseados principalmente nas propriedades de cor, textura e forma (Gonzalez and Woods, 2006).

No contexto de sistemas CBIR, um dos maiores desafios é a redução do *gap* semântico em consultas por similaridade, ou seja, a diferença entre o resultado produzido pelo sistema e o que era esperado pelos usuários (Traina et al., 2009). Isso ocorre, pois diferentes usuários podem escolher condições distintas para determinar a similaridade em um