



Universidade Federal de Uberlândia

**Programa de Pós-Graduação
Faculdade de Engenharia Elétrica**

**RECUPERAÇÃO DE IMAGENS POR CONTEÚDO:
UMA ABORDAGEM MULTIDIMENSIONAL DE MODELAGEM DE
SIMILARIDADE E REALIMENTAÇÃO DE RELEVÂNCIA**

por Emílio Zorzo Barcelos

Orientadora

Prof^ª Dr^ª Edna Lúcia Flôres

Uberlândia, MG

2009

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



Universidade Federal de Uberlândia

Programa de Pós-Graduação
Faculdade de Engenharia Elétrica

Emílio Zorzo Barcelos

RECUPERAÇÃO DE IMAGENS POR CONTEÚDO: UMA ABORDAGEM MULTIDIMENSIONAL DE MODELAGEM DE SIMILARIDADE E REALIMENTAÇÃO DE RELEVÂNCIA

Dissertação de Mestrado apresentada à Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciências – Área de Concentração: Processamento Digital de Sinais.

Orientadora

Prof^a Dr^a Edna Lúcia Flôres

**Uberlândia, MG
2009**

Emílio Zorzo Barcelos

RECUPERAÇÃO DE IMAGENS POR CONTEÚDO: UMA ABORDAGEM MULTIDIMENSIONAL DE MODELAGEM DE SIMILARIDADE E REALIMENTAÇÃO DE RELEVÂNCIA

Dissertação de Mestrado apresentada à Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciências – Área de Concentração: Processamento Digital de Sinais. Aprovação em 09 de Setembro de 2009.

Banca Examinadora

Orientadora: Prof^ª Dr^ª Edna Lúcia Flôres – UFU

Membro da Banca: Prof. Dr. Sandrerley Ramos Pires – UFG

Membro da Banca: Prof. Dr. Gilberto Arantes Carrijo – UFU

Membro da Banca: Prof. Dr. Antonio C. Paschoarelli Veiga – UFU

**Uberlândia, MG
2009**

Dados Internacionais de Catalogação na Publicação (CIP)

- B242r Barcelos, Emílio Zorzo, 1979-
Recuperação de imagens por conteúdo: uma abordagem multidimensional de modelagem de similaridade e realimentação de relevância / Emílio Zorzo Barcelos. - 2009.
148 f. : il.
- Orientadora: Edna Lúcia Flôres.
Dissertação (mestrado) – Universidade Federal de Uberlândia, Programa de Pós-Graduação em Engenharia Elétrica.
Inclui bibliografia.
1. Processamento de imagens - Técnicas digitais - Teses. I. Flôres, Edna Lúcia. II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

CDU: 621.397.331

Agradecimentos

Agradeço a Deus pelo dom da vida e pela oportunidade de poder escolher, a cada dia, como evoluir como homem, pai, marido, filho, irmão e amigo optando pelo caminho do bem e buscando dar bons exemplos durante esta jornada.

Também sou fortemente grato (à Ele) pelos obstáculos que me auxiliam no crescimento, mas principalmente pelas pessoas que comigo compartilham das dores e felicidades de cada página virada.

Obrigado à minha orientadora, pelo reconhecimento, paciência e complacência.

Obrigado aos meus pais e familiares, pela motivação e carinho.

Obrigado à minha esposa pelo amor e retaguarda.

Obrigado, meu Senhor pela filhinha maravilhosa que nos deste o prazer do convívio e a linda missão da educação.



You will always miss 100% of the shots you don't take

Wayne Gretzky

Resumo

Este trabalho apresenta uma estratégia multidimensional de modelagem de similaridade e técnica de realimentação de relevância para a minimização do problema do salto semântico, intrínseco dos sistemas *CBIR*, permitindo aos usuários a customização de suas pesquisas de acordo com seus requisitos e preferências. Propõe-se uma estratégia composta, utilizando uma abordagem multidimensional, vetorial, com agrupamento espacial e ordenada por relevância. Considerando um grupo de k características que representam os elementos em um banco de dados de imagens, a medida de similaridade entre a imagem de consulta e outra da coleção de imagens é analisada em um espaço k dimensional de acordo com suas projeções sobre os eixos x_n , onde $n = 1, 2, \dots, k$. Vários testes foram realizados com o sistema proposto utilizando um banco de imagens de testes contendo até 20.000 figuras. Os resultados obtidos mostraram que a abordagem apresentada pode aprimorar substancialmente o resultado em sistemas de recuperação de imagens.

Palavras chave: *Realimentação de relevância multidimensional; modelagem de similaridade vetorial; algoritmo genético; recuperação de imagens.*

Abstract

This work presents a multi-dimensional similarity modeling strategy and relevance feedback technique for minimizing the semantic gap intrinsic problem of *CBIR* systems by allowing users to customize their queries according to their requirements and preferences. We propose a composite strategy using a multi-dimensional, vectorial, spatially clustered, and relevance-ordered approach. Given a set of k features which represents the elements in an image database, the similarity measure between a query image and another from the image collection is analyzed in k components, and the images are ranked on a k dimensional space according to their projections over the axis x_n , where $n = 1, 2, \dots k$. System experimentation was executed thoroughly using a test image database containing up to 20,000 pictures. The experimental results have shown that the presented approach can substantially improve the outcome in image retrieval systems.

Sumário

1. Introdução	1
2. Informação e Imagens Digitais.....	6
2.1 A Era da Informação.....	6
2.2 Recuperação da Informação	8
2.3 Imagens Digitais	11
2.4 Recuperação de Imagens.....	12
2.5 Considerações Finais deste Capítulo	15
3. Recuperação de Imagens por Conteúdo	16
3.1 Introdução	16
3.2 Conceitos e Fundamentos	18
3.2.1 Domínio de Consulta.....	20
3.2.2 Concepção de Pesquisa	21
3.2.3 Estratégia de Pesquisa	24
3.3 Extração de Características	26
3.3.1 Cor	30
3.3.2 Textura	35
3.3.3 Forma.....	46
3.4 Métricas de Similaridade.....	56
3.5 Realimentação de Relevância.....	59
3.6 Medidas de Desempenho	63
3.6.1 Precisão-Revocação.....	64
3.6.2 Precisão- <i>R</i>	65
3.6.3 Precisão Média para Níveis de Revocação	65
3.6.4 <i>ANMRR (Average Normalized Modified Retrieval Rank)</i>	66

3.6.5	<i>ANSMD (Average Normalized Similarity Metric Difference)</i>	67
3.7	Apresentação de Resultados	68
3.8	Exemplos de CBIR	72
3.9	Considerações Finais deste Capítulo	82
4.	Algoritmos Genéticos	83
4.1	Introdução	83
4.2	Fundamentos	84
4.3	Componentes	86
4.3.1	Cromossomos	87
4.3.2	População	88
4.3.3	Medida de Aptidão	89
4.3.4	Evolução	91
4.3.5	Seleção	92
4.3.6	Reprodução	94
4.3.7	Mutação	99
4.3.8	Reinserção	100
4.3.9	Condições de Parada	101
4.3.10	Parâmetros de Controle	101
4.4	Problemas Práticos	103
4.5	AGs e Realimentação de Relevância	104
4.6	Considerações Finais deste Capítulo	110
5.	VIPIGraf CBIR	111
5.1	Introdução	111
5.2	Estrutura Geral	111
5.3	Descritores de Características	113
5.4	Modelo de Similaridade Escalar	115
5.5	Algoritmo Genético	118

5.6	Complexidade Computacional	121
5.7	Modelo de Similaridade Vetorial.....	122
5.8	Considerações Finais deste Capítulo	125
6.	Resultados Experimentais.....	127
6.1	Introdução	127
6.2	Bancos de Imagens Empregados.....	128
6.3	Experimentos.....	133
6.3.1	Consulta Exemplo – <i>Carros</i>	135
6.3.2	Consulta Exemplo – <i>Coleção de Flores</i>	136
6.4	Conclusões	138
6.5	Comparação ao Estado da Arte	139
7.	Conclusões, Contribuições e Trabalhos Futuros.....	141
7.1	Conclusões	141
7.2	Contribuições.....	142
7.3	Trabalhos Futuros	146
	Referências Bibliográficas.....	150

Lista de Figuras

- 2.1.** Mapa de rotas da Internet formando uma grande nuvem, como ficou conhecida mundialmente.
- 3.1.** Levantamento de publicações sobre CBIR pela IEEE, ACM e Springer, de 1995 a 2005.
- 3.2.** Contextualização dos saltos sensorial e semântico: da cena à interpretação.
- 3.3.** Fluxograma de um sistema CBIR.
- 3.4.** O portal airliners.net, o maior e mais visitado site de aviação na internet.
- 3.5.** Pesquisa por associação.
- 3.6.** Pesquisa intencional, por alvo fixo.
- 3.7.** Pesquisa categorizada, por elemento representativo de uma classe.
- 3.8.** Pesquisa com um exemplo visual.
- 3.9.** Pesquisa com um grupo como exemplos visuais.
- 3.10.** Características de baixo e alto níveis.
- 3.11.** Aplicação no campo de segurança patrimonial.
- 3.12.** Extração de características resultando em vetores de características.
- 3.13.** Alguns espaços de cores.
- 3.14.** Histogramas de cores.
- 3.15.** Fenômenos físicos da reflexão e refração em uma bolha de sabão.
- 3.16.** Exemplos de texturas encontrados na natureza.
- 3.17.** Vizinhaça de Textura.
- 3.18.** Uma forma e sua assinatura, obtida pela distância ao centro.
- 3.19.** Códigos de cadeia para 4 e 8 direções e suas representações.
- 3.20.** Aproximação por polígonos.
- 3.21.** Análise sintática.
- 3.22.** Fecho convexo.
- 3.23.** Saliências.
- 3.24.** Diferentes estratégias de realimentação de relevância.
- 3.25.** Visualização de resultados ordenados por relevância.
- 3.26.** Visualização de resultados ordenados por tempo de recuperação.
- 3.27.** Visualização de resultados de forma agrupada.

- 4.1.** Ciclo de execução básico dos AGs.
 - 4.2.** Evolução de um algoritmo genético.
 - 4.3.** Exemplo de uma roleta de seleção.
 - 4.4.** *Crossover* simples.
 - 4.5.** *Crossover* múltiplo.
 - 4.6.** *Crossover* uniforme.
 - 4.7.** *Crossover* PMX.
 - 4.8.** *Crossover* cíclico.
 - 4.9.** Mutação simples.
 - 4.10.** Uma imagem dividida em nove regiões uniformes.
-
- 5.1.** Fluxograma básico de funcionamento do CBIR.
 - 5.2.** Aplicação dos operadores de Sobel.
 - 5.3.** Separação da imagem em regiões.
 - 5.4.** Fluxograma do algoritmo genético desenvolvido no VIPIGraf CBIR.
 - 5.5.** Dois cenários avaliados da mesma forma por uma função de aptidão não baseada em ordem.
 - 5.6.** Informações escalares e vetoriais.
-
- 6.1.** Banco de imagens Vistex-60.
 - 6.2.** Banco de imagens Vistex-167.
 - 6.3.** Banco de imagens Brodatz-208.
 - 6.4.** Banco de imagens Corel-1000.
 - 6.5.** Banco de imagens BD-10000.
 - 6.6.** Banco de imagens BD-20000.
 - 6.7.** Imagens de consulta para os experimentos-exemplo.
 - 6.8.** Imagem consulta e miniaturas dos primeiros resultados produzidos pelo experimento *Carros*.
 - 6.9.** Imagem consulta e miniaturas dos primeiros resultados produzidos pelo experimento *Coleção de Flores*.

Lista de Tabelas

3.1. Principais características de textura extraídas da matriz de co-ocorrência.

4.1. Precisão de recuperação dos melhores métodos de Stejić et al.: LSP-C± e WLSP-C±.

6.1. Resultados experimentais – número de imagens relevantes e precisão-R para 50 consultas teste.

6.2. Comparação entre o desempenho do modelo multidimensional e o estado da arte.

Lista de Abreviações

AG	Algoritmos Genéticos
ANMRR	<i>Average Normalized Modified Retrieval Rank</i>
ANSMD	<i>Average Normalized Similarity Metric Difference</i>
CIE	Comissão Internacional de Iluminação
CBIR	<i>Content Based Image Retrieval</i> (Recuperação de imagens baseado em conteúdo)
CMYK	Modelo de cores <i>Cyan-Magenta-Yellow-Black</i> (Ciano, Magenta, Amarelo e Preto)
FE	<i>Feature Extraction</i> (Extração de Características)
FF	<i>Fitness Function</i> (Função de Aptidão)
HSV	Modelo de cores <i>Hue-Saturation-Value</i> (Matiz, Saturação e Valor)
HSI	Modelo de cores <i>Hue-Saturation-Intensity</i> (Matiz, Saturação e Intensidade)
IR	<i>Information Retrieval</i> (Recuperação da Informação)
LSP	<i>Local Similarity Pattern</i>
NIST	<i>National Institute of Standards and Technology</i>
PMX	<i>Partially Matched Crossover</i>
PWT	<i>Pyramid-structured Wavelet Transform</i>
PX	Pixel, <i>Picture Element</i>
Re	<i>Recall</i> (Revocação)
RF	<i>Relevance Feedback</i> (Realimentação de relevância)
RNA	Redes Neurais Artificiais
Pr	<i>Precision</i> (Precisão)
RGB	Modelo de cores <i>Red-Green-Blue</i> (Vermelho, Verde e Azul)
SRI	Sistema de Recuperação da Informação
TBIR	<i>Text Based Image Retrieval</i> (Recuperação de imagens baseado em texto)
TREC	<i>Text Retrieval Conference</i>
TWT	<i>Tree-structured Wavelet Transform</i>
VCC	Vetor de Coerência de Cores
WLSP	<i>Weighted Local Similarity Pattern</i>
WWW	<i>World Wide Web</i> (Rede mundial de computadores, A Internet)

INTRODUÇÃO

Presencia-se diariamente o crescente número de *websites*, portais, serviços, aplicações e sistemas computacionais disponíveis e conectados à Nuvem, esta que já fora chamada de “a grande rede” e popularmente conhecida por Internet. Este fenômeno ocorre em tão larga escala devido à audaz aspiração humana por informação, por comunicação. Contudo, a busca por informação relevante em meio a um cenário repleto de elementos dos mais variados temas – de histórias em quadrinhos a planejamentos e estratégias empresariais, de filosofia e auto-ajuda a tutoriais sobre práticas esportivas, dentre tantos outros – pode se tornar uma tarefa complexa e delongada. Para a consecução de tal desafio, pesquisas e sistemas de recuperação de informação são desenvolvidos e implementados em número proporcional ao crescimento da Nuvem [1].

A informação pode ser representada e difundida de diversas maneiras – seja por textos, a partir de símbolos gráficos e caracteres, ou pela fala, por meio de sinais sonoros, como também por imagens, figuras e ilustrações. A utilização de mais de uma forma de representação da informação, como por exemplo, em vídeos, onde se tem imagens, sons e também textos, é comumente denominada multimídia e se tornou bastante popular devido ao aumento da capacidade de transmissão de dados pela Nuvem.

Os sistemas de recuperação de informações mais utilizados atualmente, como o *Google* e *Yahoo!*, se baseiam na busca textual, utilizando a similaridade entre palavras,

sentenças e contextos literários com a finalidade de retornar ao usuário uma lista ordenada de conteúdos relevantes. Desta maneira, caso um paciente com cólica nos rins deseje se informar sobre casos semelhantes ao seu, diagnósticos e possíveis complicações, basta ele acessar um desses sistemas de busca, inserir o que procura saber (por exemplo, “pedra nos rins”) e executar a pesquisa, navegando pelos resultados até encontrar informações que lhe interessem.

No caso de sistemas de recuperação de imagens, o desafio é de maior complexidade. O ditado popular “uma imagem vale por mil palavras” provavelmente quer dizer que pode-se descrever uma imagem a partir de seus componentes, das cores predominantes, das emoções que um indivíduo sente ao vê-la, dos lugares aos quais ela remete, das formas presentes e por intermédio de características e qualidades que dependem da criatividade e experiência de cada indivíduo que a visualiza, sendo que dificilmente uma imagem será descrita da mesma maneira, com as mesmas palavras, mesmo pelo próprio indivíduo.

Por isso, os primeiros sistemas de busca de imagens não obtiveram grande êxito [2] na recuperação de figuras relevantes, pois utilizavam uma técnica de pesquisa textual conhecida na literatura como *Text-Based Image Retrieval* que associa determinada imagem a palavras-chave. Esta estratégia normalmente resulta em um grande número de falsos positivos e obtém resultados de precisão muito inferiores quando comparados à busca simples por informação textual, devido à complexidade de descrição da imagem em palavras, fazendo com que o usuário continue tendo o trabalho de navegar pelas imagens resultantes por período de tempo imprevisível com o objetivo de encontrar imagens que lhe interessem ou aquela exata figura a qual procura.

Por mais que esses sistemas evoluam, sempre enfrentarão o fato de que cada indivíduo é único e observa o mundo de maneira distinta, o que torna sistemas de recuperação de imagens baseados em texto pouco prováveis de atingirem melhores resultados do que

sistemas que utilizam a estratégia descrita na literatura como *Content-Based Image Retrieval (CBIR)*, que se baseia em características do conteúdo da imagem, como cores, formas e texturas, para a realização da pesquisa.

Os sistemas CBIR de pesquisa por imagens a partir de seus conteúdos geralmente funcionam da seguinte maneira: primeiramente, as características de cada imagem em um banco de imagens são extraídas, representadas em vetores e registradas em um banco de dados de características. Um aluno que precise encontrar uma série de imagens de carros para a feira de ciências de seu colégio mas só possui uma foto do carro de seu pai, acessa o sistema e ao invés de digitar a palavra “carro” como em sistemas baseados em texto, insere essa imagem de pesquisa contendo um automóvel. As mesmas características são extraídas dessa imagem de pesquisa gerando vetores de características que são comparados um a um aos registros no banco de dados e listados ordenadamente de acordo com sua similaridade. Finalmente, a lista de resultados é apresentada ao usuário, contendo imagens semelhantes, não pelo texto “carro”, mas pelas características de baixo nível contidas na imagem de pesquisa. Portanto, quanto melhores selecionados os descritores de características das imagens por seus projetistas, maior precisão e eficiência o sistema de recuperação de imagens terá.

No entanto, existe um salto semântico entre as características de baixo nível extraídas da imagem, como cor, forma e textura, do conteúdo de alto nível da mesma. Visando contornar este viés e minimizar o efeito da subjetividade inerente ao problema, utiliza-se a técnica de realimentação de relevância descrita na literatura como *Relevance Feedback*, que consiste em permitir que o usuário selecione imagens dentre a lista de resultados da pesquisa anterior que melhor se relacionam com seus objetivos e interesses. Normalmente, a realimentação é efetuada a partir da simples seleção das imagens relevantes, ou mesmo com a seleção de pesos positivos e negativos, fazendo com que imagens semelhantes sejam

posicionadas em uma nova lista de resultados de acordo com as intenções e interpretações do usuário.

Considerando o *feedback* de relevância, algoritmos de inteligência artificial e técnicas de computação evolutiva como algoritmos genéticos e redes neurais são aplicados para determinar um melhor modelo de similaridade, reduzindo o salto semântico e visando o atendimento das requisições dos usuários com maior precisão e qualidade.

Os Algoritmos Genéticos (AG) há tempos são vistos como útil ferramenta para a recuperação da informação graças ao seu poder de adaptação, podendo ser aplicados na adequação de mecanismos de consulta conforme critérios e designações dos usuários.

A apresentação das imagens resultantes da pesquisa é uma estratégia importante na popularização do próprio sistema CBIR, pois os usuários não somente desejam encontrar imagens relevantes, mas também que a pesquisa seja realizada o mais breve possível, oferecendo resultados de forma simples e ágil, com interface clara e intuitiva.

Diversas abordagens na apresentação dos resultados são sugeridas. A abordagem convencional lista as imagens por ordem de relevância, apresentando pequenas amostras das imagens, organizadas em páginas. Uma abordagem bidimensional ou tridimensional esférica também é utilizada em alguns sistemas representando as imagens da coleção próximas entre si, normalmente de acordo com sua similaridade por cores.

Neste trabalho, não somente verifica-se um aprimoramento na utilização do sistema CBIR pelos usuários, mas também uma melhora considerável na precisão dos resultados quando aplicamos uma nova abordagem multidimensional, sugerida.

O principal objetivo deste trabalho é elucidar os conceitos e metodologias aplicados a um sistema de recuperação de imagens baseado em conteúdo com realimentação de relevância utilizando um algoritmo genético e uma estratégia multidimensional para a

apresentação de resultados, delineando as idéias, boas práticas e tendências neste campo da ciência.

Na intenção de expor com clareza os conceitos, as técnicas e os resultados obtidos neste trabalho, o mesmo foi estruturado em sete capítulos, da seguinte maneira – o Capítulo 1 realiza uma breve introdução sobre os sistemas de recuperação de imagens por conteúdo, ambientando o leitor ao tema; o Capítulo 2 aborda a informação e as imagens no contexto da era digital, introduzindo conceitos sobre recuperação da informação e suas principais dificuldades; Os fundamentos e estratégias envolvidos no estudo, projeto e desenvolvimento de um sistema CBIR são ilustrados em detalhe no Capítulo 3, bem como a apresentação de alguns exemplos de implementações reais desses sistemas; o Capítulo 4 trata dos Algoritmos Genéticos, abordando sua inspiração biológica, conceitos fundamentais, seus principais componentes, problemas práticos e sua utilidade na criação de mecanismos de realimentação de relevância; o estudo detalhado da implementação do *VIPIGraf CBIR* e a apresentação das contribuições deste trabalho com a abordagem multidimensional de modelagem de similaridade, visualização e estratégia de *feedback* de relevância são realizados no Capítulo 5; o Capítulo 6 apresenta os resultados obtidos experimentalmente apontando algumas observações e diferenciais da aplicação da abordagem proposta; as conclusões são elaboradas no Capítulo 7, que também delinea as perspectivas e trabalhos futuros advindos da consecução deste trabalho.

INFORMAÇÃO E IMAGENS DIGITAIS

2.1 A Era da Informação

Os dias atuais são marcados pela habilidade dos indivíduos de transferirem informação livremente, tendo acesso ao conhecimento que em outras ocasiões seria difícil ou impossível de encontrar.

A informação transpõe cidades, países, continentes e até planetas. O pronto acesso à informação gera um movimento democrático onde as oportunidades são abertas a todos, as crises, compartilhadas entre todos, e a evolução do conhecimento otimizada. A era da informação não só traz a globalização a um patamar fidedigno, mas torna a vida das pessoas conectada de maneira que nos tornamos uma grande, complexa e diversificada rede.

A Internet foi proposta originalmente para conectar computadores em rede de maneira distribuída e auto-protégida contra falhas, pois mesmo que nós importantes de sua estrutura sejam afetados, a informação é facilmente roteada por um novo e eficiente caminho. A invenção da rede mundial de computadores (*World Wide Web*), sugerida por Tim Berners-Lee em 1989 [3], tornou a Internet uma conexão global e conseqüentes desenvolvimentos e aplicações a tornou popular como é nos dias de hoje.

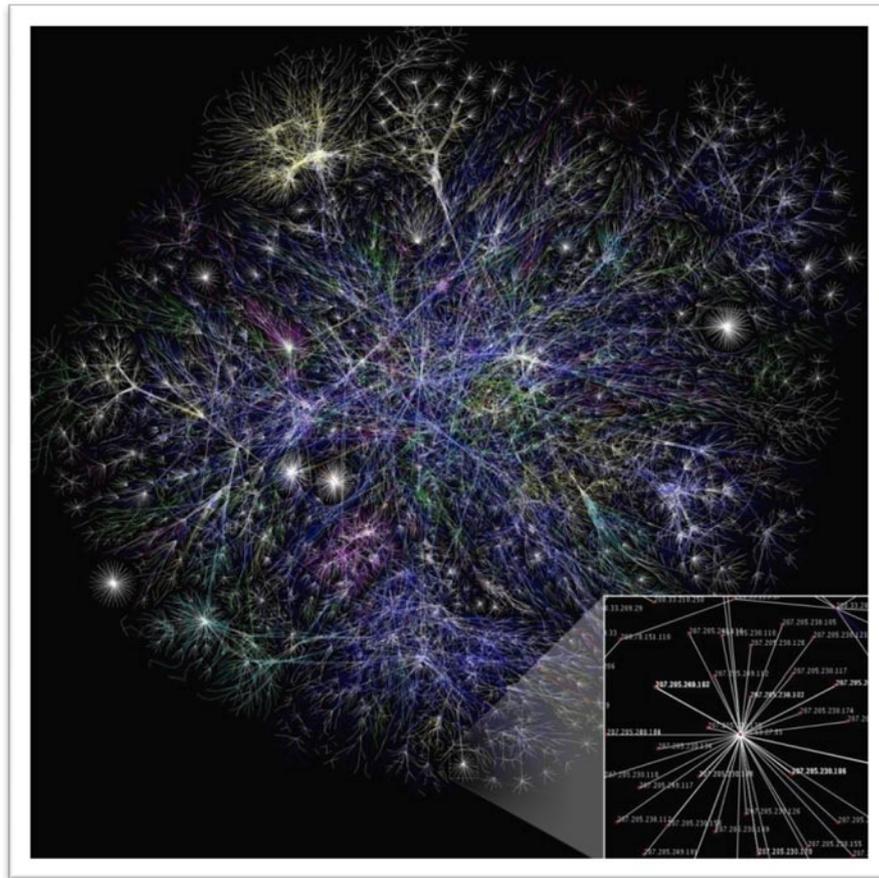


Figura 2.1. Mapa de rotas da Internet formando uma grande nuvem, como ficou conhecida mundialmente.

A motivação humana por organizar coisas é inerente [4]. Por séculos, textos de diferentes conteúdos e idiomas foram ordenados para a rápida recuperação, seja manualmente, como nas antigas *Bibliotheken*, ou automaticamente, com a ajuda de sistemas computacionais, em bibliotecas e acervos eletrônicos atuais.

Em meio a esse emaranhado de documentos e arquivos de dados, imagens, sons e vídeos, sistemas de organização, e principalmente, de recuperação da informação se fazem fundamentais. De que adianta possuir um grande acervo, se o tempo para acessar certo documento se delonga de maneira insatisfatória? [4] A informação tem que estar presente, mas também tem de ser ágil, de pronto acesso.

Por isso, novas técnicas, abordagens e algoritmos são desenvolvidos pela crescente comunidade científica desta área do conhecimento [2], com a finalidade de satisfazer a audaz aspiração humana por informação, por comunicação.

2.2 Recuperação da Informação

Estudos realizados na década de 90 [5] apontam que a forma preferida de obtenção de informações da maioria das pessoas era por meio do contato com outras pessoas. Entretanto, conforme os sistemas de recuperação de informação foram otimizados e sua efetividade aprimorada a níveis de qualidade satisfatórios para os seus usuários, a busca por informações na internet se tornou um padrão e geralmente, a fonte preferida de informação. Isso fez com que a recuperação da informação deixasse de ser uma disciplina primariamente acadêmica para ser o fundamento para o acesso à informação de milhões de pessoas, a todo o momento.

O significado do termo *Recuperação da Informação* pode ser bastante amplo. O ato de abrir uma agenda para verificar os aniversariantes do mês pode ser uma forma de recuperação de informação. Entretanto, cientificamente, recuperação da informação pode ser definida como o processo de encontrar documentos sem uma estruturação definida, pertencentes a grandes coleções que satisfaçam uma necessidade de informação [6]. Há algum tempo atrás, essa área da ciência era de interesse de bibliotecários, arquivologistas, advogados e pesquisadores, mas atualmente, é material de estudo para um número crescente de pessoas e organizações.

Devido à sua característica multidisciplinar, sistemas de recuperação da informação, ou comumente denominados *SRI*s, são normalmente desenvolvidos por profissionais de diversas áreas como engenharia, ciência da computação, matemática, arquitetura da

informação, ciência da informação, psicologia cognitiva, lingüística, estatística e física, entre outros.

Historicamente, em 1992, o Departamento de Defesa dos Estados Unidos da América, juntamente com o Instituto Nacional de Padrões e Tecnologia (*NIST*) realizaram o *Text Retrieval Conference, TREC* (<http://trec.nist.gov>), a primeira grande conferência do segmento. Daí em diante, a introdução de motores de busca na internet ampliou a necessidade de sistemas de recuperação em escalas grandiosas e ao final dos anos 90, esses se tornaram os mais comuns e provavelmente os melhores exemplares de modelagem, pesquisa e implementação de recuperação da informação.

Um processo de recuperação da informação inicia quando um usuário entra com uma *consulta* ao sistema. Consultas são formalizações sobre necessidades de informações e não identificam unicamente um documento em uma coleção, mas um conjunto deles compatíveis à pesquisa, provavelmente com diferentes graus de relevância [7]. Dependendo da aplicação, esses objetos podem ser documentos de texto, áudio, imagens ou vídeos e geralmente, esses arquivos não são diretamente armazenados no sistema de recuperação da informação, mas sim representados por conjuntos de índices e características, preservando-se também sua localização na rede.

A recuperação da informação se diferencia da recuperação de dados, pois no primeiro, tem-se um processo de indução, probabilístico, que busca documentos relevantes, e no segundo, um processo de dedução, determinístico, em que o resultado é composto por documentos similares.

A qualidade e efetividade desses sistemas são normalmente avaliadas por meio de duas métricas: Precisão (*Precision*) e Revocação (*Recall*) [8]. A precisão mede o percentual de documentos relevantes à necessidade de informação do usuário em relação ao total de

documentos resultantes da consulta. Já a revocação representa a fração de documentos relevantes resultantes da consulta em relação ao total de relevantes da coleção.

Um exemplo clássico de recuperação de informação seria encontrar peças da extensa obra de William Shakespeare que contenham a presença de *Brutus* e *César*, mas não *Calpurnia*. Uma maneira de realizar tal tarefa seria ler todo o conteúdo, do início ao fim, anotando a participação ou não de cada um dos personagens procurados em cada peça. A forma mais simples de recuperação assistida por computador seria a realização de uma leitura pelos documentos. Este processo é conhecido como *grepping* devido ao comando *grep* do sistema operacional Unix, que desempenha esta tarefa.

O *grepping* é altamente efetivo, especialmente devido à velocidade de processamento dos computadores mais modernos, porém, mesmo as melhores máquinas teriam problemas para consultar uma coleção de pouco menos que um milhão de palavras, como é o caso da obra de Shakespeare; imagine então o funcionamento desse procedimento em gigantescas coleções de documentos digitais.

Portanto, um sistema de recuperação de informação deve conter algumas características importantes para que além de efetivo, seja eficiente. São elas,

- a habilidade de processar rapidamente grandes coleções de documentos;
- permitir operações de consulta flexíveis; e
- a apresentação de resultados ordenados em um *ranking*.

Um grande problema em sistemas de recuperação da informação está na dificuldade humana, observada em várias ocasiões e situações, de expressar seus desejos com precisão por meio de palavras. Niels Henrik David Bohr [9], com notada ironia, diz para um indivíduo “nunca se expressar mais claramente do que é capaz de raciocinar”.

Tome como exemplo o desejo de encontrar um perfeito retrato em uma coleção. Qualquer tentativa de expressar o que faz um retrato “perfeito” acaba por subestimar a beleza da imaginação.

Quando a necessidade é a organização de imagens, as pessoas têm tradicionalmente superado as máquinas para a maioria das tarefas. Naturalmente, a interpretação do que se vê é de difícil caracterização, e ainda de maior complexidade para ensinar a um computador. Contudo, a partir da última década, tentativas ambiciosas foram, e continuam sendo desenvolvidas para fazer computadores entenderem, indexarem e caracterizarem imagens, com notável progresso [10].

2.3 Imagens Digitais

Em diversas áreas como o comércio, a academia, o governo, hospitais e centros de saúde, grandes coleções de imagens digitais são criadas. Em todo o mundo e a cada momento, estima-se que inúmeras imagens digitais são geradas pelas pessoas. Muitas dessas imagens e coleções são produtos da digitalização de coleções existentes, na forma analógica, de fotografias, diagramas, desenhos, rascunhos, pinturas e impressões. Outras são o resultado da utilização de sistemas e aplicações computacionais para a criação de imagens, já no formato digital. Imagens são produzidas por satélites, operações de vigilância, imageamentos biométricos, experimentos científicos e sistemas médicos, de entretenimento e informação, entre outros.

Uma imagem digital é uma representação bidimensional em *bits* de uma imagem e é composta por um número finito de valores digitais denominados *pixels*, nome originado da língua inglesa para o termo *picture element*, que representa um pequeno ponto, ou o menor

elemento, desta imagem. Portanto, uma imagem digital possui um número fixo de linhas (m) e colunas (n) de pixels, como uma matriz repleta de zeros e uns. Multiplicando-se os valores de m por n tem-se o total de pixels dessa imagem digital, métrica a qual denomina-se *resolução*. Comumente, as imagens digitais são transmitidas e arquivadas de forma comprimida, sendo o *JPEG* a mais popular.

A princípio, a maneira de pesquisar essas coleções de imagens seria por indexação de palavras-chave e categorias ou pela simples procura, imagem a imagem. No primeiro caso, o custo para implementar uma sistemática para categorizar todas as imagens das coleções e manter novas inserções atualizadas pode se tornar exorbitante, inviabilizando sua aplicação [11]. Contudo, o dispendioso trabalho de procurar por imagens em uma coleção uma a uma é uma opção inaceitável considerando os avanços tecnológicos e o estado atual da ciência.

Essas dificuldades se tornaram grande oportunidade para uma nova concentração de estudos nas áreas de processamento digital de imagens e afins denominada Recuperação de Imagens.

2.4 Recuperação de Imagens

Sistemas de recuperação de imagens são sistemas de recuperação de informação cujo principal objetivo é de organizar arquivos de imagens digitais. Dois principais paradigmas são aplicados na implementação de tais sistemas, o primeiro, mais antigo, que se baseia em informações textuais, meta-dados e anotações para a consecução de sua tarefa, enquanto o segundo, que em contrapartida, utiliza informações visuais para recuperar imagens que atendam às necessidades de informação das pessoas.

Tais sistemas têm aplicações educacionais, culturais, de segurança, de entretenimento, comerciais, de proteção a direitos e propriedade intelectual, entre outros. Esse é um fator essencial que motiva inúmeras pesquisas e desenvolvimentos nas áreas de recuperação de imagens, visão computacional, reconhecimento de padrões e análise de imagens. Estas recentes áreas de pesquisa anunciam aplicações de cunho tecnológico com ganhos sociais imensuráveis. Como exemplo, podem ser citados sistemas para auxílio ao diagnóstico médico e reconhecimento de tumores e lesões, prevenção ao crime por meio de reconhecimento de faces e biometria, aplicações militares para o reconhecimento de alvos, observação espacial e controle sobre áreas de desmatamentos, queimadas e de culturas não permitidas, etc.

A recuperação de imagens tem sido uma área de pesquisa bastante ativa desde a década de 70, com linhas provenientes de duas grandes comunidades científicas – administração de bancos de dados e visão computacional. Estas vertentes estudam a recuperação de imagens de diferentes ângulos, sendo uma baseada em texto, e a outra baseada em conteúdo.

Sistemas de recuperação de imagens baseadas em texto, conhecidos na literatura como *Text-based Image Retrieval* (TBIR), foram os primeiros representantes dessa modalidade de sistemas de recuperação da informação devido aos desenvolvimentos na área de consultas textuais, utilizando a similaridade entre palavras, sentenças e contextos literários com a finalidade de retornar ao usuário uma lista ordenada de conteúdos relevantes.

Apesar de tratar-se de imagens e não textos, a aplicação dessa metodologia resultou em alguns casos de sucesso. Revisões de trabalhos de recuperação de imagens baseados em anotações textuais podem ser encontrados em Chang and Hsu [12] e H. Tamura [13].

Apesar do seu improvável sucesso para muitas aplicações, essa abordagem permanece aplicável para imagens de significado semântico especial como coleções de fotos de museus e de pinturas famosas.

No entanto, existem duas principais dificuldades, principalmente quando trata-se de bancos de imagens de grande porte, na ordem de dezenas ou centenas de milhares de componentes. Primeiramente, o vasto montante de trabalho requerido para a anotação manual de categorias, palavras-chave e meta-dados não pode ser desconsiderado. A outra dificuldade, de caráter essencial, se encontra no rico conteúdo das imagens e a subjetividade da percepção humana, o que pode causar o insucesso de tais sistemas devido à imprecisão das anotações.

De maneira análoga, o serviço de recuperação de imagens provido por motores de busca como o Google utiliza os nomes, palavras e contextos ao redor de imagens encontradas na Internet para descrevê-las permitindo uma busca textual com anotação automatizada. Mesmo assim, indícios de ineficiência são destacados nesta abordagem, primordialmente pelo fato de que em várias ocorrências, textos próximos a uma imagem não a descrevem com precisão, muito menos com fidelidade.

Para superar estas dificuldades, foi proposta a recuperação de imagens baseada em conteúdo, reconhecida na literatura como *Content-Based Image Retrieval*, ou CBIR.

Este método, ao invés de anotações textuais para a caracterização das imagens, utiliza de informações de baixo nível e características como cor, forma e textura para a indexação automática de cada imagem em uma coleção. Inúmeras aplicações, científicas e comerciais, já foram desenvolvidas sob esta perspectiva e seus resultados demonstram grande superioridade frente aos sistemas de recuperação de imagens baseados em texto [11].

Contudo, mesmo com o enfoque da mídia sobre o assunto [14] e o crescente número de publicações, o desafio de implementar e trazer um sistema de recuperação de imagens baseado em conteúdo ao público comum da forma como a tecnologia de pesquisa textual o fez ainda está para ser alcançado.

Um dos principais obstáculos neste campo de estudo é o salto semântico entre as características de baixo nível escolhidas para a representação das imagens e conceitos de alto

nível, também tratados como a leitura visual ou percepção de certo indivíduo frente determinada imagem. Esta dificuldade une cientistas de áreas como visão computacional, inteligência artificial, recuperação da informação, interação e interface homem-máquina, sistemas de bancos de dados, mineração de dados, teoria da informação, estatística e psicologia, visando uma maior compreensão sobre o problema e aprimoramento conseqüente dos resultados.

Em um sistema CBIR, imagens são comparadas com base na similaridade entre a requisição do usuário (*query*) e as características extraídas da coleção de imagens.

2.5 Considerações Finais deste Capítulo

Este capítulo abordou a informação e as imagens no contexto da era digital, a era da informação, introduzindo conceitos sobre recuperação da informação e suas principais dificuldades nos dias de hoje.

Além disso, o capítulo apresentou a crescente utilização de imagens digitais e acervos gigantescos de imagens, tornando necessária a implementação de sistemas de organização e recuperação de imagens, estes, um tanto mais complexos do que *SRI*s textuais convencionais.

O capítulo seguinte trata os sistemas de recuperação de imagens baseados no conteúdo visual com maior especificidade, delineando seus conceitos e fundamentos, metodologias para extração de características, métricas de similaridade e desempenho, bem como estratégias de realimentação de relevância e visualização para o atingimento de maiores níveis de qualidade desses sistemas.

RECUPERAÇÃO DE IMAGENS POR CONTEÚDO

3.1 Introdução

O compartilhamento online de imagens se tornou extremamente popular com serviços como o Flickr [15], que hospeda centenas de milhões de imagens com conteúdo diverso. Também crescente é o interesse pela área e o número de publicações sobre recuperação de imagens [11], como mostra a Figura 3.1 a seguir.

Os anos de 1994 a 2000 podem ser caracterizados como a fase inicial da pesquisa e desenvolvimento de recuperação de imagens por conteúdo. O progresso realizado durante esse período é claramente apresentado em Smeulders et al. [16], que teve grande influência sobre os desenvolvimentos na área e continua influenciando trabalhos futuros.

Como abordado anteriormente neste trabalho, a dificuldade e o custo de se produzir anotações textuais ricas e confiáveis para grandes coleções de imagens, bem como a subjetividade associada a essas anotações, reafirmam o grande interesse pela recuperação de imagens pelo conteúdo visual nos dias de hoje.

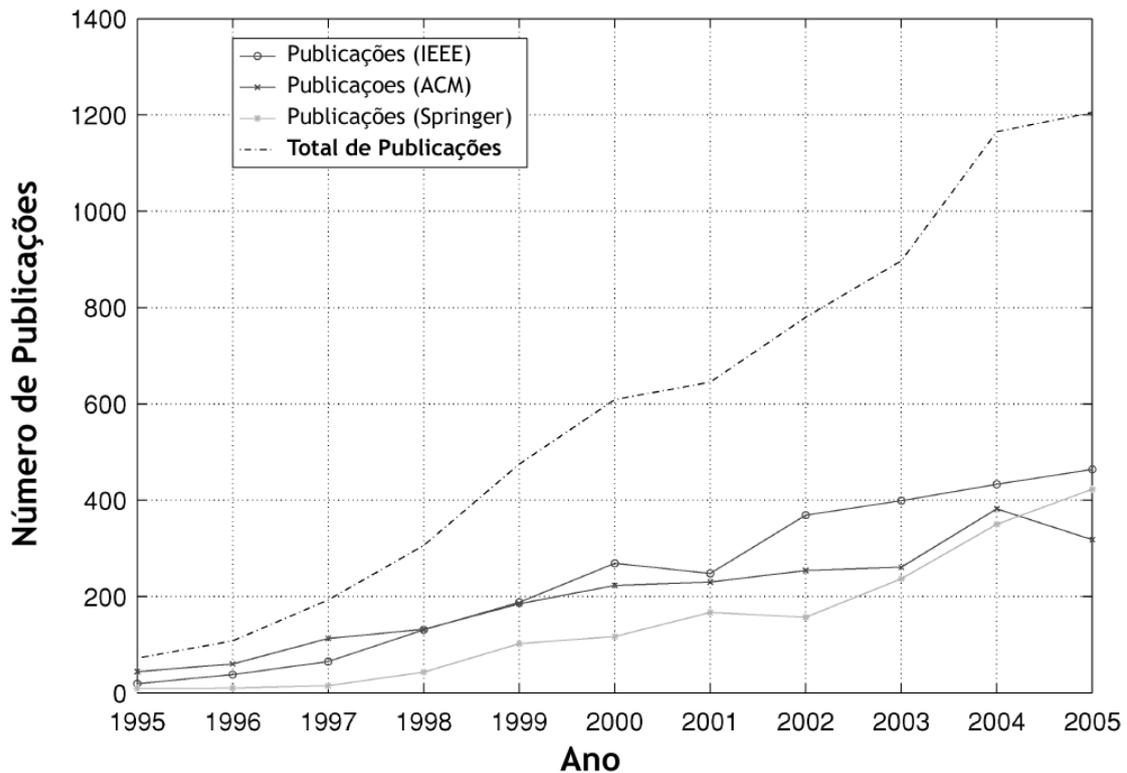


Figura 3.1. Levantamento de publicações sobre CBIR pela IEEE, ACM e Springer, de 1995 a 2005.

Uma deficiência em todas as abordagens atuais, segundo Datta et al. [11], é a credibilidade depositada à similaridade visual no julgamento da similaridade semântica. Isto ocorre devido aos problemas dos saltos sensorial e semântico.

O **salto sensorial** representa a disparidade entre o objeto no mundo real e a informação descrita (computacionalmente), derivada pelo registro da referida cena [11].

O **salto semântico** advém da falta de coincidência entre a representação matemática extraída dos dados visuais de uma imagem e sua interpretação por um usuário em uma determinada situação [11]. Os *gaps* sensorial e semântico são ilustrados na Figura 3.2.

Na seqüência deste capítulo, conceitos fundamentais são apresentados, bem como os descritores de características mais utilizados na literatura. Além disso, o capítulo trata as medidas de similaridade e cálculos de distância entre vetores, a abordagem da realimentação de relevância, estratégias de visualização e apresentação de resultados, finalizando com as

medidas de desempenho e a apresentação de alguns exemplos práticos de implementações de sistemas CBIR.



Figura 3.2. Contextualização dos saltos sensorial e semântico: da cena à interpretação.

3.2 Conceitos e Fundamentos

Vários sistemas de recuperação de imagens podem ser descritos conceitualmente como apresentado na Figura 3.3, na seqüência. O principal objetivo de um sistema CBIR é

encontrar imagens *relevantes* à necessidade do usuário, a partir de características visuais extraídas a priori. Pode-se observar que a interpretação semântica é fator essencial para a recuperação de imagens e as estratégias de caracterização, modelagem, pesquisa, similaridade, visualização e realimentação desses sistemas determinam conjuntamente sua eficiência e popularidade, representando portanto, a principal dificuldade desta área de estudo.

Considere uma coleção I composta por n imagens, k características visuais c_k são **extraídas** de cada elemento i dessa coleção resultando em índices e valores que são estruturados e construídos conforme uma estratégia de **caracterização e modelagem** dessas imagens. Essas representações das imagens são **armazenadas** em um banco de dados de características C . Então, de acordo com a metodologia do sistema CBIR, uma **pesquisa** q é efetuada ao sistema, visando a recuperação de elementos da coleção de imagens que satisfaçam a necessidade de informação do usuário. A busca gera uma ação de conferência e comparação entre as informações da pesquisa q e cada registro no banco de dados de características C , por **similaridade**, resultando na **apresentação** dos supostos documentos relevantes para a apreciação do usuário, que opcionalmente, pode **avaliar** cada imagem recuperada, tanto positiva como negativamente, **realimentando** o sistema para assim refinar sua pesquisa e conseqüentemente obter melhores e melhores resultados, a cada nova realimentação.

Geralmente as interfaces desses sistemas compõem-se de duas partes – a formulação da busca e a apresentação de resultados.

A literatura também apresenta as estratégias citadas acima caracterizadas em três classes, fundamentais para um sistema CBIR. São elas,

- Extração de características visuais;
- Indexação; e
- Projeto do sistema de recuperação.

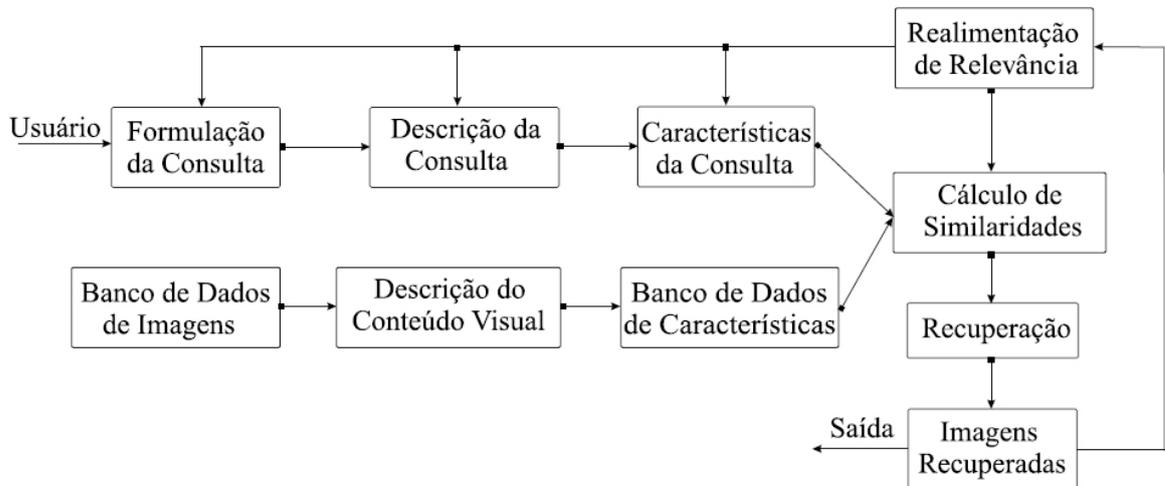


Figura 3.3. Fluxograma de um sistema CBIR.

Antes de abordar as técnicas e metodologias de um sistema de recuperação de imagens por conteúdo visual, convém classificá-los segundo algumas características que determinam a aplicação de certas estratégias em detrimento de outras, como o domínio da consulta, a concepção e a estratégia de pesquisa.

3.2.1 Domínio de Consulta

Dado a um sistema a propriedade de recuperação de imagens por conteúdo (e não por texto), uma classificação importante reconhecida na literatura aborda o domínio de consulta destas imagens, podendo um sistema CBIR ser *restrito* ou *amplo*.

Domínios de imagens **restritos** geralmente possuem variabilidade limitada e características visuais bem definidas, como por exemplo, a coleção de imagens relacionadas à aviação Airlines.Net [17].

Por outro lado, domínios **amplos** tendem a ter alta variabilidade e imprevisibilidade, como exemplo, imagens na *web*, que tornam a generalização muito mais complexa.



Figura 3.4. O portal airliners.net, o maior e mais visitado site de aviação na internet.

3.2.2 Concepção de Pesquisa

Existem três principais categorias de pesquisa em sistemas CBIR – pesquisa por associação, pesquisa intencional e pesquisa categorizada.

A primeira, denominada **pesquisa por associação**, ocorre quando não existe intenção clara por parte do usuário em encontrar determinada figura, mas pelo contrário, a pesquisa é refinada iterativamente conforme os resultados são apresentados durante sua navegação pelo sistema.

Um exemplo de aplicação dessa classe de pesquisa seria apresentar uma série de imagens de animais e conforme o usuário navega e interage com o sistema, se interessa por

tigres asiáticos albinos, e assim, ao clicar em algumas dessas figuras, um maior número de semelhantes vão sendo apresentados, como ilustra a Figura 3.5.

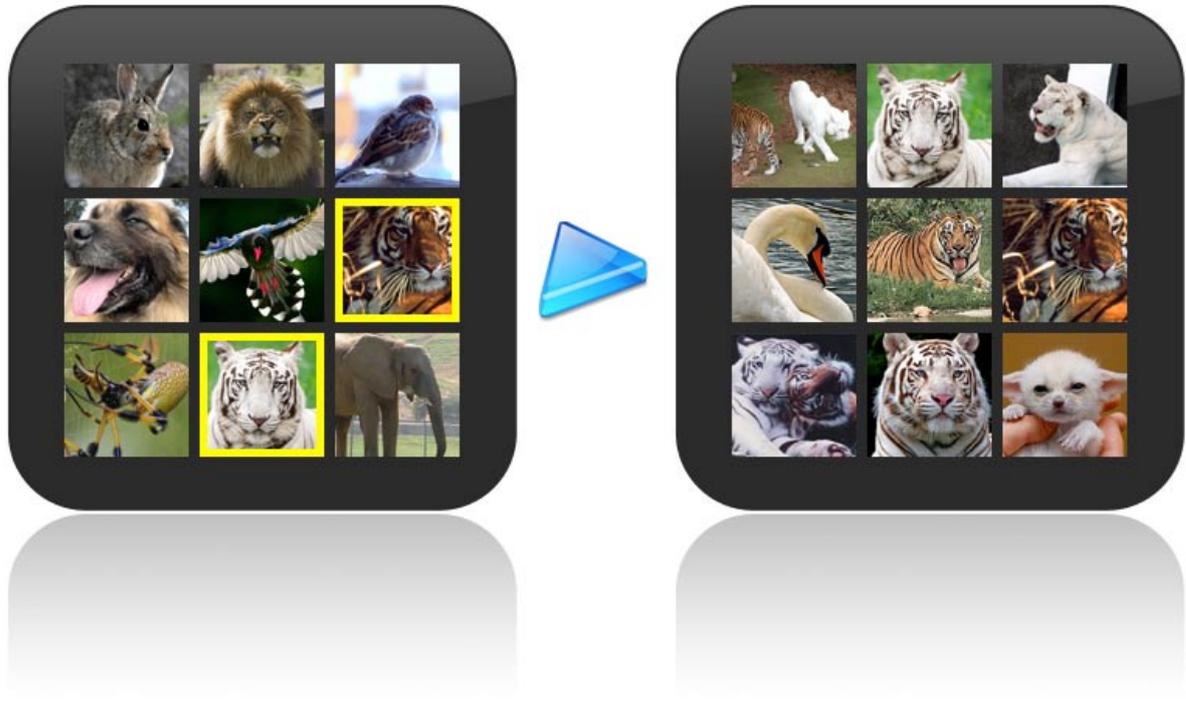


Figura 3.5. Pesquisa por associação.

Quando existe uma imagem específica a ser encontrada, classifica-se a consulta como **pesquisa intencional**. Um exemplo da aplicação dessa forma de busca é quando um usuário deseja encontrar uma imagem em alta resolução da *Monalisa*, obra de arte de Leonardo Da Vinci, para a apreciação pessoal.

A Figura 3.6 ilustra esse exemplo em um sistema pelo qual o colecionador pode desenvolver rascunhos e traços que remetem à obra de arte vislumbrada.

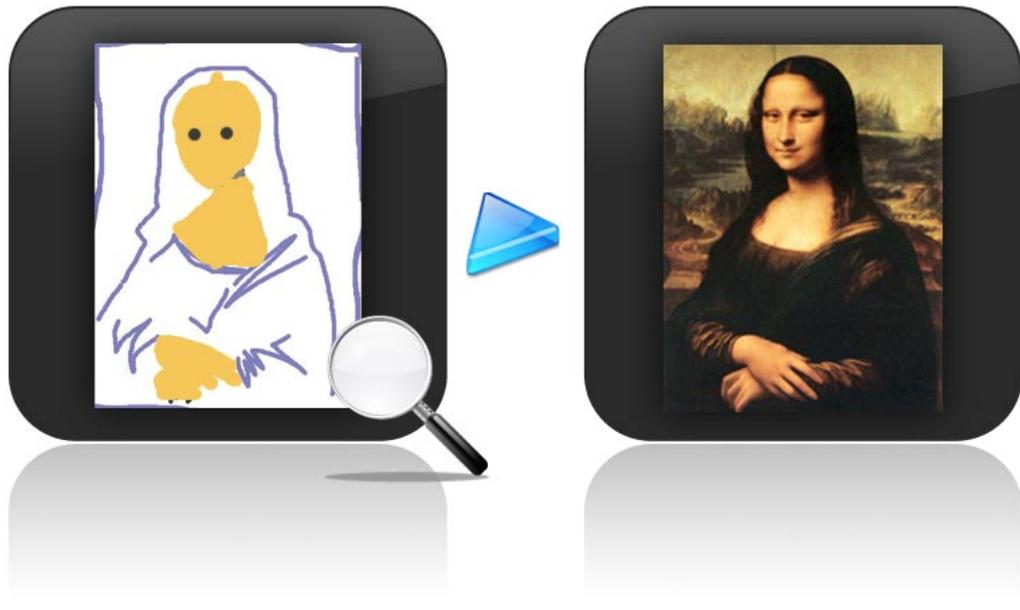


Figura 3.6. Pesquisa intencional, por alvo fixo.

A **pesquisa categorizada** é assim nomeada pela sua característica de resultar em pelo menos uma imagem que representa determinada classe semântica, para ilustrar um parágrafo ou seção de texto. Esta forma de pesquisa é utilizada, por exemplo, quando se deseja encontrar uma imagem de uma ilha para representar o cenário político nacional que endossa livremente despesas de viagens a parlamentares, como ilustração de um artigo a ser publicado.



Figura 3.7. Pesquisa categorizada, por elemento representativo de uma classe.

3.2.3 Estratégia de Pesquisa

Um sistema de recuperação de imagens é fortemente associado à sua estratégia de pesquisa. Dentre os sistemas publicados, tanto acadêmica quanto comercialmente, pode-se observar três principais modelos de estratégia de pesquisa, seja por **navegação** entre os elementos da coleção, por **busca direta** ou por **procura por exemplo visual**, este podendo ser uma imagem ou rascunho.

A navegação entre os elementos da coleção, conhecida na literatura por *browsing*, constitui da simples apresentação de exemplares do banco de imagens. O usuário seleciona as figuras que lhe são relevantes à medida que interage com o sistema.

A busca direta se baseia na informação de palavras-chave ou em termos de características extraídas da imagem, como por exemplo, cor e forma. O usuário digita as características desejadas com a finalidade de encontrar exemplares semelhantes. Por exemplo, ao pesquisar por imagens de uma Ferrari, o usuário poderá informar “quadrado” e “vermelho”. O sistema traduz os conceitos em valores de características para então realizar sua pesquisa junto ao banco de imagens.

A procura por exemplo visual funciona da seguinte forma – o usuário insere uma imagem exemplo ou desenha um esboço que melhor representa sua necessidade de informação. Esta figura é caracterizada pelos mesmos descritores de características das imagens pertencentes à coleção e é então comparada e ordenada de acordo com o modelo de similaridade aplicado. Um exemplo de aplicação dessa estratégia é a inserção de uma imagem de um foragido da justiça, digitalizada a partir de um retrato-falado, para a imediata apresentação às vítimas, de fotografias de suspeitos, já fichados, facilitando assim a segurança dos direitos dos cidadãos e aprimorando a atuação das polícias.

A procura por exemplo visual pode ser empregada utilizando a inserção de **múltiplas imagens** exemplo com o objetivo de caracterizar um grupo. Isso significa que caso o usuário

possua a fotografia de uma praia e o desenho de um cavalo e deseje encontrar uma imagem de um cavalo galopando pela praia, a inserção das imagens exemplo conjuntamente provavelmente terá maior eficiência do que um exemplar por si só.



Figura 3.8. Pesquisa com um exemplo visual.



Figura 3.9. Pesquisa com um grupo como exemplos visuais.

3.3 Extração de Características

A extração de características é a base da recuperação de imagens por conteúdo. Em processamento de imagens pode-se conceituar o termo *extração de características* como um método de captura de conteúdos visuais de imagens para indexação e pesquisa [18].

Essas características são comumente denominadas de primitivas ou *características de baixo nível* de uma imagem e retratam matematicamente aspectos visuais humanos como cor, forma e textura. Já as *características de alto nível* são aquelas que representam descrições semânticas das imagens, como objetos, ações e até sentimentos.



Figura 3.10. Características de baixo e alto níveis.

Grande parte das pesquisas nessa área do conhecimento se restringe ao estudo de características de baixo nível [19], pois a extração de características de alto nível representa uma tarefa complexa, com relativamente poucos resultados práticos.

Existem características com domínios específicos, como em rostos, impressões digitais e padrões de íris. Essas características são muito bem aplicadas em sistemas com domínio de consulta restrito, como anteriormente conceituado na subseção 3.2.1 deste capítulo. Por exemplo, um sistema CBIR de aplicação no campo da segurança patrimonial muito popular nos dias de hoje (como mostrado na Figura 3.11) se baseia na captura de uma imagem do dedo polegar ou indicador do indivíduo para acesso à áreas específicas em empresas, clubes ou organizações. Características específicas sobre impressões digitais denominadas minúcias são extraídas da imagem exemplo e então comparadas com as imagens de um banco de dados. Caso o grau de similaridade da imagem de pesquisa seja superior a um limiar de segurança, por exemplo, de 90%, o acesso é liberado ao indivíduo, desde que ele possua as credenciais para a efetivação de tal acesso. Caso contrário, a imagem pode ser registrada em outro banco de imagens de possíveis tentativas de acesso irregulares.



Figura 3.11. Aplicação no campo de segurança patrimonial.

Em sistemas com amplo domínio de consulta, geralmente utiliza-se características como cor, forma e textura para representar as imagens. A grande questão é quais são as características extraídas que desenvolverão uma recuperação mais significativa. Neste contexto, deve-se levar em consideração as *condições invariantes mínimas* que modelam a intenção do usuário, reduzindo o salto sensorial devido às distorções acidentais, cortes, oclusões, etc.

Como pré-requisito, uma característica deve:

- conter informação suficiente a respeito da imagem;
- ser de fácil computação, viabilizando o rápido acesso, principalmente em grandes coleções; e
- se relacionar bem com as características de percepção visual humanas.

Por causa da subjetividade e da complexidade da criatividade humana, não existe uma única representação de característica que seja considerada a melhor. Contudo, uma boa descrição de imagem deve ser caracterizada por alguns fatores, como a precisão de recuperação, a separabilidade, a compactabilidade, suas aplicações, complexidade computacional e robustez, entre outros [20].

A maioria dos sistemas CBIR pesquisados utilizam características de cor e textura, alguns usam características de forma e um pequeno número adota informações de posicionamento. A recuperação baseada em cor geralmente traz imagens com cores similares. Já a recuperação por textura nem sempre traz imagens pelas quais se observa claramente a mesma textura original, a menos que o banco de dados possua várias imagens com certa textura dominante. A pesquisa baseada na característica de forma às vezes traz resultados surpreendentes, mas aparentemente, os descritores de forma utilizados não demonstram ser dos mais efetivos [21].

Comumente, na literatura, utiliza-se o termo “vetor de características” (*image features array*) para uma descrição matemática de características de baixo nível como cor, forma e textura.

Como citado anteriormente neste capítulo, um vetor de características tem por finalidade representar os aspectos visuais significativos de uma imagem, como apresentado pela Figura 3.12 abaixo.



Figura 3.12. Extração de características resultando em vetores de características.

Vetores de característica podem representar a imagem como um todo, bem como regiões ou segmentos da mesma. Existem duas principais maneiras para dividir uma imagem em partes – a simples repartição em regiões com dimensões similares ou a partir de técnicas de segmentação, visando a obtenção de objetos mais semanticamente significantes (como, por exemplo, pessoas, animais, etc.)

A seguir, elucidamos os principais descritores de imagens para estudo e análise.

3.3.1 Cor

De acordo com a literatura, pode-se afirmar que **cor** é a característica mais utilizada em recuperação de imagens [11], por ser um dos principais aspectos da visão humana, sendo o histograma de cores sua representação mais comum.

A cor é percebida pelo olho quando a luz é refletida por um objeto. Em 1801, o físico Thomas Young [22] propôs o modelo tricromático para a percepção de cores em seres humanos, afirmando que todas as sensações de cor percebidas são combinações das intensidades dos estímulos recebidos pelos três tipos de cones, células fotossensíveis que compõem a retina.

Segundo Young, os valores máximos de sensibilidade se situavam próximos do vermelho, verde e azul. Seu sistema, juntamente com colaborações de outros importantes cientistas, fora adotado pela Comissão Internacional de Iluminação (CIE) em 1931, denominando as três cores como *cores primárias*.

Existem várias formas de representar as cores. Para a realização de tal tarefa, utiliza-se um espaço de cor como base para a representação. De acordo com a aplicação, certo espaço de cor pode apresentar-se mais adequado ou viável frente aos demais.

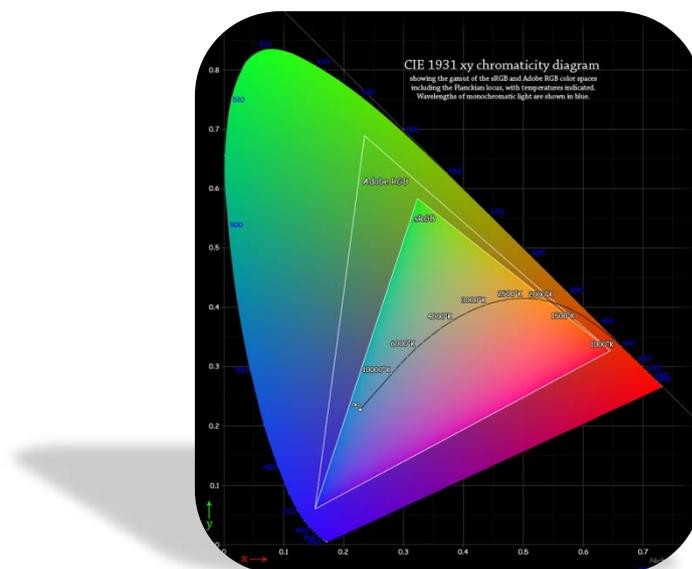


Figura 3.13. Alguns espaços de cores.

Dentre os espaços de cores, pode-se citar os modelos *RGB*, *CMYK*, *XYZ*, *YCbCr*, *HSV*, *HSL*, *HSI*, $L^*u^*v^*$ e $L^*a^*b^*$ como os mais populares. A Figura 3.13 ilustra alguns espaços de cores utilizados em processamento digital de imagens.

a. Histograma de Cores

Um histograma mede a frequência de determinado componente considerando um conjunto de elementos da mesma classe. Logo, o histograma de cores descreve a ocorrência das cores em uma imagem (Figura 3.14). Para a obtenção de melhor compactabilidade e desempenho, normalmente os histogramas apresentam a frequência de um intervalo de valores ao invés de um único valor de cor. Outras vezes, ocorre a fixação do número de cores utilizadas, seja pela maior frequência dessa cor na coleção de imagens, ou pelo seu poder de discriminação frente as demais.

Entre os problemas dos histogramas de cores pode-se destacar sua alta dimensionalidade e o fato de não considerarem a localização espacial das cores. Assim, imagens muito diferentes podem ter representações semelhantes. Caso adotada a divisão das imagens em regiões, observa-se o aumento da dimensionalidade e da carga computacional necessária para a comparação dos descritores.

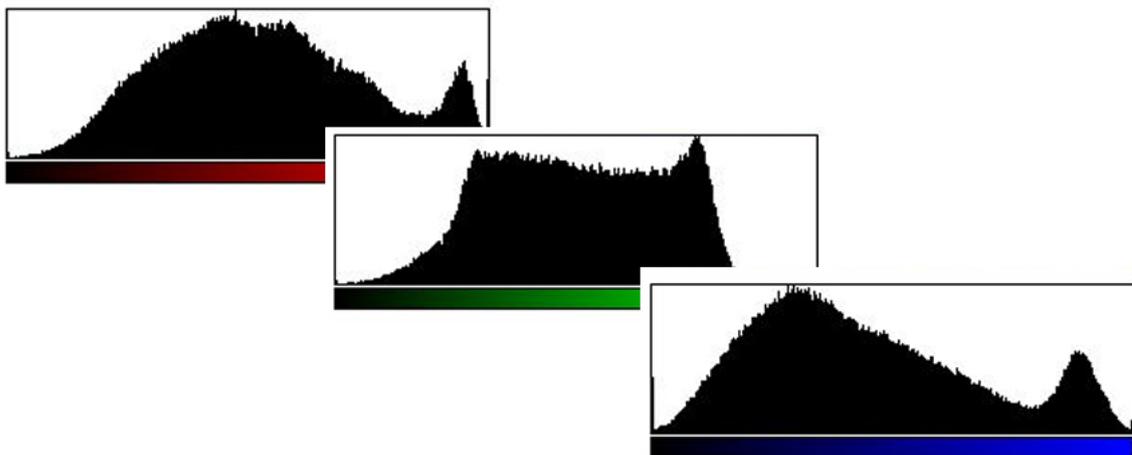


Figura 3.14. Histogramas de cores.

b. Momentos de Cores

Originalmente apresentados em 1995 por Stricker e Orengo [23], *momentos de cores* caracterizam as imagens segundo suas distribuições de cores por meio de três medidas estocásticas – média (*mean*), desvio padrão (*standard deviation*) e inclinação (*skewness*) – por isso o termo “momentos”.

Cada uma dessas medidas é computada para cada um dos canais de cores de uma imagem. Desta forma, para imagens pertencentes ao espaço de cor *RGB*, calcula-se a média, desvio padrão e inclinação para os canais *R*, *G* e *B*, respectivamente.

Devido a sua alta compactabilidade e por não considerar informações espaciais, este descritor tem poder de discriminação reduzido. Portanto, seu desempenho pode ser significativamente aprimorado quando aplicado a regiões de imagens, caracterizando localmente a região.

Seja N o número de pixels de uma imagem e p_{ij} o valor do j -ésimo pixel no i -ésimo canal de cor, os três primeiros momentos são calculados, respectivamente, a partir das Equações (3.1), (3.2) e (3.3) a seguir.

$$E_i = \frac{1}{N} \sum_{j=1}^N p_{ij} \quad (3.1)$$

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^2} \quad (3.2)$$

$$s_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^3} \quad (3.3)$$

c. Vetor de Coerência de Cores

Nesta abordagem, cada pixel de uma imagem é classificado como coerente ou incoerente. Um pixel será coerente se ele fizer parte de uma região de coloração uniforme (de tamanho pré-definido), caso contrário, ele será classificado como incoerente. Este foi o primeiro passo na tentativa de incorporar informações espaciais em um descritor de cor [24].

O número de pixels coerentes α_k e incoerentes β_k são calculados para cada cor k . O vetor de coerência de cores é formado por pares de coerência (α_k, β_k) , conforme a Equação (3.4) a seguir.

$$VCC_{Imagem} = \{ (\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_n, \beta_n) \} \quad (3.4)$$

O histograma de cores da imagem pode ser obtido pelos valores do vetor de coerência de cores seguindo a Equação (3.5) seguinte.

$$Histograma_{Imagem} = \{ \alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_n + \beta_n \} \quad (3.5)$$

Geralmente, os vetores de coerência de cores produzem melhores resultados de recuperação que os histogramas de cores, especialmente para imagens com muitas regiões de cores uniforme ou muita textura. Outro fator importante é que histogramas de cores e vetores de coerência de cores produzem melhores resultados quando aplicados ao espaço de cor *HSV* [25].

d. Correlograma de Cores

O correlograma de cores é um histograma tridimensional que caracteriza a distribuição das cores e a correlação espacial entre pares de cores. A primeira e a segunda dimensão do histograma representam as cores de qualquer par de pixels e a terceira dimensão, a distância espacial entre eles [26].

Um correlograma de cores pode ser tratado como uma tabela indexada por pares de cores, onde a k -ésima entrada para (i, j) especifica a probabilidade de um pixel de cor j se encontrar a uma distância k de outro pixel de cor i na imagem.

Seja H o conjunto de pixels de uma imagem e $H_{c(j)}$ o conjunto de pixels de cor $c(j)$, então, o correlograma desta imagem é definido como

$$\gamma_{i,j}^k = Pr[p_2 \in H_{c(j)}, |p_1 - p_2| = k] \quad (3.6)$$

onde:

$$i, j \in \{1, 2, 3, \dots, N\},$$

$$k \in \{1, 2, 3, \dots, d\},$$

$|p_1 - p_2|$ é a distância entre os pixels p_1 e p_2 , e

Pr é a função probabilidade.

Se forem consideradas as possibilidades de pares de cores para uma imagem com milhões de cores, o correlograma pode se tornar inviável devido ao seu tamanho. Neste caso, o autocorrelograma é aplicado. Este, por sua vez, captura somente a correlação espacial entre os pixels de cores idênticas reduzindo a dimensão e complexidade do descritor.

Comparado ao histograma e ao vetor de coerência de cores, o autocorrelograma de cores produz melhores resultados, porém é computacionalmente mais caro devido a sua alta dimensionalidade [27].

e. Características Invariantes de Cores

A cor percebida pelos olhos humanos é o resultado de fenômenos físicos como a reflexão da luz sobre a superfície de um determinado material (vide Figura 3.15). Contudo, outros fatores como mudanças de iluminação, as orientações das superfícies dos objetos, o

posicionamento e a angulação das câmeras, e a refração, entre outros, influenciam consideravelmente na variação das cores.



Figura 3.15. Fenômenos físicos da reflexão e refração em uma bolha de sabão.

Por isso, representações invariantes de cores foram propostas e introduzidas na recuperação de imagens por conteúdo. Gevers et al. [28] apresentam um conjunto de características invariantes de cores derivada das propriedades de reflexão de objetos.

Quando aplicadas na recuperação de imagens, essas características invariantes de cor produzem representações do conteúdo das imagens independentes da iluminação e geometria da cena, mas acabam por introduzir certa perda no poder de discriminação entre imagens.

3.3.2 Textura

Textura se refere a padrões visuais, normalmente repetitivos ou periódicos, com propriedades de homogeneidade que não são originárias de apenas uma cor ou de certa intensidade. Textura é uma característica de praticamente todas as superfícies, incluindo

nuvens, árvores, tijolos, cabelo e tecidos, entre outros. Ela contém informações importantes sobre o arranjo estrutural dessas superfícies e sua relação com o ambiente ao seu redor.

Texturas são caracterizadas pela relação entre os pixels da imagem, diferentemente de cor, que é uma característica própria de cada pixel. Elas apresentam diversas propriedades como homogeneidade, randomicidade, granularidade, aspereza, etc. e são resultantes dos princípios de combinações de cores (ou tons de cinza) que formam as imagens. Textura é uma característica fortemente discriminante do sistema visual humano e está presente em quase todo lugar na natureza, como ilustra a Figura 3.16.



Figura 3.16. Exemplos de texturas encontrados na natureza.

Devido à sua importância e utilidade em campos de estudo como Reconhecimento de Padrões e Visão Computacional, ricos resultados e pesquisas já existiam, facilitando sua aplicação à Recuperação de Imagens por Conteúdo.

Encontra-se na literatura três formas fundamentais para a descrição de texturas – estrutural, estocástica e estatística [29].

A descrição **estrutural** se baseia na concepção de que as texturas são compostas por características dispostas regularmente, segundo regras bem definidas. Telhados, estruturas celulares, colméias de abelhas e paredes de tijolos são alguns exemplos dessa descrição.

No entanto, devido à complexidade da identificação automatizada dessas características e sua aplicabilidade reduzida a texturas estritamente uniformes, a abordagem estrutural se torna, de certa forma, bastante limitada.

Métodos estruturais para a descrição de texturas compreendem grafos, operadores morfológicos e diagramas de Voronoi [30].

De maneira distinta, a descrição **estocástica** define parâmetros que governam um processo estocástico para a modelagem de texturas. A análise de textura é definida por um modelo que estima certos parâmetros de modo que o processo estocástico possa reproduzir o padrão de textura. Esses parâmetros servem como características de textura para a recuperação, classificação e segmentação dessas imagens.

Um problema observado na utilização da abordagem estocástica é que várias das texturas naturais não seguem as restrições de um modelo estocástico em particular, como cadeias randômicas de Markov, cadeias de Gibbs ou fractais. Esses métodos são bastante úteis na síntese de textura, mas nem sempre satisfatórios em processos de recuperação e classificação de imagens.

Já a abordagem **estatística** descreve as propriedades das texturas por métricas estatísticas, fortemente baseadas nos aspectos da percepção humana de textura. Dentre as

propriedades de texturas, pode-se citar *contraste, correlação, entropia, densidade, aspereza, rugosidade, uniformidade, regularidade, linearidade, direcionalidade e frequência*.

Pela semelhança à percepção humana, descrições estatísticas se tornaram as mais populares para a caracterização de texturas. A seguir, alguns desses descritores são expostos.

a. Matriz de Co-ocorrência

A matriz de co-ocorrência é a abordagem mais tradicional para a análise de textura, proposta em 1979 por Haralick [31], se fundamentando na caracterização de uma imagem de textura a partir de dados estatísticos associados à ocorrência simultânea de pixels da imagem com os mesmos níveis de cinza.

Seja a distância D , definida por coordenadas polares (d, α) , onde d é a magnitude e a orientação é $\alpha = \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$, uma matriz $P_D(i, j)$ que contabiliza a co-ocorrência de pixels com níveis de cinza i e j , pode ser definida como:

$$P_D(i, j) = Pr \{(I(p_1) = i) \wedge (I(p_2) = j) \wedge (\|p_1 - p_2\| = D)\} \quad (3.7)$$

onde:

Pr é a probabilidade,

p_1 e p_2 são as posições de dois pixels em determinada imagem e

$I(p)$ é a intensidade do pixel p .

Os valores de d são definidos de acordo com a granularidade das imagens da coleção a ser descrita.

Após calculada, um conjunto de 14 características de textura [31] podem ser extraídas da matriz de co-ocorrência. Essas características apresentam boa capacidade de discriminação de texturas, porém possuem um alto custo computacional, que deve ser levado em consideração.

As cinco características mais utilizadas em sistemas de recuperação de imagens são listadas na Tabela 3.1, onde μ_i e μ_j são as médias e σ_i e σ_j os desvios padrões de P .

Tabela 3.1. Principais características de textura extraídas da matriz de co-ocorrência.

Característica	Equação
<i>Contraste</i>	$c_1 = \sum_i \sum_j (i - j)^2 P(i, j)$
<i>Correlação</i>	$c_2 = \frac{1}{(\sigma_i \sigma_j)} \sum_i \sum_j (i - \mu_i)(j - \mu_j) P(i, j)$
<i>Energia</i>	$c_3 = \sum_i \sum_j P^2(i, j)$
<i>Entropia</i>	$c_4 = \sum_i \sum_j P(i, j) - \log P(i, j)$
<i>Homogeneidade</i>	$c_5 = \sum_i \sum_j \frac{P(i, j)}{(1 + i - j)}$

b. Vizinhança de Textura

A vizinhança de textura é calculada a partir dos tons de cinza de uma imagem, examinando a vizinhança de cada pixel (vizinhança de oito). Após análise, obtém-se uma estimativa probabilística para cada posição relativa de que um pixel vizinho tenha maior intensidade de brilho do que o pixel central [32].

A Figura 3.17 apresenta a utilização da vizinhança de textura para uma imagem exemplo, aplicada sob o espaço de cor *HSI*. Melhores informações descritivas podem ser extraídas do canal *I*.

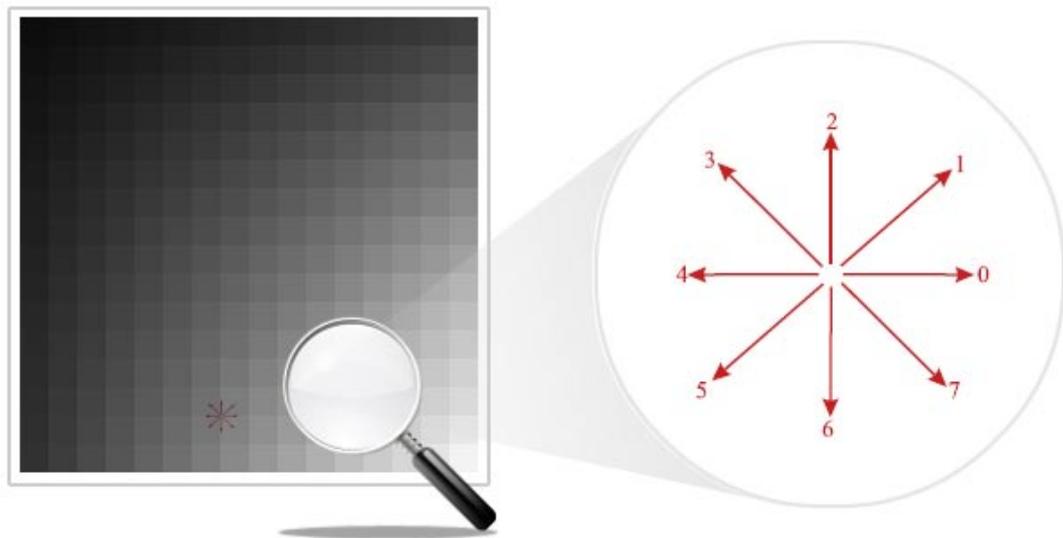


Figura 3.17. Vizinhaça de Textura.

c. Características de Tamura

Motivados por estudos psicovisuais sobre a percepção humana, Tamura et al. [33] exploraram a representação de texturas por um ângulo diferente, incluindo medidas de *rugosidade*, *contraste*, *direcionalidade*, *semelhança de linhas*, *regularidade* e *aspereza*.

A principal distinção entre a representação de Tamura e a matriz de co-ocorrência é que todas as propriedades de texturas em seu descritor são visualmente significativas, ao contrário de algumas extraídas da matriz de co-ocorrência, como “entropia”. Esta diferença torna a representação proposta por Tamura bastante atrativa para a recuperação de imagens por prover uma interface mais amigável e atributos mais compreensíveis ao público comum.

A seguir são descritas as três propriedades de Tamura consideradas mais populares – rugosidade, contraste e direcionalidade.

A **rugosidade** é uma medida da granularidade da textura. Para calcular essa característica, parte-se do cálculo das médias $A_k(x, y)$ dos níveis de cinza em cada vizinhança 2^k , $k = 2, 3, \dots, L$, para cada pixel (x, y) . Então, calcula-se a diferença entre pares de médias vizinhas não sobrepostas (h) horizontal e (v) verticalmente, como mostra as Equações abaixo:

$$E_{k,h}(x, y) = |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)| \quad (3.8)$$

$$E_{k,v}(x, y) = |A_k(x, y + 2^{k-1}) - A_k(x, y - 2^{k-1})| \quad (3.9)$$

Posteriormente, o tamanho da vizinhança $S_{ótimo}$ que maximiza o valor de E é calculado para cada pixel, a partir da Equação (3.10).

$$S_{ótimo}(x, y) = 2^k \quad (3.10)$$

Então, a rugosidade é obtida pela Equação (3.11) a partir da média dos valores $S_{ótimo}$ encontrados, onde $m \times n$ representa as dimensões da imagem.

$$Rugosidade = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{ótimo}(i, j) \quad (3.11)$$

O **contraste** caracteriza-se pela diferença entre os níveis de cinza ou de intensidade dos pixels e pode ser representado da seguinte maneira:

$$Contraste = \frac{\sigma}{\sqrt[4]{\alpha_4}} \quad (3.12)$$

onde,

o curtose é obtido por $\alpha_4 = \mu_4 / \sigma^4$,

μ_4 é o quarto momento sobre a média e

σ^2 é a variância.

O contraste pode ser calculado para toda a imagem, bem como para regiões específicas.

A **direcionalidade** é a distribuição de frequência de bordas locais orientadas em relação aos seus ângulos direcionais. O processo de cálculo da direcionalidade parte da realização de uma convolução entre a imagem e os operadores direcionais descritos na Equação (3.13).

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad e \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad (3.13)$$

Em seguida, calcula-se um vetor gradiente para cada pixel, pelo qual se extrai os valores de magnitude $|\Delta G|$ e ângulo θ , como representado nas Equações (3.14) e (3.15).

$$|\Delta G| = \frac{\Delta_H + \Delta_V}{2} \quad (3.14)$$

$$\theta = \tan^{-1} \left(\frac{\Delta_V}{\Delta_H} \right) + \pi/2 \quad (3.15)$$

onde:

Δ_H e Δ_V são as diferenças horizontais e verticais da convolução.

Um histograma de direções contendo os valores de θ é então construído para valores de magnitude superiores a um limiar T , pré-definido.

Esse histograma apresenta fortes picos para imagens com características altamente direcionais, e apresenta forma aplanada para imagens sem forte orientação. O histograma pode ser resumido por um valor único de direcionalidade. A Equação (3.16) apresenta a forma de cálculo deste valor.

$$Direccionalidade = \sum_{\forall p \in P} \sum_{\phi \in \omega_p} (\phi - \phi_p)^2 H_D(\phi) \quad (3.16)$$

onde,

P é o conjunto de picos do histograma de direções,

ω_p é o conjunto de intervalos direcionais entre dois vales que delimitam um pico e

ϕ_p é a direção onde se encontra o pico.

d. Características da Filtragem de Gabor

O filtro de Gabor é amplamente utilizado para a extração de características de imagens, especialmente quando trata-se de textura [27]. Apesar de existirem várias propostas de caracterização de textura com base na filtragem Gabor, a idéia principal se mantém praticamente inalterada.

A Equação (3.17) define uma função de Gabor bidimensional onde σ_x e σ_y são os desvios padrões dos núcleos Gaussianos ao longo das direções x e y , respectivamente.

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[-1/2 \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j\omega x \right] \quad (3.17)$$

Assim sendo, um conjunto de filtros de Gabor podem ser obtidos conforme as Equações (3.18) a (3.20), a partir de dilatações e rotações, onde $a > 1$, $\theta = n\pi/K$, $n = 0, 1, \dots, K - 1$, $m = 0, 1, \dots, S - 1$, K é o número de orientações e S é o número de escalas.

$$g_{mn}(x, y) = a^{-m} g(x', y') \quad (3.18)$$

$$x' = a^{-m} (x \cos \theta + y \sin \theta) \quad (3.19)$$

$$y' = a^{-m} (-x \sin \theta + y \cos \theta) \quad (3.20)$$

Seja uma imagem $i(x, y)$, sua transformada de Gabor é definida como

$$W_{mn}(x, y) = \int i(x, y) g_{mn}^*(x - x_1, y - y_1) dx_1 dy_1 \quad (3.21)$$

onde * indica o complexo conjugado.

Por fim, a média μ_{mn} e o desvio padrão σ_{mn} da magnitude de $W_{mn}(x, y)$, ou seja,

$$f_{Gabor} = [(\mu_0, \sigma_0), \dots, (\mu_{mn}, \sigma_{mn}), \dots, (\mu_{S-1 K-1}, \sigma_{S-1 K-1})] \quad (3.22)$$

podem ser utilizados para representar as características de textura de uma imagem.

e. Características da Transformada Wavelet

A transformada *wavelet*, semelhante à filtragem de Gabor, resulta em uma abordagem multi-resolução para a análise e classificação de textura, se decompondo em um sinal com uma família de funções bases $\psi_{mn}(x)$ obtidas por meio de translações e rotações de uma *wavelet* mãe $\psi(x)$, conforme a Equação (3.23).

$$\psi_{mn}(x) = 2^{-m/2} \psi(2^{-m} x - n) \quad (3.23)$$

onde m e n são os parâmetros de dilatação e translação, respectivamente.

Um sinal $f(x)$ pode ser representado como,

$$f(x) = \sum_{m,n} c_{mn} \psi_{mn}(x) \quad (3.24)$$

O cálculo da transformada *wavelet* de um sinal bidimensional consiste de procedimentos de filtragem recursiva e de sub-amostragem. A cada nível o sinal é decomposto em quatro sub-bandas de frequências denominadas *LL*, *LH*, *HL* e *HH*, onde *L* e *H* denotam baixa frequência (*Low*) e alta frequência (*High*), respectivamente.

As principais transformadas *wavelet* aplicadas para a análise de textura são descritas na literatura como *pyramid-structured wavelet transform (PWT)* e *tree-structured wavelet transform (TWT)* [34]. A *PWT* decompõe recursivamente a banda *LL*, enquanto a *TWT* decompõe as demais bandas, quando necessário. Após a decomposição, o descritor de característica pode ser construído a partir dos valores das médias e dos desvios padrões da distribuição de energia de cada sub-banda de frequência, em cada nível.

f. Características Wold

Uma outra abordagem para a descrição de textura em termos de propriedades baseadas na percepção visual humana, conforme descrito anteriormente neste capítulo, é a decomposição Wold, proposta em 1996 por Liu et al. [35].

Nesta abordagem, três componentes – harmonia, evanescência e indeterminismo – correspondem à periodicidade, direcionalidade e aleatoriedade de texturas, respectivamente.

A decomposição Wold bidimensional possibilita que um campo aleatório discreto e homogêneo $\{ y(m,n), (m,n \in \mathbb{Z}) \}$ seja decomposto em três componentes mutuamente ortogonais, como mostrado na Equação (3.25).

$$y(m,n) = i(m,n) + d(m,n) = i(m,n) + h(m,n) + e(m,n) \quad (3.25)$$

onde,

$i(m,n)$ é o componente indeterminístico e

$d(m,n)$, o componente determinístico, que por sua vez pode ser decomposto em

$h(m,n)$, componente harmônico e

$e(m,n)$ componente evanescente.

No domínio do espaço, os três componentes ortogonais podem ser obtidos estimando a probabilidade máxima, que consiste em ajustar um processo auto-regressivo de alta ordem, que minimiza uma função custo e resolve um conjunto de equações lineares.

Ao aplicar a transformada de Fourier na Equação (3.25), temos que

$$F_y(\xi, \eta) = F_i(\xi, \eta) + F_d(\xi, \eta) = F_i(\xi, \eta) + F_h(\xi, \eta) + F_e(\xi, \eta) \quad (3.26)$$

onde as $F(\xi, \eta)$ representam as funções de distribuição espectral das suas respectivas funções no domínio espacial.

No domínio da frequência, componentes *Wold* podem ser obtidos pelos limiares globais de magnitudes do espectro de Fourier das imagens.

3.3.3 Forma

O conceito de forma faz parte do aprendizado humano desde seus primeiros meses de vida, pela identificação de simples objetos [11]. A cognição humana tem na percepção de forma o principal meio para o reconhecimento de objetos. Contudo, a extração de características de forma é um dos aspectos mais difíceis da recuperação de imagens por conteúdo, decorrente da complexidade de segmentar objetos de interesse presentes nas imagens, mantendo sua representação invariante à translação, rotação e escala.

Uma série de técnicas para a descrição de formas foi, e continua sendo desenvolvida, podendo esses métodos ser classificados em dois grupos distintos – o primeiro, que se baseia no contorno, e o outro, baseado em regiões. Essas abordagens ainda podem ser qualificadas entre métodos no domínio do espaço ou no domínio da frequência.

I. Técnicas baseadas no contorno

Técnicas baseadas no contorno [27] investigam somente as informações de bordas (fronteiras) e supõem que essas tenham sido detectadas previamente. As abordagens para a modelagem do contorno de formas podem ser classificadas em contínua ou discreta.

Uma abordagem é **contínua** caso um vetor de característica numérico derivado de toda a borda é utilizado para descrever a forma e medidas de similaridade entre formas são normalmente calculadas por métricas de distância. Descritores simples nesta categoria incluem – *excentricidade (comprimento do eixo maior / comprimento do eixo menor)*, *orientação do eixo maior* e *energia da curvatura*. Demais descritores nessa categoria incluem *a curvatura de espaço de escala e casamento elástico*.

Abordagens **discretas** são aquelas que segmentam as bordas em primitivas utilizando critérios específicos onde a representação final é geralmente uma *string* ou um grafo e sua medida de similaridade é obtida por operações de casamento de *strings* ou grafos [36]. Essas representações podem ser codificadas como uma *string* da forma:

$$S = s_1, s_2, \dots, s_n \quad (3.27)$$

onde s_i pode ser um elemento de um código da cadeia, um lado de um polígono ou conter atributos da forma como comprimento, curvatura, orientação, etc.

A seguir, os principais descritores de forma baseados no contorno são listados, conceituados e definidos. Primeiramente, dos itens *a* a *d*, as principais abordagens contínuas são apresentadas. Então, dos itens *e* a *g*, algumas das abordagens discretas mais aplicadas na literatura também são apresentadas.

a. Histograma de Direção de Bordas

O histograma de direção de bordas é um dos descritores de forma mais populares, principalmente devido a sua simplicidade [27]. A técnica não realiza a segmentação de objetos.

O processo compreende duas etapas. Primeiramente, uma imagem é pré-processada para a extração das fronteiras, bem como suas direções. Para a realização de tal tarefa, operadores de *Canny*, *Sobel* ou derivações *Gaussianas* podem ser utilizados. Então, um histograma é formado pela frequência das direções das fronteiras.

Histogramas de bordas são representações bastante compactas e invariantes à operações de translação e escala.

b. Assinatura de Forma

Uma assinatura de forma é uma representação por meio de uma função matemática unidimensional obtida a partir dos pixels de fronteira em relação à uma referência, como a distância ao centro, ângulo tangente e curvatura, dentre outros [37]. (Figura 3.18).

Assinaturas de forma normalizadas são invariantes a translação e escala, porém possuem bastante sensibilidade a ruído.

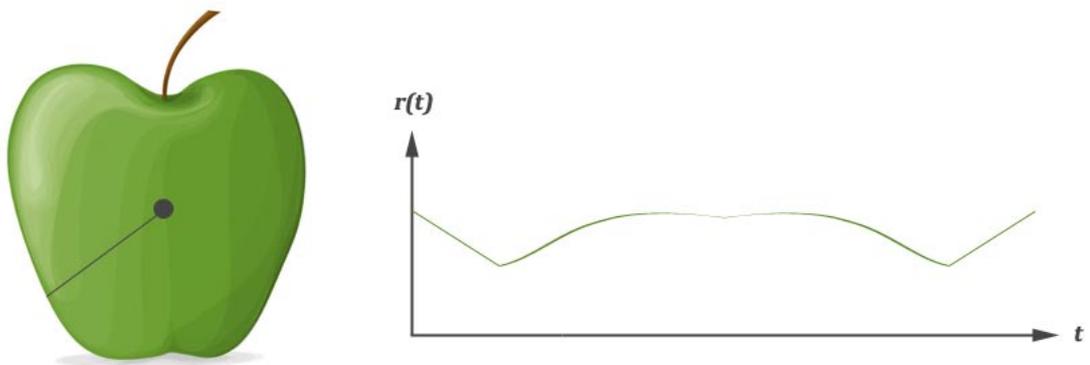


Figura 3.18. Uma forma e sua assinatura, obtida pela distância ao centro.

c. Momentos de Bordas

Os momentos de bordas [37] são utilizados para reduzir as dimensões das representações de bordas por assinatura, por adotar os valores de momentos e seu momento central. Assumindo que a borda de uma forma fora representada por uma assinatura z , o r -ésimo momento m_r e o momento central μ_r podem ser estimados por:

$$m_r = \frac{1}{N} \sum_{i=1}^N |z(i)|^r \quad (3.28)$$

$$\mu_r = \frac{1}{N} \sum_{i=1}^N |z(i) - m_1|^r \quad (3.29)$$

onde N é o número de pixels de borda.

Quando os momentos de borda são normalizados se tornam invariantes às transformações de rotação, translação e escala, tornando-os representações bastante úteis em sistemas CBIR. Os momentos normalizados são calculados pelas Equações (3.30) e (3.31):

$$\bar{m}_r = \frac{m_r}{(\mu_2)^{r/2}} \quad (3.30)$$

$$\bar{\mu}_r = \frac{\mu_r}{(\mu_2)^{r/2}} \quad (3.31)$$

d. Transformadas Espectrais

A técnica de descrição de forma a partir de transformadas espectrais inclui descritores de Fourier e descritores *Wavelet*, derivados de suas respectivas transformadas, dada sua assinatura de forma unidimensional, ou mesmo a partir dos próprios pontos de contorno, diretamente [37].

Os descritores de Fourier consistem na obtenção de coeficientes que representam a fronteira da forma em questão.

Transformadas espectrais resultam em representações compactas e de fácil casamento, pouco sensíveis a ruídos.

e. Código de Cadeia

A representação por códigos de cadeia é uma caracterização estrutural de forma que descreve um contorno fechado, a partir de uma seqüência de segmentos de linhas de tamanho unitário e orientação definida [38].

Sua implementação consiste em uma sobreposição da borda a uma grade e cada ponto de borda é aproximado para o ponto da grade mais próximo, obtendo assim uma borda amostrada, conforme representada na Figura 3.19.

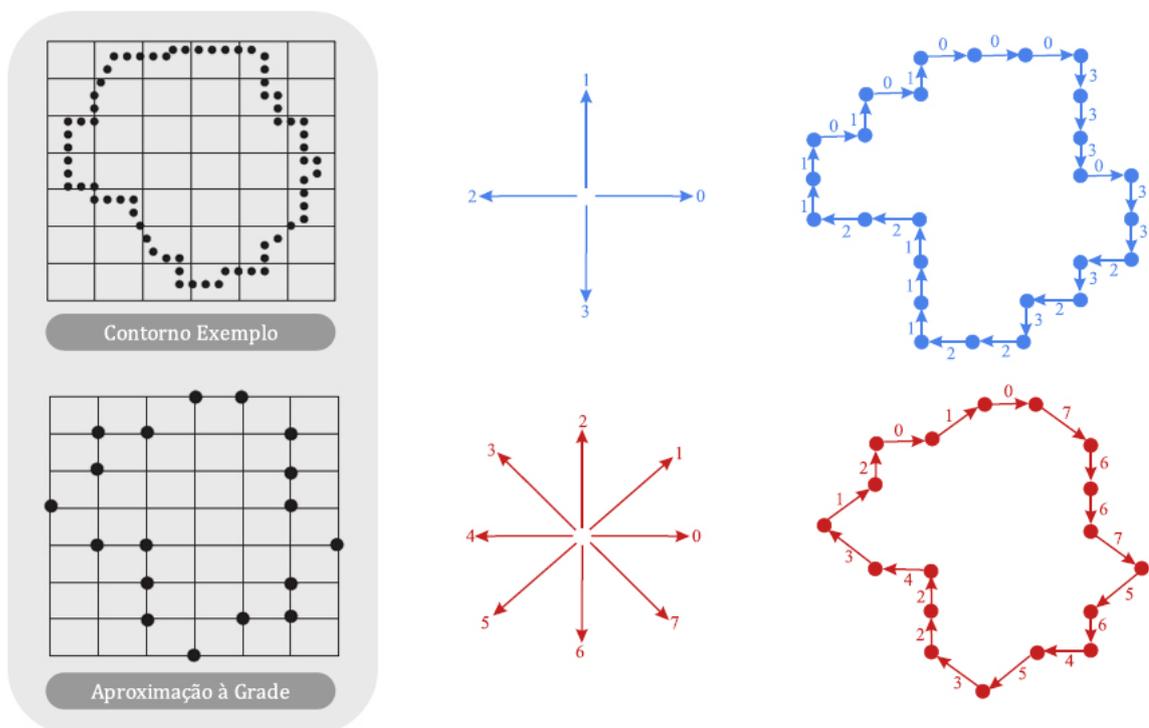


Figura 3.19. Códigos de cadeia para 4 e 8 direções e suas representações.

Códigos de cadeia podem ser gerados a partir de 4, 8 ou N direções e deve ser independente da escolha de seu ponto de partida, ou seja, o código é tratado como uma seqüência circular de números, sendo que o ponto inicial é definido de modo que o número formado possua a menor magnitude possível.

f. Aproximação por Polígonos

Na abordagem de aproximação por polígonos, como o nome indica, um polígono representa por aproximação um contorno fechado na imagem. Por questões de otimização, a técnica busca a representação da fronteira com o menor número possível de segmentos poligonais, considerando a precisão desejada.

O polígono resultante deve possuir um perímetro mínimo para uma fronteira conexa S , de maneira que para que essa otimização seja alcançada, polígonos que possuam forma semelhante à original e o menor número possível de lados devem ser encontrados.

Geralmente, um procedimento geométrico é aplicado para a aproximação por polígonos, a partir da sobreposição da borda em questão por um conjunto de células de tamanho pré-definido para que assim o polígono possa ser traçado dentro dessas células [37].

A Figura 3.20 ilustra esse processo.

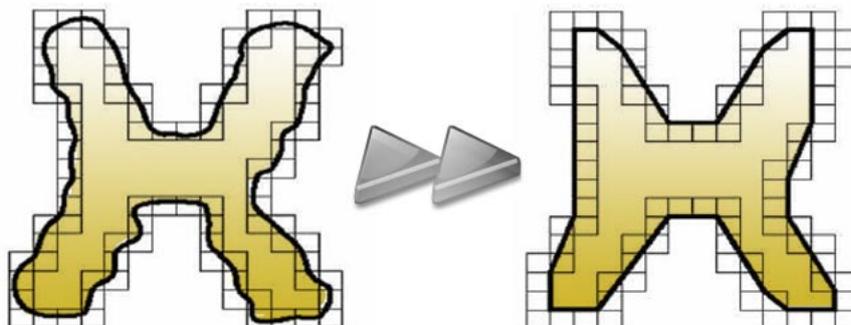


Figura 3.20. Aproximação por polígonos.

g. Análise Sintática

A análise sintática parte do pressuposto que a composição de uma fronteira é semelhante à composição de uma linguagem, portanto, nesta abordagem, as formas são representadas por primitivas e seus relacionamentos descritos a partir de uma gramática, conforme apresentado por Zhang et al. em 2004 [39].

A Figura 3.21 apresenta uma aplicação da análise sintática. Nela, um cromossomo pode ser representado por sua fronteira, que por sua vez, pode ser descrito por algumas primitivas que são denominadas (a, b, c, d) .

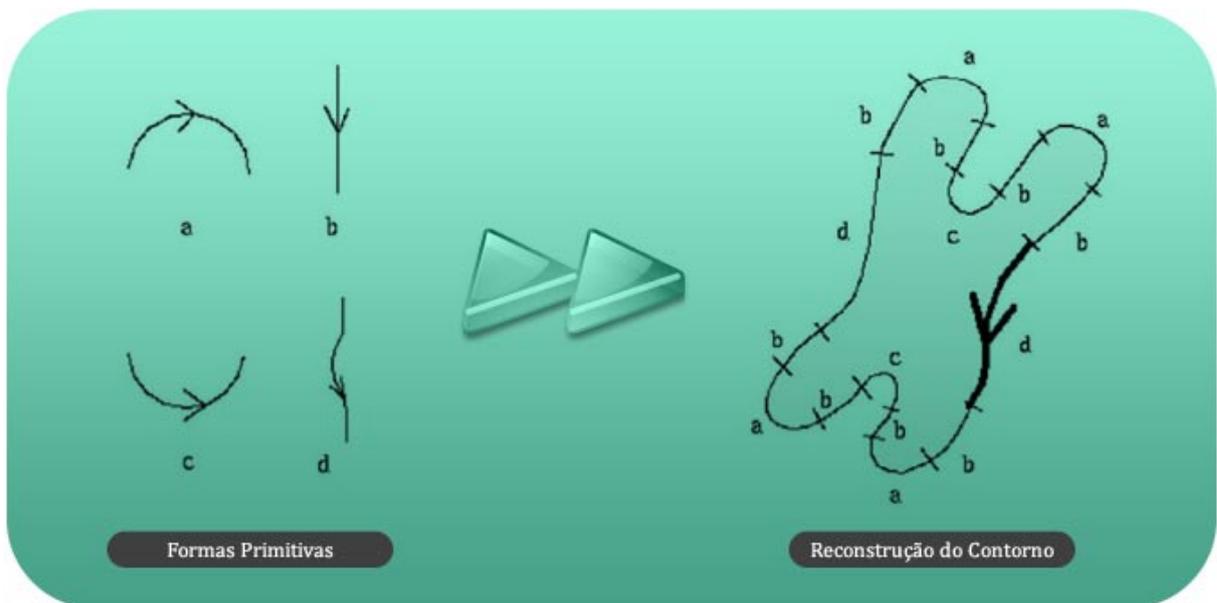


Figura 3.21. Análise sintática.

Uma “palavra” (*string*) composta pelas primitivas adotadas representa a forma final,

$$S = dbabcbabdbabcbab \quad (3.32)$$

Na análise sintática, a similaridade entre formas é mensurada por métodos de casamento de strings, bem conhecidos e apresentados na literatura [36].

II. Técnicas baseadas em regiões

Nas técnicas baseadas em regiões, todos os pixels dentro de uma região da imagem são considerados para a representação de formas, ao invés de utilizar somente as informações das fronteiras. Aqui também, as técnicas destacam-se em dois grupos, classificadas em métodos globais ou métodos estruturais.

Os **métodos globais** tratam a forma como um todo, medindo a distribuição dos pixels por regiões, sendo menos prováveis de serem afetados por ruídos ou variações e resultando em uma descrição de um vetor de características numérico cuja similaridade é mensurada por distâncias. Descritores simples enquadrados nessa categoria incluem – *área*, *circularidade* ($\text{perímetro}^2 / \text{área}$) e *excentricidade*. Demais abordagens nessa categoria incluem descritores genéricos de Fourier, Métodos de Grid e Matrizes de Forma [39].

Os **métodos estruturais** decompõem as regiões da imagem para formarem as descrições.

A seguir, os principais descritores de forma baseados em regiões são listados, conceituados e definidos. Primeiramente, nos itens *a* e *b*, as principais abordagens globais são apresentadas, e depois, nos itens *c* e *d*, algumas das abordagens estruturais são apresentadas.

a. Momentos Geométricos

Seja o objeto R representado como uma imagem binária, então os momentos centrais de ordem $p + q$ são definidos como

$$\mu_{p,q} = \sum_{(x,y) \in \mathbb{R}} (x - x_c)^p (y - y_c)^q \quad (3.33)$$

onde (x_c, y_c) é o centro do objeto [27].

O momento central pode ser normalizado para se tornar invariante a escala conforme apresentado nas Equações (3.34) e (3.35).

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\gamma} \quad (3.34)$$

$$\gamma = \frac{p + q + 2}{2} \quad (3.35)$$

Destas equações, um conjunto de momentos geométricos invariantes a translação, rotação e escala podem ser derivados. As equações de sete momentos geométricos são apresentados por Long et al. [27].

b. Momentos Zernike

A técnica de momentos Zernike utiliza funções de bases ortogonais e são menos sensíveis a ruídos do que os momentos geométricos, mantendo a qualidade de invariação quanto às operações de translação, rotação e escala.

Momentos Zernike bidimensionais são dados por:

$$A_{m,n} = \frac{m + 1}{\pi} \int f(x, y) [V_{m,n}(x, y)]^* dx dy \quad (3.36)$$

onde,

$$x^2 + y^2 \leq 1$$

$$m = 0, 1, 2, \dots, \infty$$

$f(x,y)$ é uma função que descreve a imagem,

$V_{m,n}$ denota a função de base Zernike e $*$ o seu complexo conjugado, e

n é um inteiro que representa a dependência angular ou rotação.

c. Fecho Convexo

Uma região r é convexa se e somente se para quaisquer dois pontos x_1 e $x_2 \in r$, o segmento de reta x_1x_2 estiver completamente dentro da região r . O fecho convexo de uma região é a menor região convexa H que satisfaz a condição $R \subset H$. A deficiência convexa D da região R é obtida pela diferença $H - R$ [39].

Regiões e deficiências convexas podem ser descritas em forma de árvores de dados, conforme mostrado na Figura 3.22.

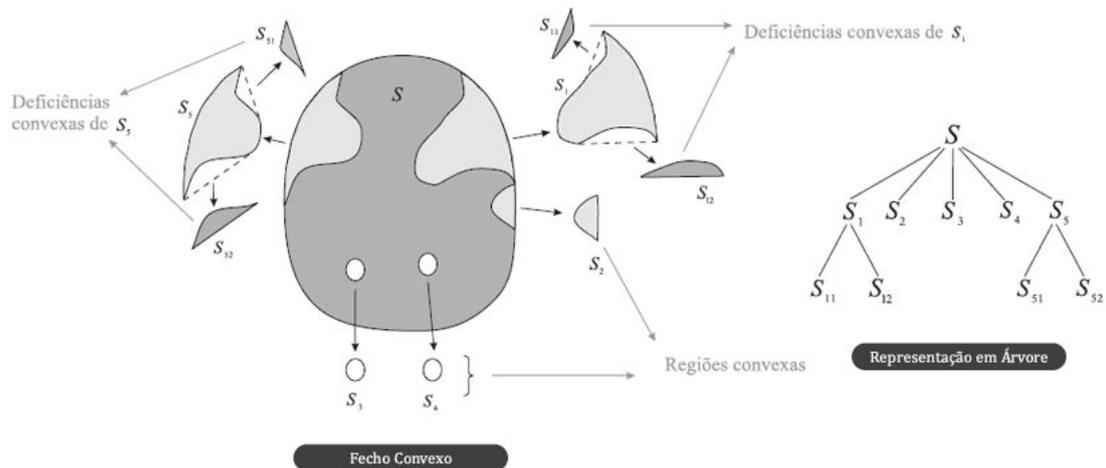


Figura 3.22. Fecho convexo.

d. Saliências

Saliências são utilizadas para descrever pontos de fronteira de alta curvatura ou as extremidades dos esqueletos de forma [40]. A Figura 3.23 apresenta três ilustrações de saliências de uma folha, onde se tem os pontos de saliência do contorno, as saliências do esqueleto interno e as saliências do esqueleto externo.

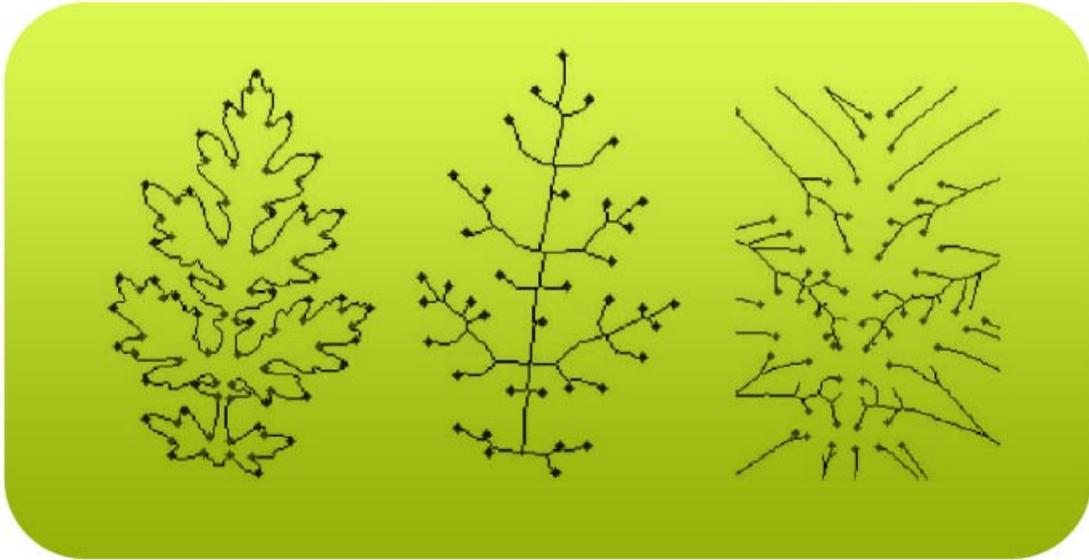


Figura 3.23. Saliências.

3.4 Métricas de Similaridade

A similaridade é uma estimativa de semelhança entre imagens e geralmente representa dois fatores fundamentais – a similitude entre características e o grau de importância de cada uma destas. Um modelo de similaridade representa uma estratégia de como mensurar e tratar cada característica e sua representatividade de forma a otimizar a resposta do sistema de recuperação de imagens, conforme os requisitos do usuário.

Como introduzido na seção anterior, as imagens podem ser representadas por múltiplas características como cor, forma e textura, sejam estas extraídas da imagem ou sub-regiões da mesma. Devido ao fato de, além das características terem diferentes importâncias, e as regiões disporem de representatividades distintas, um modelo de similaridade que agregue esses pormenores faz-se fundamental para o efetivo desenvolvimento e projeto de um sistema de recuperação de imagens por conteúdo.

Boa parte dos inúmeros modelos de similaridade apresentados na literatura se baseia na utilização de médias ponderadas de similaridades parciais de características e regiões e o que geralmente as diferencia é a seleção de uma técnica para a apuração dos pesos, seja ela uma heurística, a inserção de dados pelo próprio usuário ou o auto-ajuste de pesos por mecanismos de realimentação de relevância.

A partir do momento que as características das imagens foram extraídas e armazenadas em vetores de características, métricas que comparam estes vetores retornando valores de similaridade se fazem necessárias.

As medidas de similaridade são geralmente baseadas em medidas de distância. As mais importantes e efetivas são apresentadas a seguir.

a. Distância de Minkowski

A distância de Minkowski é aplicada e de utilização apropriada quando as coordenadas dos vetores de características são independentes entre si e de igual importância. A Equação (3.37) apresenta sua formulação matemática.

$$D(X, Y) = \sum_{i=1}^N [|X_i - Y_i|^\lambda]^{1/\lambda} \quad (3.37)$$

A **distância Euclidiana** é um caso particular da distância de Minkowski onde $\lambda = 2$:

$$D(X, Y) = \sqrt{\sum_{i=1}^N (X_i - Y_i)^2} \quad (3.38)$$

A **distância em quadras**, conhecida na literatura como *City-block*, também é um caso especial da distância de Minkowski, para $\lambda = 1$:

$$D(X, Y) = \sum_{i=1}^N |X_i - Y_i| \quad (3.39)$$

b. Distância Quadrática

A distância quadrática foi introduzida para resolver a especificidade da distância de Minkowski de tratar as coordenadas dos vetores de características de forma inteiramente independente, não levando em consideração suas representatividades, ou seja, o fato de que determinadas coordenadas podem corresponder a características perceptivas mais importantes que outras. Assim sendo, a distância quadrática entre X e Y é obtida pela equação abaixo.

$$D(X, Y) = \sqrt{(X - Y)^T A (X - Y)} \quad (3.40)$$

onde,

$$A = [a_{ij}] \text{ e}$$

a_{ij} denota a similaridade entre características de índices i e j .

c. Distância de Mahalanobis

A distância de Mahalanobis é aplicada e apropriada nos casos onde as coordenadas dos vetores de características são independentes e de diferentes importâncias, tal como na distância quadrática. Entretanto, nessa abordagem de similaridade, a matriz de covariância S dos vetores de características é utilizada, como mostra a Equação (3.41).

$$D(X, Y) = \sqrt{(X - Y)^T S^{-1} (X - Y)} \quad (3.41)$$

d. Outras distâncias

Várias outras medidas de distâncias são encontradas na literatura. Medidas de **Canberra**, **Chebyshev**, **Bottleneck**, **Hausdorff**, **Fréchet**, **Earth Movers**, **Banach-Mazar** e **Cosseno** são parte da imensa lista de medidas de distâncias utilizadas para o cálculo da similaridade entre vetores de características de imagens digitais. No entanto, a medida de Minkowski é a mais aplicada em sistemas de recuperação de imagens [27].

3.5 Realimentação de Relevância

Em sistemas de recuperação da informação, o objetivo é apresentar documentos relevantes e úteis ao usuário, resultantes de uma consulta ao sistema. Quando estes sistemas tratam de imagens ao invés de textos, a complexidade é amplamente aumentada devido aos problemas dos saltos sensorial e semântico, já apresentados neste trabalho.

Sistemas CBIR devem interpretar o conteúdo visual das imagens de busca de forma a apresentar respostas relevantes com imagens similares contidas no banco de imagens, ordenando-as de maneira que os documentos classificados como mais semelhantes à necessidade de informação do usuário sejam apresentados à frente dos demais, com menor similaridade.

A interpretação de conteúdos visuais faz-se complexa devido à difusa percepção humana que compreende seus conhecimentos, suas experiências, suas emoções, sua imaginação e tantas outras qualidades objeto de estudo, principalmente de ciências humanas como a Psicologia. Dessa forma, programar uma máquina para realizar eficientemente uma tarefa que nós seres humanos temos dificuldade de contextualizar se torna um desafio.

A dificuldade do processo, além da extração de informações, é de como utilizá-las para decidir o grau de relevância das imagens. Como apresentado anteriormente, a caracterização de imagens por cor, forma e textura é o primeiro passo para que um sistema tenha métricas para classificar e ordenar documentos conforme sua estratégia de similaridade.

O conceito de relevância pode ser considerado o cerne de qualquer tarefa de recuperação da informação. Entretanto, é o usuário quem melhor conhece o que é relevante no contexto de uma busca em particular. Por isso, surgiu a necessidade por uma aplicação que compreendesse a integral participação do usuário no processo de recuperação, a qual foi denominada Realimentação de Relevância, também reconhecida na literatura por *relevance feedback* [6].

Conforme Baeza-Yates et al. [7] e Salton et al. [41], pode-se definir realimentação de relevância como uma forma controlada de modificar uma consulta a um sistema de recuperação de informações. A idéia principal é a de utilizar informações fornecidas pelo usuário acerca da relevância dos documentos previamente recuperados com a finalidade de aprimorar a qualidade dos resultados recuperados.

A realimentação de relevância é motivada pelo fato de que a maioria dos usuários desses sistemas encontra dificuldades na formulação de consultas que atendam de imediato seus requisitos de recuperação. Geralmente, isto ocorre pelo fato desses usuários não conhecerem a coleção de documentos do sistema de recuperação o suficiente, ou pela falta de familiaridade com o ambiente de recuperação, causando complicações na execução de buscas satisfatórias.

Sob a abordagem da realimentação de relevância, a primeira operação de recuperação é considerada uma mera tentativa, que associada às informações inseridas pelo usuário, refinam o escopo da busca iterativamente até atenderem os propósitos do usuário.

Considerando especificamente os sistemas CBIR, um exemplo de aplicação da realimentação de relevância é apresentado a seguir. A consulta inicial pode ser uma imagem exemplo que retrata diversos elementos como uma floresta, uma lagoa, montanhas e uma casa de campo, sendo que o real desejo do usuário é de recuperar cenas que contenham embarcações, mas por falta de uma imagem dessa classe, apresenta outra que contém uma lagoa, na expectativa de encontrar o que procura.

A princípio, a não ser por um lance de sorte, pode-se dizer que o sistema não terá êxito na primeira apresentação de resultados, pela impossibilidade de se especular, aproximar, ou adivinhar o que é relevante ao usuário neste momento.

Contudo, devido à caracterização visual das imagens, a probabilidade de que o sistema retorne algumas imagens contendo somente casas de campo, somente montanhas, somente florestas ou somente lagoas é considerável. Assim sendo, após apresentados os resultados, o usuário pode selecionar somente as imagens que contenham lagoas, mares e documentos que se assemelham à seu propósito final de busca.

A pesquisa é então refinada, ponderando características componentes dessas imagens selecionadas a partir da realimentação de relevância de maneira a torná-las mais significantes perante as demais.

Após consecutivas inserções de relevância pelo usuário, há uma boa chance de que o usuário encontre imagens de embarcações, como desejado a priori. Isto significa que o sistema teve sucesso na recuperação de imagens relevantes.

Devido ao modo de operação de um sistema CBIR, há uma subjetividade inerente ao processo de busca. Quando o usuário fornece como consulta uma imagem exemplo, o sistema percebe essa imagem como um conjunto de características de baixo nível enquanto o usuário a percebe pela sua semântica. A mesma imagem de consulta pode ser interpretada de formas distintas, até mesmo pelo mesmo indivíduo, em momentos diferentes.

Dessa forma, a utilização de mecanismos de realimentação de relevância, geralmente herdados de sistemas de recuperação de informação textual, se faz bastante interessante por tornar o usuário parte presente e fundamental ao sucesso do sistema de recuperação sem causar efeitos colaterais. Com a aplicação da realimentação de relevância observa-se uma notável amenização do *gap* entre conceitos de alto nível e características de baixo nível de imagens.

Dentre as mais populares abordagens de realimentação de relevância encontradas na literatura [42], observa-se a utilização de duas formas de ponderação:

- pesos positivos, quando a imagem é de certa forma relevante; e
- pesos positivos e negativos, objetivando não somente refinar a pesquisa, mas também penalizar imagens que não se relacionam com o propósito do usuário.

A aplicação dessas formas de ponderação pode ser efetuada pela simples seleção ou clique em um botão, como maneira de informar se o documento é ou não relevante, ou por meio de uma escala, onde o usuário tem o controle de quão relevante ou não, ou quanto penalizará determinada imagem em uma recuperação. A Figura 3.4 apresenta tais abordagens.

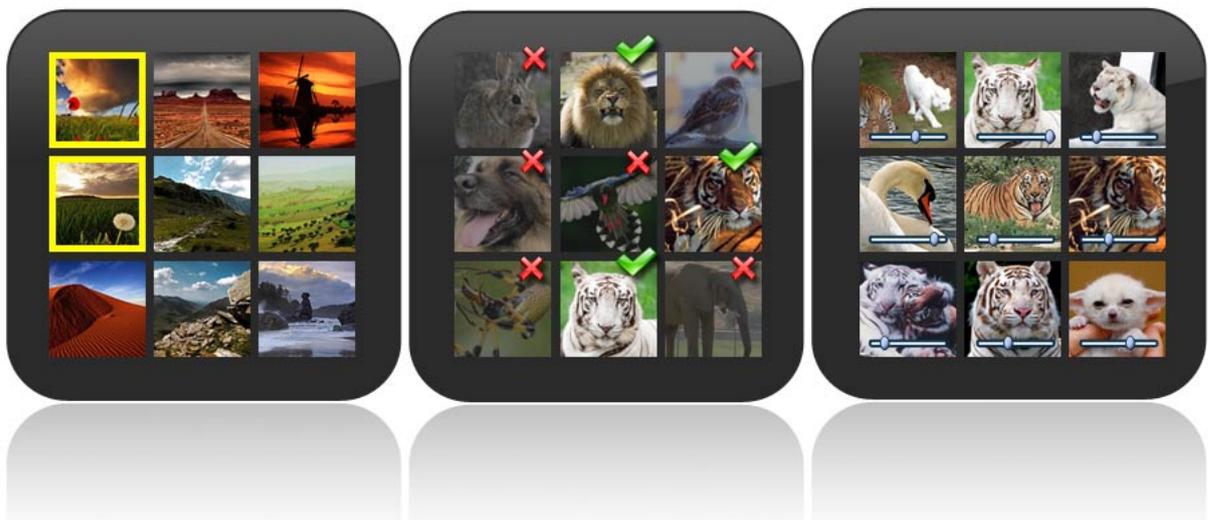


Figura 3.24. Diferentes estratégias de realimentação de relevância.

Na tentativa de capturar os requerimentos do usuário, os mecanismos de realimentação de relevância modificam a representação da consulta ou então alteram a medida de similaridade empregada para recuperar e ordenar os documentos da coleção de referência [42].

Dentre as maneiras de modificação da representação da consulta pode-se citar a computação de uma nova representação a partir da técnica conhecida como *query point movement*, apresentada por Ciocca et al. em 1999 [43] ou pela expansão da consulta, reconhecida na literatura como *query expansion*, que normalmente utiliza múltiplas representações da técnica anterior, sendo chamada de *multiple query points movement*, proposta por French et al. em 2004 [44].

As formas de modificação da medida de similaridade entre imagens são voltadas principalmente para o reajuste de pesos de características e de regiões conforme o conceito de relevância por parte do usuário. Destas, diversas abordagens e ferramentas empregadas podem ser citadas – técnicas estatísticas (como correlação, variância, etc.), redes neurais artificiais, algoritmos genéticos, programação genética e redes bayesianas, entre outros.

3.6 Medidas de Desempenho

Medidas de desempenho são métricas utilizadas para comparar sistemas de recuperação de imagens por conteúdo de acordo com os resultados obtidos de consultas realizadas com inúmeros quesitos de recuperação, visando o aprimoramento desses sistemas, sua eficácia e eficiência, além da capacidade de classificar esses sistemas como mais ou menos adequados para determinadas aplicações.

As medidas mais comumente utilizadas para avaliar o desempenho de um CBIR são, como citado a priori, a eficácia (se tem sucesso na execução da tarefa a qual fora desenhado) e a eficiência (quão bem realiza a tarefa). Também se pode associar a eficácia ao nível de exatidão da recuperação frente aos propósitos iniciais do usuário ou da consulta e a eficiência ao custo computacional do processo.

As principais medidas de eficácia de recuperação utilizadas na análise de desempenho de sistemas de recuperação de imagens por conteúdo visual são apresentadas a seguir.

3.6.1 Precisão-Revocação

O critério quantitativo mais utilizado para a avaliação de sistemas de recuperação de informação é a curva precisão-revocação, ou *precision-recall* [7].

Seja $Pr(q)$ a medida de precisão para uma consulta q , $Re(q)$ a medida de revocação para a mesma consulta, M o total de imagens recuperadas, $Ra(q)$ o número de imagens relevantes recuperadas e $A(q)$ o total de imagens relevantes na coleção.

Então, a **precisão** para uma consulta q pode ser definida como a razão entre o número de imagens relevantes dentre o total de imagens recuperadas e a **revocação** para a mesma consulta q , é obtida pela razão entre o número de imagens relevantes recuperadas do total de imagens relevantes da coleção, como demonstram as Equações (3.42) e (3.43) abaixo.

$$Pr(q) = \frac{Ra(q)}{M} \quad (3.42)$$

$$Re(q) = \frac{Ra(q)}{A(q)} \quad (3.43)$$

Essas equações avaliam a precisão e a revocação de uma única consulta q . Usualmente, os sistemas de recuperação de imagens por conteúdo visual são avaliados para

várias consultas distintas, para que se tenha uma medida estatisticamente confiável. Nesses casos a abordagem adotada é o cálculo da média de Precisão e da média de Revocação para as n consultas realizadas.

3.6.2 Precisão- R

A métrica de desempenho Precisão- R é aplicada quando se quer estabelecer a precisão de imagens relevantes recuperadas dentre os R primeiros resultados do *ranking*.

Um exemplo de utilização desta medida de desempenho seria a avaliação das 100 primeiras imagens recuperadas por dois sistemas CBIR distintos. Após a recuperação de imagens pelos sistemas, pode-se analisar o número de imagens relevantes recuperadas entre as 100 primeiras, em cada um deles, obtendo assim o valor de Precisão- R , para $R = 100$.

3.6.3 Precisão Média para Níveis de Revocação

A precisão média para níveis de revocação, derivada da análise de *precisão-revocação*, utiliza o cálculo da precisão média em níveis fixos de revocação, como por exemplo $(0,00; 0,10; 0,20; 0,30; \dots; 1,00)$ ou $(0,25; 0,50; 0,75)$.

O cálculo da precisão média para n consultas é obtido pela Equação (3.44), onde $\overline{Pr}(r)$ é a precisão média no nível de revocação r , N_q é o número de consultas de teste e $Pr_i(r)$ é a precisão no nível de revocação r para a i -ésima consulta.

$$\overline{Pr}(r) = \sum_{i=1}^{N_q} \frac{Pr_i(r)}{N_q} \quad (3.44)$$

3.6.4 ANMRR (*Average Normalized Modified Retrieval Rank*)

O ranking de recuperação médio normalizado e modificado, conhecido na literatura como *average normalized modified retrieval rank*, ou simplesmente ANMRR, foi inicialmente proposto pelo *Moving Picture Element Group (MPEG)* como uma métrica de desempenho alternativa, resumindo a performance do sistema em um único número que leva em conta a posição no ranking das imagens relevantes recuperadas, devido ao fato dos sistemas CBIR serem usualmente avaliados principalmente para baixos níveis de revocação.

Seja n_Q o número de consultas realizadas e n_I o número de imagens relevantes na coleção. Então, para uma consulta q , considerando $A(q)$ como o conjunto de imagens relevantes na coleção e $|A(q)|$ como o número de elementos de $A(q)$, ANMRR é calculado a partir das Equações (3.45) a (3.47), abaixo.

$$ANMRR = \frac{1}{n_Q} \sum_{q=1}^{n_Q} \frac{MRR(q)}{C(q) + 0,5 - \frac{|A(q)|}{2}} \quad (3.45)$$

onde,

$$MRR(q) = \frac{1}{|A(q)|} \left[\sum_{k=1}^{n_I} Rank(k, q) \right] - 0,5 - \frac{|A(q)|}{2} \quad (3.46)$$

$$C(q) = \min \{4 \times |A(q)|, 2 \times \max_{k=1}^{n_Q} |A(k)|\} \quad (3.47)$$

A função $Rank(k, q)$ fornece a utilidade de uma imagem que aparece na k -ésima posição do ranking da consulta q ,

$$Rank(k, q) = \begin{cases} k, & \text{se } k \leq C(q) \text{ e a } k \text{ ésima imagem } \in A(q) \\ C(q) + 1, & \text{se } k > C(q) \text{ e a } k \text{ ésima imagem } \in A(q) \\ 0, & \text{caso contrário} \end{cases} \quad (3.48)$$

O valor de *ANMRR* está compreendido no intervalo $[0, 1]$, sendo um baixo valor indicador de uma alta taxa de recuperação. Idealmente, $ANMRR = 0$, se todas as imagens relevantes aparecerem nas primeiras posições do *ranking*. O pior caso, quando $ANMRR = 1$, representa ineficácia de recuperação, onde nenhuma imagem relevante aparece entre os primeiros resultados do *ranking*.

3.6.5 *ANSMD (Average Normalized Similarity Metric Difference)*

A avaliação de desempenho de um sistema CBIR a partir do cálculo do erro absoluto entre a medida de similaridade corrente e a ideal para todas as imagens recuperadas, relevantes ou não, foi proposta em 2006 por Doulamis et al. [42].

Conceitualmente, uma métrica de similaridade retorna um valor próximo a 1 (ou 100%) para todas as imagens relevantes e um valor próximo a zero para imagens irrelevantes. Desta maneira, o erro absoluto para uma consulta q , obtido por $e(q)$, pode ser definido como

$$e(q) = \frac{1}{M} \sum_{k \in S_M} |d_n(f_q, f_k) - \tau(q, k)|_2 \quad (3.49)$$

onde,

$d_n(f_q, f_k)$ expressa a medida de similaridade corrente e normalizada entre q e a k -ésima imagem das M primeiras recuperadas,

$\tau(q, k)$ refere-se a medida de similaridade ideal entre as respectivas imagens, e

a norma L_2 (distância Euclidiana) é representada por $|\cdot|_2$.

A medida final utilizada como métrica de desempenho após a realização de várias consultas é o erro médio sobre todas as consultas, sendo que quanto menor o erro médio, melhor é o desempenho do sistema.

Assumindo que Q diferentes consultas tenham sido realizadas, temos que

$$ANSMD = \frac{1}{Q} \sum_{q=1}^Q e(q) \quad (3.50)$$

onde *ANSMD* é a sigla para *average normalized similarity metric difference*, sendo definida como a simples média dos erros absolutos para cada uma das Q pesquisas.

3.7 Apresentação de Resultados

O projeto de um sistema de recuperação de imagens por conteúdo compreende uma série de etapas importantes para que a aplicação final implementada tenha êxito no propósito que fora concebido, no contexto de um sistema de recuperação da informação. Desde a escolha dos descritores de imagens, a técnica de armazenamento, a forma de elaboração de consultas pelos usuários, as medidas de distância empregadas, a modelagem da similaridade entre imagens adotada, até a utilização e implementação de mecanismos de realimentação de relevância, os projetos de CBIR se diferenciam pelas decisões tomadas a cada etapa e pelas estratégias adotadas, tornando-os projetos de software um tanto complexos.

Além das decisões de projeto de como matematicamente representar as imagens em uma coleção e como avaliar a similaridade entre um par de imagens, baseando-se em suas abstrações numéricas descritivas, a visualização dos resultados de consulta é considerado um dos principais fatores na aceitação e popularidade de um sistema de recuperação, segundo Datta et al. [11].

De todas as implementações analisadas, pode-se classificar as estratégias de visualização de resultados de pesquisa em cinco abordagens principais, listadas e conceituadas a seguir.

a. Apresentação Ordenada por Relevância

Essa abordagem apresenta os resultados da pesquisa em ordem (*ranking*) de acordo com a similaridade, medida numérica da relevância entre a imagem de consulta e as imagens recuperadas do banco de imagens. A estratégia de visualização de resultados ordenada por relevância é a mais popular entre os sistemas CBIR da atualidade. A Figura 3.25 apresenta esta abordagem em uma consulta por águias.

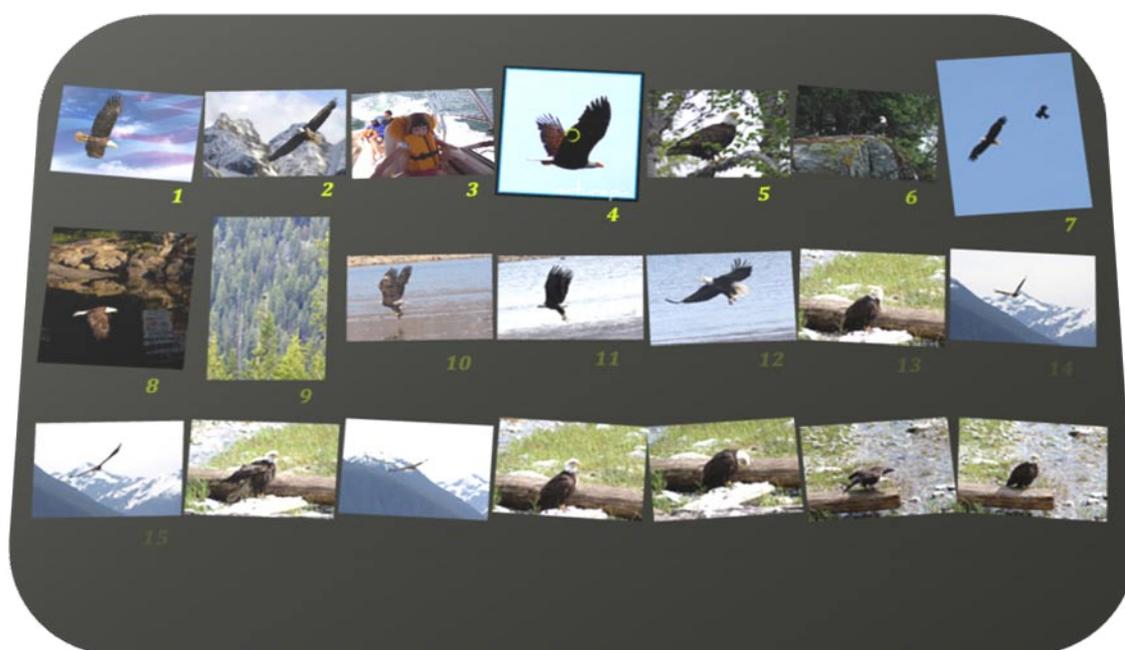


Figura 3.25. Visualização de resultados ordenados por relevância.

b. Apresentação Ordenada por Recuperação (*por Tempo*)

Esse modelo apresenta os resultados da pesquisa conforme os documentos considerados relevantes são encontrados na coleção, em ordem cronológica, não seguindo a ordem direta por similaridade.

Uma vantagem dessa estratégia é o tempo de resposta do sistema, que pode superar outras abordagens. Contudo, a apresentação de resultados sem a utilização de um ranking

pode tornar o sistema um pouco confuso para o usuário comum, justificando a utilização dessa estratégia em casos específicos onde o tempo de resposta pode ser fator crucial da aplicação. A Figura 3.26 apresenta a abordagem ordenada por tempo.

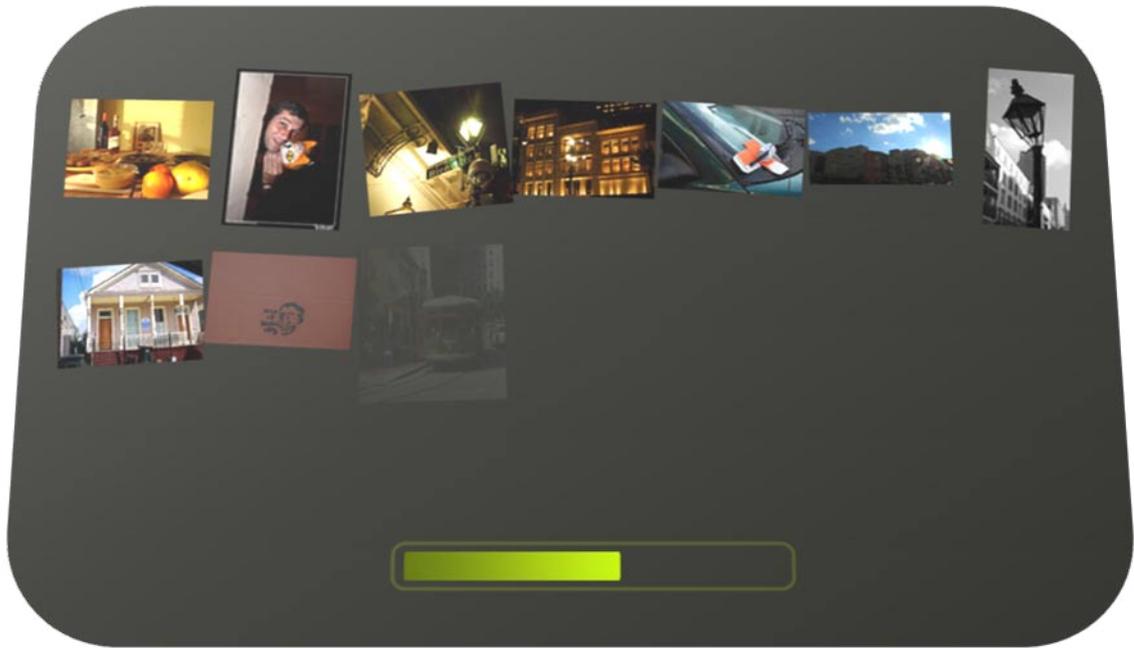


Figura 3.26. Visualização de resultados ordenados por tempo de recuperação.

c. Apresentação por Agrupamento

Nessa técnica, as imagens são agrupadas pelos seus meta-dados ou aspectos visuais [45], gerando uma forma espacial intuitiva de apresentação de resultados. A apresentação por agrupamento vem sendo empregada para o aprimoramento do desempenho da recuperação, normalmente associada à mecanismos de realimentação de relevância. A Figura 3.27 a seguir ilustra esta abordagem.



Figura 3.27. Visualização de resultados de forma agrupada.

d. Apresentação por Hierarquia

Utilizada principalmente para visualização de arquivos de grandes coleções de artes ou de cunho educacional, a apresentação por hierarquia é aplicada se meta-dados podem ser associados às imagens a partir de estruturas de dados conhecidas como árvores, que representam relações entre itens de coleções de maneira hierárquica.

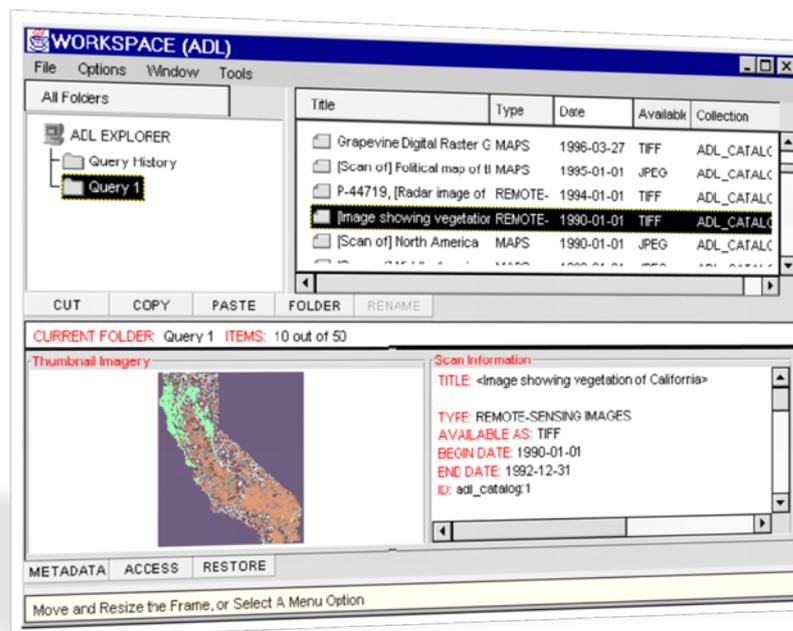
e. Visualização Composta

A visualização composta é a combinação de quaisquer das técnicas de apresentação citadas nesta seção.

3.8 Exemplos de CBIR

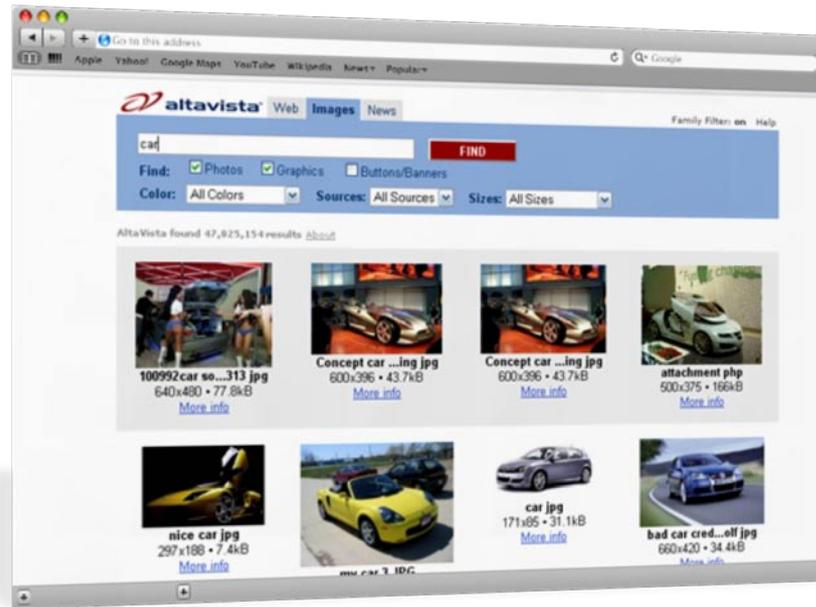
Inúmeras são as aplicações de sistemas CBIR, tanto academicamente quanto comercialmente. Veltkamp [10] apresenta e compara 58 sistemas CBIR apontando suas principais características. A seguir, as principais características, estratégias empregadas e aplicações de alguns sistemas de recuperação de imagens por conteúdo são descritas.

a. Alexandria Digital Library



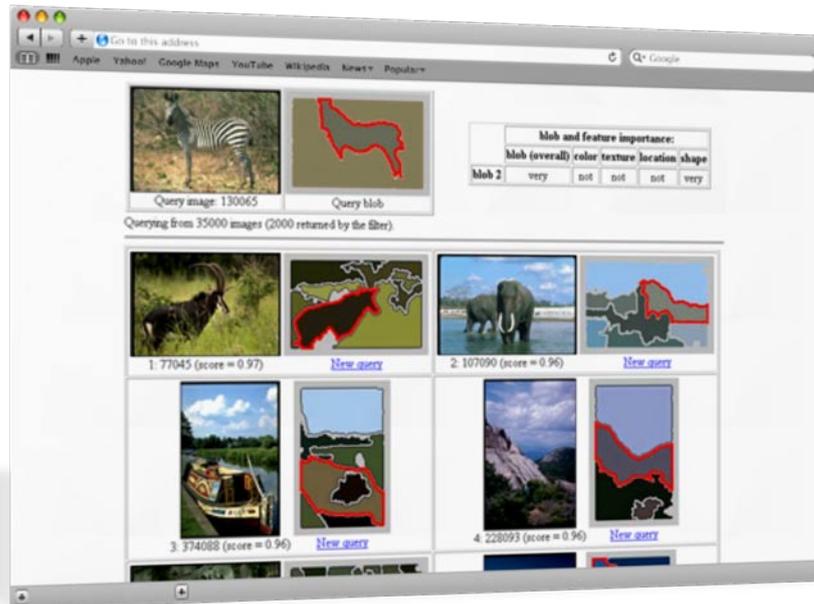
<i>Desenvolvido por</i>	Universidade da Califórnia, Santa Bárbara.
<i>Acessível pelo endereço</i>	http://www.alexandria.ucsb.edu
<i>Elaboração de Consultas</i>	A partir de um mapa interativo bidimensional, o usuário marca uma região do planeta de interesse que contenha ou sobreponha imagens no banco de dados adicionando parâmetros textuais e palavras-chave.
<i>Apresentação de Resultados</i>	As imagens recuperadas são apresentadas por meio de miniaturas e o local onde fora capturada é retratado no mapa.
<i>Principais Aplicações</i>	Suporte a ciência básica a partir da inclusão de imagens aéreas, de satélites, de sensoriamento remoto, bancos de dados sísmicos, etc.

b. AltaVista Photofinder



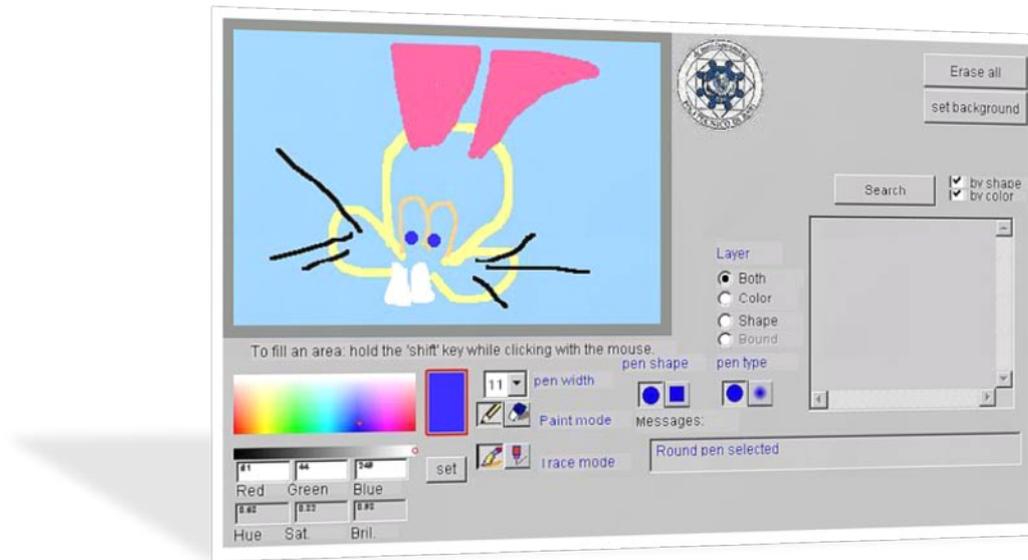
<i>Desenvolvido por</i>	DEC Research Lab, companhia comprada pela AltaVista
<i>Acessível pelo endereço</i>	http://www.altavista.com/image/default
<i>Elaboração de Consultas</i>	Primeiro, o usuário digita uma palavra para iniciar a busca. A imagem recuperada pode apresentar uma <i>link</i> com a palavra 'similar', que se for clicada, lista uma série de imagens similares.
<i>Medidas e Modelo de Similaridade</i>	O cálculo da similaridade é baseado nas características visuais como cores, formas e texturas dominantes. Maiores detalhes sobre os descritores de características não são informados.
<i>Apresentação de Resultados</i>	Resultados apresentados a partir de miniaturas, sem ordem explicitamente estabelecida.

c. Blobworld



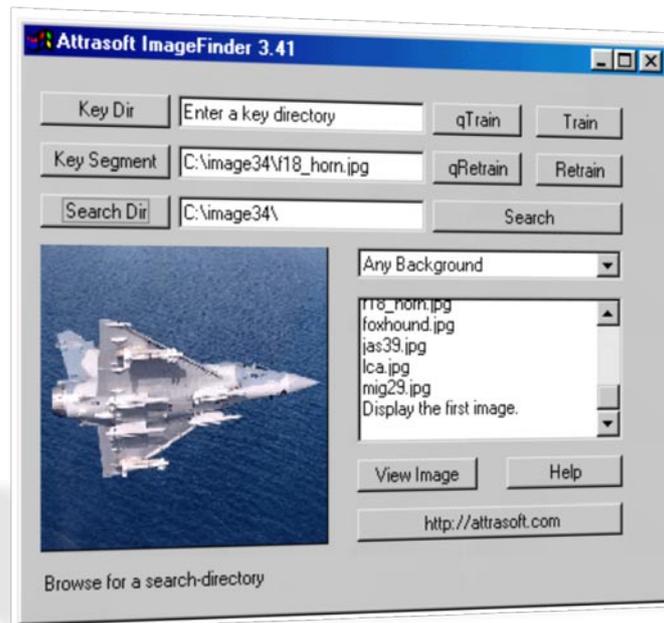
<i>Desenvolvido por</i>	Universidade da Califórnia, Berkeley.
<i>Acessível pelo endereço</i>	http://elib.cs.berkeley.edu/photos/blobworld
<i>Disponível ao público</i>	Até 2007
<i>Descritores de Características</i>	Cor, textura, localização e forma.
<i>Elaboração de Consultas</i>	O usuário seleciona uma categoria limitando o espaço de consulta e então marca uma região (<i>blob</i>) indicando sua importância ('certa' ou 'grande') e a importância de cada característica. Mais de uma região pode ser marcada para consulta.
<i>Medidas e Modelo de Similaridade</i>	Distância quadrática para histograma de cores, distância Euclidiana para textura (contraste médio), distância entre centróides para localização das regiões, todas combinadas em um somatório final.
<i>Apresentação de Resultados</i>	Ordenados linearmente e apresentados juntamente com uma versão segmentada da imagem recuperada.
<i>Principais Aplicações</i>	Recuperação para uma coleção de 10.000 imagens do banco <i>Corel Stock Photos</i> .

d. DrawSearch



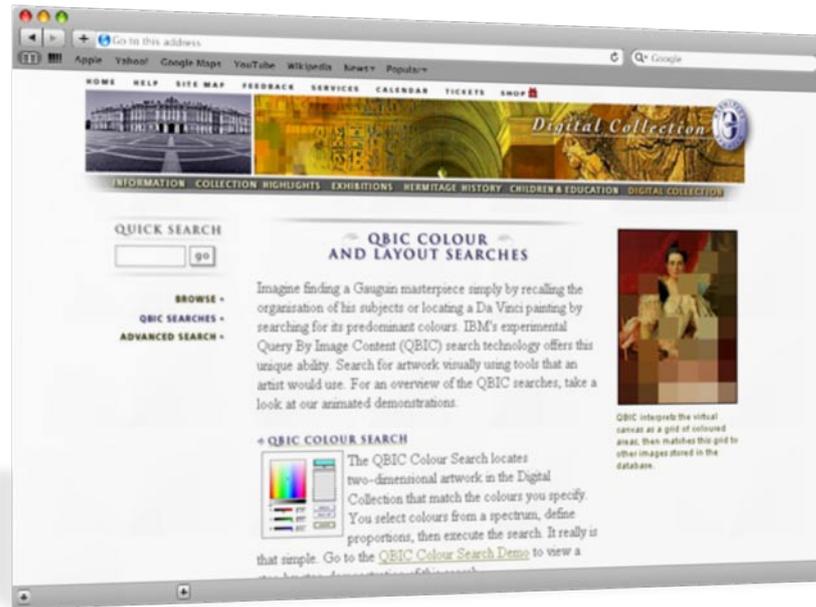
<i>Desenvolvido por</i>	Universidade Técnica de Bari, Itália.
<i>Descritores de Características</i>	Cor, textura e forma.
<i>Elaboração de Consultas</i>	O usuário desenha um rascunho objeto de sua consulta com opções de seleção de cores e atributos para o pincel
<i>Medidas e Modelo de Similaridade</i>	As imagens são divididas em 16 sub-imagens (4x4) e seus vetores de características são comparados a partir da soma ponderada das distâncias calculadas para descritores de cor e forma.
<i>Apresentação de Resultados</i>	Ordenada por relevância.
<i>Realimentação de Relevância</i>	Presente somente para textura.

e. ImageFinder (*AttraSeek*)



<i>Desenvolvido por</i>	Attrasoft Inc.
<i>Acessível pelo endereço</i>	http://atrasoft.com
<i>Elaboração de Consultas</i>	Inserção de imagem exemplo pelo usuário.
<i>Medidas e Modelo de Similaridade</i>	Utiliza a técnica 'Boltzman Machine', um tipo probabilístico de rede neural artificial, sendo que esta rede deve primeiramente ser treinada.
<i>Apresentação de Resultados</i>	Uma página com uma lista textual de links para as imagens recuperadas.

f. QBIC



<i>Desenvolvido por</i>	IBM Almaden Research Center, Califórnia.
<i>Acessível pelo endereço</i>	http://www.qbic.almaden.ibm.com
<i>Histórico</i>	Foi o primeiro CBIR comercial.
<i>Descritores de Características</i>	Cor, forma e textura
<i>Elaboração de Consultas</i>	Baseado em imagens exemplo, inclusive permitindo a inserção de rascunhos.
<i>Medidas e Modelo de Similaridade</i>	Distância Euclidiana para a cor média, para textura e forma e distância quadrática do histograma para histogramas de cores (alto custo computacional), computando um somatório final que representa o casamento entre a imagem consulta e componentes da coleção.
<i>Apresentação de Resultados</i>	Ordenado por relevância, apresentando o escore final opcionalmente.
<i>Realimentação de Relevância</i>	Presente.
<i>Principais Aplicações</i>	Consulta em uma base de marcas, para questões de segurança e proteção dos direitos sobre as marcas. <i>The State Hermitage Museum: Digital Collection - Powered by IBM:</i> eleito recentemente o melhor site na Rússia – realiza pesquisas em imagens de obras de artes famosas.

g. Outros sistemas *CBIR* encontrados na Internet



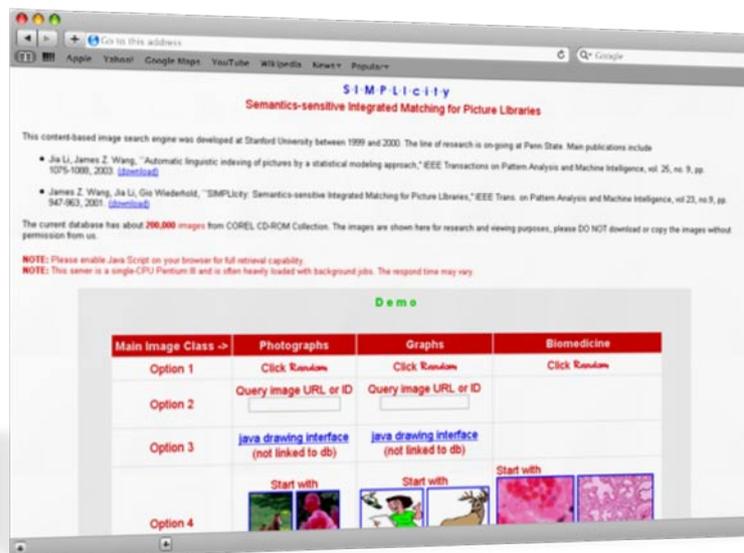
CIREs

Desenvolvido por

University of Texas, Austin.

Acessível pelo endereço

<http://amazon.ece.utexas.edu/~qasim/research.htm>



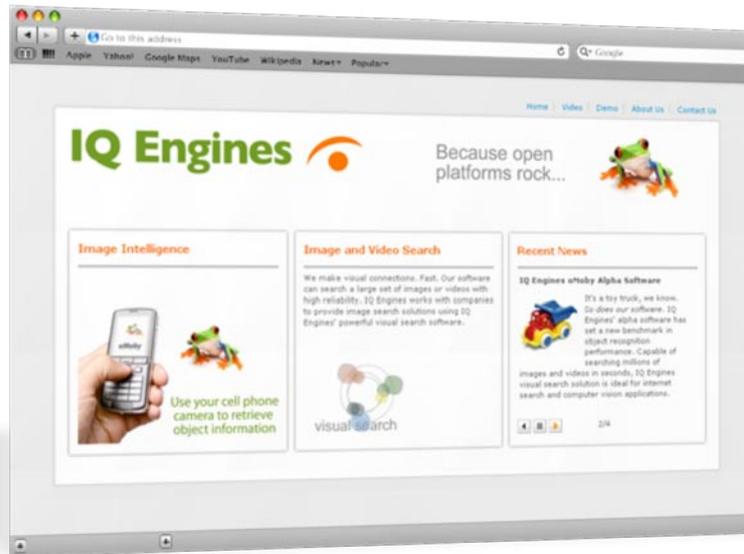
SIMPLiCity

Desenvolvido por

Penn State University

Acessível pelo endereço

http://wang14.ist.psu.edu/cgi-bin/zwang/regionsearch_show.cgi



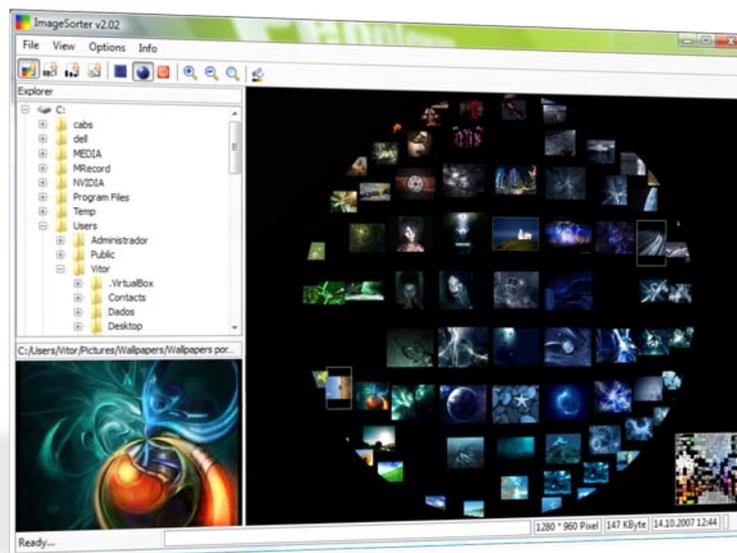
IQ Engines

Desenvolvido por

University of California, Berkeley e Davis

Acessível pelo endereço

<http://www.iqengines.com>



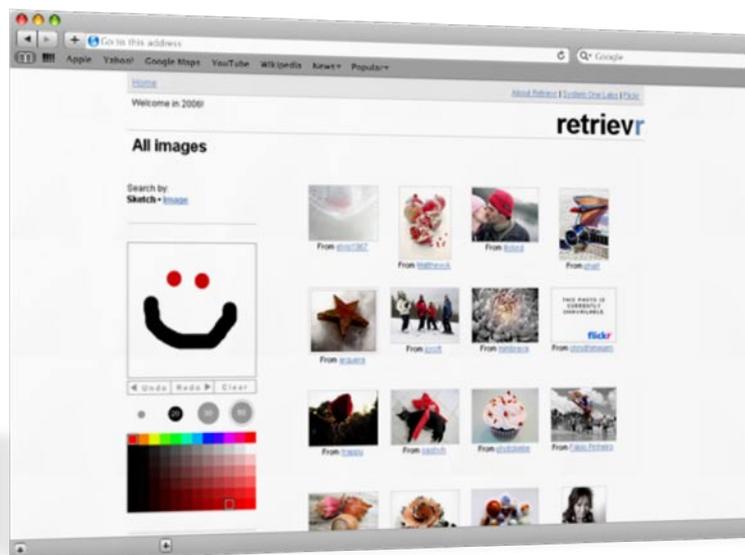
ImageSorter

Desenvolvido por

Kompetenzzentrum Mensch-Maschine-Kommunikation

Acessível pelo endereço

<http://mmk1.f4.fhtw-berlin.de/Projekte/ImageSorter>



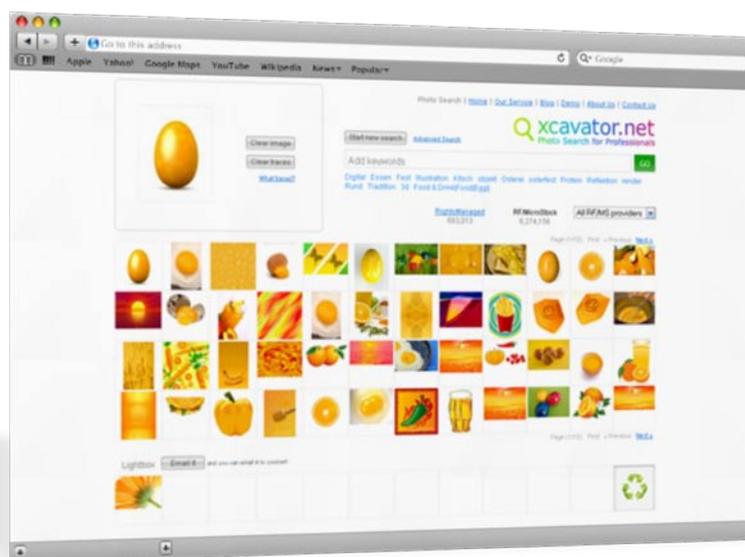
Retrievr

Desenvolvido por

System One Labs

Acessível pelo endereço

<http://labs.systemone.at/retrievr>



Xcavator

Desenvolvido por

CogniSign

Acessível pelo endereço

<http://www.xcavator.net>



Picitup

Desenvolvido por

Picitup, Israel

Acessível pelo endereço

<http://www.picitup.com/picitup/index.jsp>



oSkope Visual Search

Desenvolvido por

oSkope Media, Zurique

Acessível pelo endereço

<http://www.oskope.com>

3.9 Considerações Finais deste Capítulo

Como descrito em detalhe ao longo deste capítulo, em sistemas CBIR, imagens são comparadas com base na similaridade entre a requisição do usuário (*query*) e as características extraídas da coleção de imagens.

Também foram apresentados alguns conceitos fundamentais como o domínio de consulta, concepção e estratégia de pesquisa. Os principais descritores de características de imagens para cor, textura e forma encontrados na literatura foram abordados e as medidas de similaridade elucidadas, assim como os cálculos de distância entre vetores de características.

Tópicos essenciais para a consecução de um sistema CBIR como a realimentação de relevância, estratégias de visualização e apresentação de resultados foram abordados, finalizando o conteúdo deste capítulo com a análise das medidas de desempenho e a exposição de 14 exemplos práticos de implementações de sistemas de recuperação de imagens baseado em conteúdo de cunho acadêmico e comercial.

O Capítulo 4 abrange temas, conceitos e fundamentos relacionados aos algoritmos genéticos. Os componentes e principais decisões de um projeto de desenvolvimento de um AG são analisados e seus problemas práticos apresentados. Também é abordado neste próximo capítulo a aplicação de um algoritmo genético para a implementação da realimentação de relevância em sistemas CBIR.

ALGORITMOS GENÉTICOS

4.1 Introdução

Algoritmos Genéticos são métodos computacionais de busca, otimização e aprendizado de máquina, baseados nos mecanismos biológicos da evolução natural e da genética. Em sua abordagem, uma população de possíveis soluções para um problema, mesmo de complexidade elevada e de difícil compreensão humana, evolui de acordo com operadores probabilísticos concebidos a partir de metáforas biológicas, de modo que há uma tendência de que, na média, os indivíduos representem soluções cada vez melhores, à medida que o processo evolutivo continua [46].

Os algoritmos genéticos têm demonstrado ser excelentes ferramentas para a solução de problemas de otimização e busca, com diversas aplicações publicadas na literatura com resultados expressivos. Sua utilização no contexto de sistemas CBIR de recuperação de imagens por conteúdo visual é recente, porém demonstra um potencial de aplicação interessante.

Neste capítulo, os principais fundamentos dos algoritmos genéticos são apresentados, seus componentes são descritos e analisados, alguns problemas práticos em sua utilização são

discutidos e a aplicação desses algoritmos no que tange a realimentação de relevância em sistemas CBIR é avaliada.

4.2 Fundamentos

Desde os primórdios, os seres humanos percebem, experimentam e aprendem com o seu ambiente a todo o momento, buscando na natureza inspirações para a criação de tecnologias que aprimorem sua qualidade de vida. Alguns exemplos de invenções baseadas na observação do meio ambiente seriam aviões, submarinos e sonares, criados a partir da análise e estudo de pássaros, peixes e morcegos, respectivamente [47].

A Inteligência Artificial [48], campo de estudo e elaboração de técnicas e metodologias que emulam aspectos da inteligência humana em máquinas, é um grande exemplo de ciência que se inspira fortemente em projetos naturais para seu desenvolvimento. Tome como exemplo as *Redes Neurais Artificiais*, modelos baseados no neurônio e cérebro humanos, tanto em sua constituição física, a partir de conexões sinápticas, quanto em seu funcionamento. A observação de colônias de formigas, cupins e abelhas e seu comportamento na realização de tarefas gerou uma técnica computacional desenhada principalmente para casos de otimização, denominada “colônia de insetos”. Diversas outras abordagens se inspiram na natureza, e os algoritmos genéticos também se enquadram nesse grupo.

Os AGs, como são conhecidos na comunidade científica, foram inicialmente propostos por Holland [49] em 1975, a partir de seus estudos sobre a teoria da evolução das espécies de Darwin. Segundo Holland, a evolução natural é como a aprendizagem, ou seja, um processo de adaptação, sendo que elas apenas se diferem no período de tempo em que ocorrem, uma podendo levar inúmeras gerações, enquanto a outra se passa em um instante.

A seleção natural é um processo comumente observado na natureza. Quando há escassez de recursos em uma determinada população, sejam eles alimento, espaço ou outro recurso essencial, a competição é imposta naturalmente, e assim, os indivíduos mais preparados têm maior chance de sobrevivência. Isso ocorre pois as características dos indivíduos são passadas entre as gerações, e conseqüentemente, certos indivíduos pertencentes àquela população podem ter herdado características que os fazem mais propensos à adaptação. Neste caso, os seres sobreviventes possuirão certas características genéticas mais acentuadas do que gerações anteriores.

Para Holland, a evolução natural poderia ser implementada em máquinas desde que fosse desenvolvido um modelo matemático que a suportasse. Dado um problema de difícil resolução, a população inicial seria composta por possíveis soluções selecionadas ao acaso. Dessa forma, a evolução da população por várias gerações fora transcrita como um processo iterativo de melhoramento das soluções do referido problema, permitindo a sobrevivência dos mais aptos a partir de representações probabilísticas de seleção associadas a uma medida heurística que avaliasse a aptidão dos indivíduos, ou seja, da qualidade da solução que estes representam.

Os AGs se diferenciam de outros procedimentos de busca e otimização em quatro princípios básicos:

- Muitos algoritmos genéticos operam em um espaço de soluções codificadas, não diretamente no espaço de busca;
- Os AGs operam em um conjunto de pontos (população de indivíduos), não a partir de um único ponto isolado;
- Os algoritmos genéticos não necessitam de conhecimentos auxiliares além da representação das soluções e do cálculo estimado de suas qualidades; e
- AGs utilizam regras de transição probabilísticas, não determinísticas.

Portanto, a utilização de um algoritmo genético como ferramenta auxiliar em um sistema de recuperação de imagens por conteúdo demonstra ser interessante, e de fato, traz resultados surpreendentes, se devidamente projetado [50].

4.3 Componentes

Genericamente, um algoritmo genético segue um fluxograma básico, apresentado a seguir, na Figura 4.1 [51]. Dado um problema de busca ou otimização τ e eleita uma representação das possíveis soluções desse problema, gera-se uma população inicial de forma aleatória contendo n indivíduos.

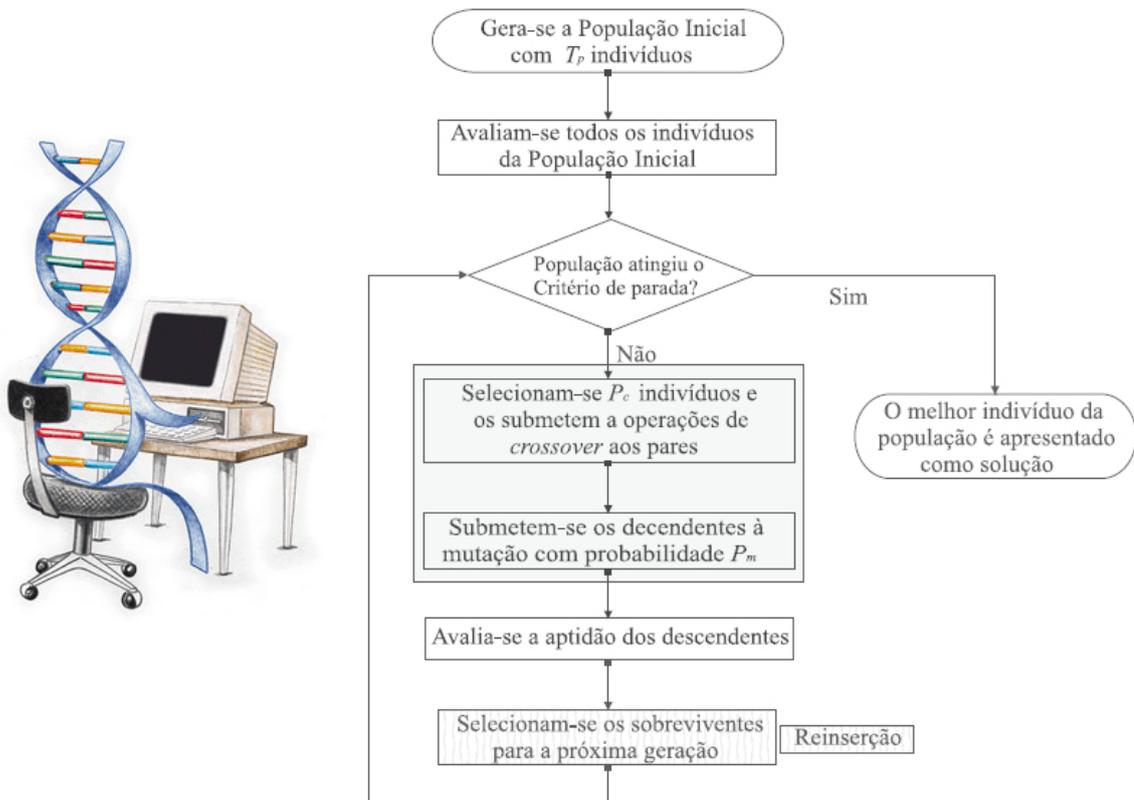


Figura 4.1. Ciclo de execução básico dos AGs.

Os indivíduos dessa população são avaliados segundo sua aptidão e caso algum desses atinja um nível satisfatório pré-estabelecido e considerado uma “boa solução”, o algoritmo é interrompido e o melhor indivíduo é eleito como a solução para o problema z . Ao contrário, integrantes da população são selecionados para a realização de algumas operações genéticas conhecidas como *cruzamento* e *mutação*, concebendo novos indivíduos.

Esses descendentes serão avaliados segundo sua aptidão e juntamente com seus pais, serão classificados de maneira que os n indivíduos mais aptos sobreviverão para a próxima geração, fechando o ciclo de execução do AG.

A seguir, os fundamentos e etapas de um algoritmo genético são apresentados e conceituados.

4.3.1 Cromossomos

O **espaço de busca** é um conjunto, espaço ou região que compreende as soluções possíveis ou viáveis do problema e a primeira etapa para a aplicação de um AG para solução de um problema é a escolha de uma representação matemática finita das possíveis soluções, pertencentes ao espaço de busca.

As primeiras **representações** utilizadas em algoritmos genéticos compreendiam o alfabeto binário $\{0, 1\}$, contudo, representações baseadas em caracteres, números inteiros e reais são amplamente utilizadas, de forma que, segundo Silva et al. [50], recentes pesquisas demonstram que alfabetos de alta-cardinalidade como números reais são mais efetivos na solução de problemas utilizando AGs, devido ao seu maior poder de expressão.

À essas representações das soluções, dá-se o nome de **indivíduos**, estes, pertencentes a uma **população**, um conjunto de elementos pertencentes ao espaço de busca.

Outra nomenclatura amplamente observada na literatura é a utilização do termo **cromossomo** para identificar as representações, os indivíduos. Desta forma, a linguagem

utilizada nessa técnica de computação evolutiva passa a ser bastante semelhante a da Genética, já que os AGs têm forte inspiração biológica.

Apesar de muitos trabalhos terem sido desenvolvidos nesse campo de estudo, não há uma receita para uma boa codificação, ou representação de um cromossomo. Para que bons resultados práticos sejam alcançados, os AGs devem ser desenvolvidos sob medida para a solução do problema em questão e experimentos devem ser conduzidos de forma a obter-se um desempenho aprimorado.

Portanto, a melhor estratégia a ser empregada é a utilização de conhecimentos específicos do domínio do problema em questão para a codificação de uma representação cromossômica adequada.

4.3.2 População

A população compreende todos os indivíduos da respectiva geração. A formação de uma **população inicial** é imprescindível para a execução do algoritmo, pois é a partir dela que todo o processo se passa.

Uma abordagem bastante utilizada para a seleção dos cromossomos pertencentes à população inicial é a aplicação de processos heurísticos ou a utilização de uma seleção aleatória entre os elementos do espaço de busca.

Um quesito importante da população inicial é sua cobertura em relação ao espaço de busca. Quanto mais “espalhados” os indivíduos da população inicial estiverem em relação ao espaço de busca, melhores são as possibilidades de que o algoritmo genético evolua para uma solução próxima, senão, a ótima. Caso os cromossomos sejam posicionados muito próximos entre si, geralmente observa-se uma evolução à pontos ótimos locais, causando uma repentina cessão na evolução da aptidão, de forma que os indivíduos tendem a se tornarem idênticos após consecutivas gerações.

Fazendo uma analogia à natureza, a evolução não ocorre sem a presença da diversidade [49], pois as diferentes características e graus de aptidão distintos são necessários para que possa ocorrer a seleção natural. Contrariamente, também é observado o caso de espécimes em risco de extinção, que devido ao pequeno número de indivíduos, mesmo com os esforços de organizações para a realização da proteção desses animais a partir da procriação em cativeiro, nota-se a implantação de problemas anteriormente não reconhecidos, devido ao alto grau de similaridade genética entre os animais sobreviventes.

4.3.3 Medida de Aptidão

A medida de aptidão indica quão bem adaptado está cada indivíduo da população em relação ao ambiente. Alternativamente, pode ser obtida matematicamente por uma função, um simulador ou um especialista humano.

A especificação de uma medida de aptidão apropriada se faz crucial para a consecução de um AG, pois é essencial que essa métrica seja altamente representativa do problema em questão, de forma a diferenciar na proporção adequada, as más soluções das boas.

Caso exista pouca precisão na avaliação, soluções de grande qualidade podem ser perdidas durante a execução do AG, por serem classificadas equivocadamente, e conseqüentemente, o algoritmo demandará maior período de tempo explorando o espaço de busca com soluções pouco promissoras, encontrando por fim, uma solução que não represente boa qualidade [46].

Segundo Michalewicz [52], alguns fatores devem ser considerados na elaboração de uma medida de aptidão:

- Determinação se o problema é relacionado à maximização ou minimização;
- Definição se os ambientes são determinísticos ou não;

- A possibilidade da medida de aptidão ser alterada dinamicamente durante a execução do algoritmo;
- A avaliação da aptidão deve consumir pouco tempo de processamento, de forma a viabilizar a simplificação da medida de aptidão em alguns casos;
- A medida de aptidão deve alocar diferentes valores de aptidão para indivíduos distintos, permitindo uma melhor eficiência na execução dos operadores de seleção;
- Deve contemplar e considerar as restrições do problema ao máximo; e
- Pode incorporar múltiplos objetivos.

Geralmente o maior custo computacional de um algoritmo genético ocorre a partir do cálculo da medida de aptidão, uma vez que esta avalia todos os indivíduos, geração a geração. No intuito de minimizar essa carga computacional, alguns cuidados especiais foram propostos por Haupt et al. [53], como:

- Não gerar indivíduos idênticos na população inicial;
- Garantir que a reprodução não gere indivíduos idênticos; e
- Criar uma memória para os AGs, descartando os indivíduos gerados anteriormente e considerados inaptos ou “menos aptos”.

Na prática, observa-se que desses critérios, apenas o primeiro é amplamente levado em conta na elaboração dos algoritmos genéticos [54].

4.3.4 Evolução

Geralmente os AGs partem de soluções selecionadas randomicamente no início da busca iterativa e conforme evolui, valores particulares para cada gene podem começar a prevalecer. Desta forma, assim que a variância dos valores de aptidão decresce, a população converge.

A análise da convergência pode ser realizada por meio de diversas medidas. Goldberg [46] utilizou o desvio padrão dos valores de aptidão de cada indivíduo para análise. Outras medidas para a análise da convergência, ou evolução de um AG, são o melhor indivíduo da geração, o pior indivíduo da população ou a média de aptidão daquela geração, como ilustra a Figura 4.2.

Idealmente, a população converge para a solução ótima. Contudo, várias são as situações práticas onde esse fato não se confirma, geralmente devido aos problemas da convergência prematura ou término lento, tratados na seção 4.4 deste capítulo.

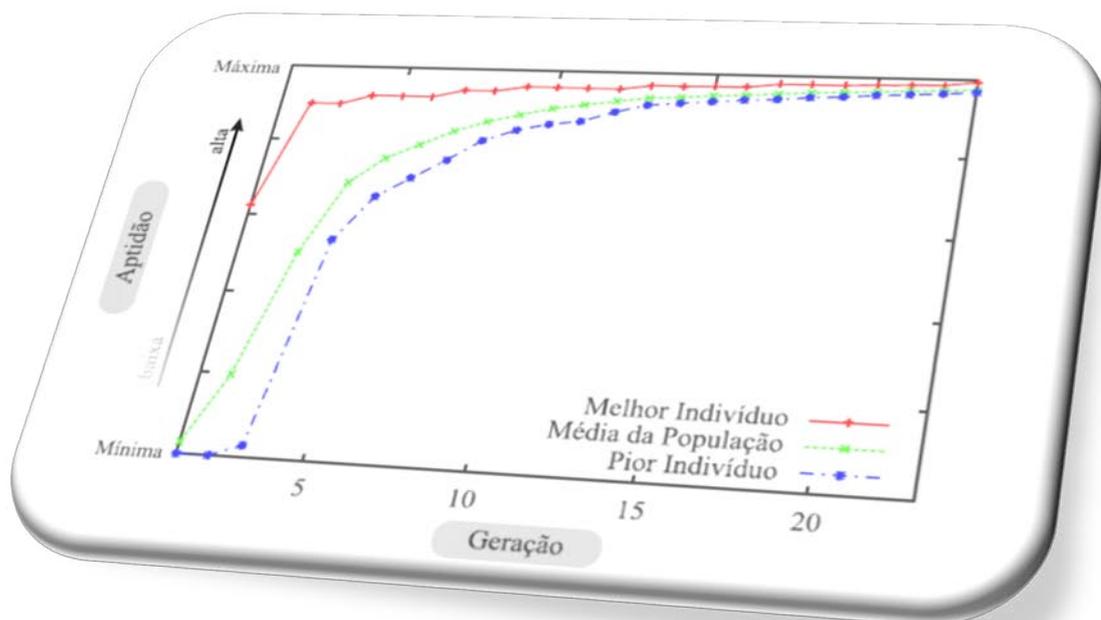


Figura 4.2. Evolução de um algoritmo genético.

4.3.5 Seleção

A operação genética da seleção é realizada após a medição de aptidão dos indivíduos, desempenhando o papel da seleção natural na evolução biológica, simulando uma seleção prioritária onde os organismos melhores adaptados sobrevivem e podem se reproduzir.

Existem várias formas de realizar a seleção dos indivíduos de uma população, dependendo do método de seleção adotado. No entanto, é certo que os indivíduos melhores adaptados terão uma maior probabilidade de sobrevivência e reprodução do que os demais, menos adaptados. Alguns operadores de seleção são apresentados a seguir:

a. Seleção estocástica com reposição

O método de seleção estocástica com reposição, proposto inicialmente por Holland [49] e também conhecido como *Método de Seleção por Roleta*, é o mais utilizado em implementações de algoritmos genéticos. A cada indivíduo da população corrente é atribuído uma fatia de uma roleta imaginária, sendo o tamanho desta fatia proporcional a aptidão do indivíduo, conforme ilustra a Figura 4.3.

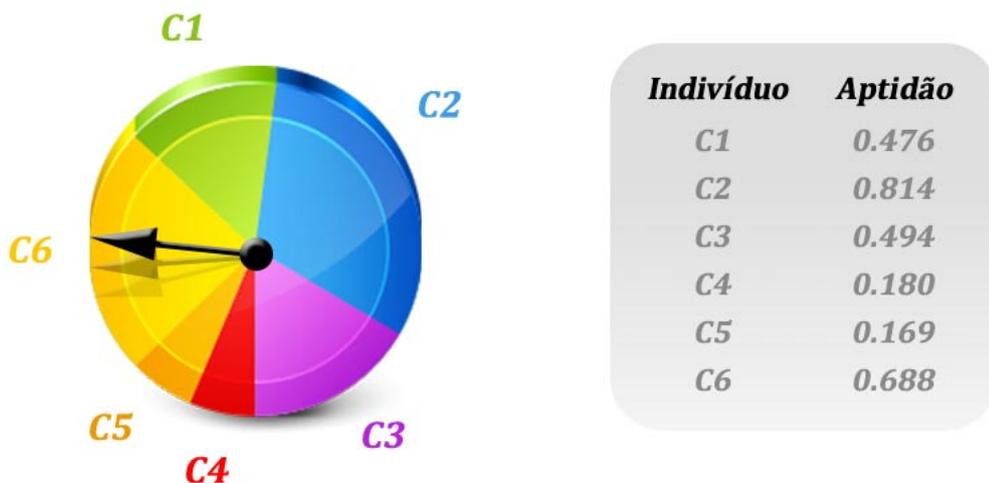


Figura 4.3. Exemplo de uma roleta de seleção.

Simula-se um giro de um ponteiro nessa roleta para a seleção de um indivíduo, de forma que se f_j é a aptidão do cromossomo j na população corrente, a probabilidade p_j do indivíduo ser selecionado é obtida pela Equação (4.1).

$$p_j = \frac{f_j}{\sum_{k=1}^{T_p} f_k} \quad (4.1)$$

onde

T_p é o número de indivíduos na população e

f_k é a aptidão do k -ésimo indivíduo da população.

b. Seleção por torneio simples

Na seleção por torneio simples ocorre a promoção de um torneio entre um grupo de N indivíduos tomados aleatoriamente na população. O indivíduo de maior aptidão do grupo é selecionado, enquanto os demais são descartados. Nesse método a pressão seletiva pode ser controlada pelo tamanho dos grupos do torneio.

c. Seleção por torneio estocástico

A seleção por torneio estocástico é similar ao torneio simples, sendo diferente no aspecto de que os indivíduos dos grupos são selecionados pelo método da roleta ao invés de serem aleatoriamente tomados da população corrente.

d. Seleção por truncamento

Na seleção por truncamento, um subconjunto dos melhores indivíduos é escolhido e seus componentes selecionados com a mesma probabilidade.

e. Seleção por ordenação

Nesse modelo de seleção, os indivíduos são ordenados segundo suas aptidões em um ranking crescente, ou seja, do pior para o melhor. A seguir, cada indivíduo j recebe uma probabilidade de seleção p_j tomada de uma distribuição aplicada às posições dos cromossomos no ranking, seguindo a restrição que

$$\sum p_j = 1 \quad (4.2)$$

As distribuições mais utilizadas são, respectivamente, a linear e a exponencial, apresentadas nas Equações (4.3) e (4.4), onde $pos(j)$ é a posição do indivíduo j no ranking.

$$p_j = a \, pos(j) + b, \quad a > 0 \quad (4.3)$$

$$p_j = a^{b \, pos(j)+c}, \quad a > 0 \text{ e } b > 0 \quad (4.4)$$

f. Seleção elitista

A seleção elitista, ou simplesmente *Elitismo*, seleciona os N melhores indivíduos da população corrente para fazerem parte da próxima geração, uma vez que, probabilisticamente, existe a chance desses não serem selecionados. Esse modelo é normalmente acoplado a outros métodos de seleção na tentativa de se aumentar a velocidade de convergência do AG.

4.3.6 Reprodução

A reprodução é um processo inspirado na recombinação biológica, ou seja, na troca de material genético entre os pais na geração de filhos. Esse processo também é referido na literatura como cruzamento, ou *crossover*.

Após serem escolhidos a partir dos métodos de seleção apresentados na subseção anterior, os indivíduos são organizados em pares para que o cruzamento seja efetivado, dando origem a filhos. Um par selecionado para o *crossover* (pais) originará um par de descendentes (filhos).

A frequência de recombinações é controlada pela probabilidade de cruzamento p_c , ou taxa de *crossover*, que denota o percentual da população que será submetida ao processo de reprodução, a cada geração.

O modo em que a operação de reprodução é realizada depende do domínio e das restrições do problema em questão. Nesta categoria, existem três tipos de operações de cruzamento: simples, múltiplo e uniforme.

a. *Crossover* simples

No *crossover* simples, um ponto dos cromossomos é selecionado ao acaso onde ambos os pais serão “cortados” de maneira que a primeira parte do pai A é combinada com a segunda parte do pai B, formando um dos filhos. O mesmo ocorre na formação do segundo filho, que recebe a primeira parte do pai B combinada a segunda parte do pai A. A Figura 4.4 apresenta um exemplo dessa operação genética.

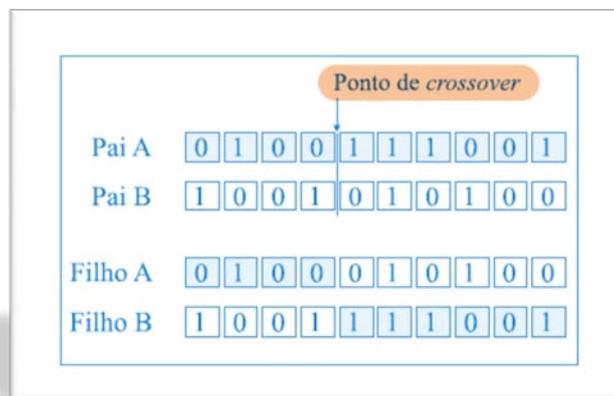


Figura 4.4. *Crossover* simples.

b. *Crossover múltiplo*

Similar ao operador de cruzamento simples, o crossover múltiplo utiliza dois ou mais pontos de cruzamento, selecionados aleatoriamente. Os segmentos de genes, definidos pelos pontos de corte, são alternadamente tomados para a geração dos filhos, conforme ilustra a Figura 4.5.

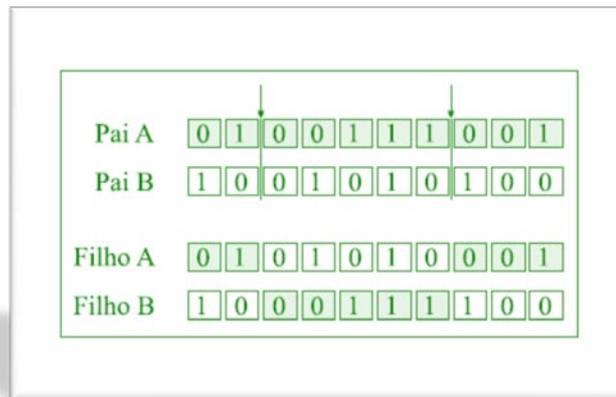


Figura 4.5. Crossover múltiplo.

c. *Crossover uniforme*

Esse operador genético de reprodução é caracterizado por realizar o cruzamento a partir de uma máscara de mesmo tamanho do cromossomo contendo informações de qual pai fornecerá cada gene ao primeiro filho, sendo que o segundo é gerado pelo complemento da máscara. A figura 4.6 apresenta um exemplo.

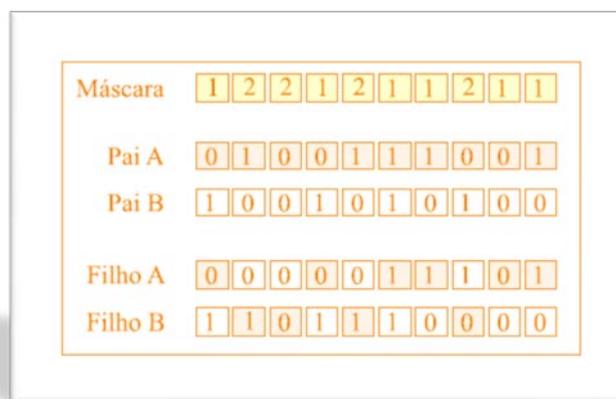


Figura 4.6. Crossover uniforme.

Em alguns domínios de problema, a reprodução não pode gerar genes repetidos, como é o exemplo do problema do caixeiro viajante, onde conhecidas as cidades a serem visitadas, busca-se um trajeto que passe por todas elas uma única vez de maneira a formar o menor percurso possível.

Naturalmente, uma representação para esse problema seria um cromossomo de tamanho igual ao número de cidades a serem visitadas contendo algarismos representativos de cada uma das cidades como conteúdo genético, sendo que sua ordem no cromossomo representa a ordem de visitação.

Como a solução do problema requer uma única visita a cada cidade, necessita-se de uma operação de cruzamento que não gere genes repetidos. Encontra-se na literatura operadores de *crossover* que levam em consideração alguns desses requisitos de cruzamento, baseados em problemas particulares e de domínios específicos. Dois dos mais populares são apresentados a seguir.

d. *Crossover* PMX

Com sua sigla proveniente do inglês (*Partially Matched Crossover*) e também conhecido como cruzamento parcial, o *crossover* PMX ocorre a partir da seleção aleatória de dois pontos de cruzamento, realizando o intercâmbio genético entre os pais, conforme o operador de cruzamento múltiplo com dois pontos.

Finalizado o processo, os genes repetidos externos aos pontos de corte são substituídos por seus alelos no outro parceiro. Na biologia, alelos referem-se às variantes de um mesmo gene. Neste contexto, o termo alelo é utilizado para denotar a mesma posição genética nos cromossomos pais. A Figura 4.7 ilustra este operador genético de reprodução.

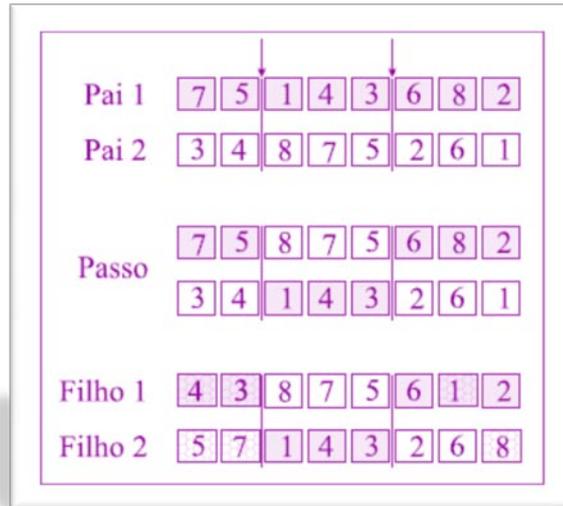


Figura 4.7. Crossover PMX.

e. Crossover cíclico

A operação *crossover* cíclico é realizada a partir de uma única posição de forma que os genes são trocados entre os parceiros até formar um ciclo. Os demais genes fora do ciclo não são trocados. A Figura 4.8 exemplifica essa operação. Começando da primeira posição de cromossomos, troca-se 7 por 3, 3 por 5, 5 por 4 e 4 por 7, completando o ciclo 7–3–5–4–7.



Figura 4.8. Crossover cíclico.

4.3.7 Mutação

A mutação é uma operação que simplesmente modifica randomicamente alguma característica genética de um indivíduo. Esse operador é importante pois permite a criação de novas características que não existiam ou que apareciam de forma irrelevante na população em análise.

A mutação introduz diversidade genética à população, de forma a assegurar que a probabilidade de se chegar a qualquer ponto do espaço de busca não será nula. Contudo, sua utilização demasiada pode transfigurar as características correntes dos indivíduos na geração, tornando o AG puramente probabilístico. Por isso, geralmente, o operador de mutação é aplicado aos indivíduos com uma taxa de mutação p_m pequena.

A mutação, bem como a reprodução, deve ser realizada de forma a não violar as restrições do problema. Alguns procedimentos de mutação são propostos na literatura como a substituição de um gene por outro gerado ao acaso, perturbação de genes, permutação de um gene, etc. A Figura 4.9 ilustra a mutação simples.

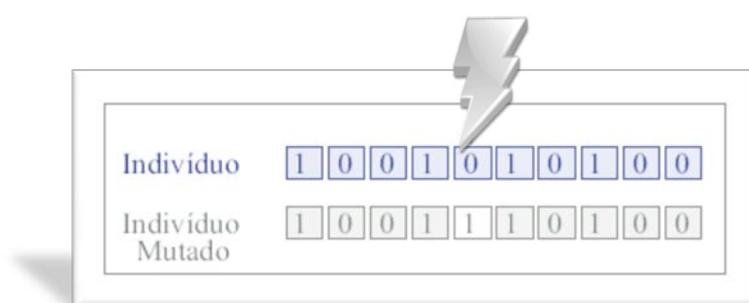


Figura 4.9. Mutação simples.

4.3.8 Reinservação

Após o processo de reprodução (cruzamento e mutação) e avaliadas as aptidões dos descendentes, são aplicados mecanismos de seleção que elegerão os sobreviventes para a próxima geração. Os principais métodos de reinservação são:

a. Reinservação pura

A reinservação pura substitui completamente a população pelos filhos gerados. Essa estratégia é geralmente acompanhada pelo elitismo.

b. Elitismo

No elitismo, uma parte da população é mantida para a próxima geração, no intuito da manutenção das melhores soluções. Geralmente o elitismo é um procedimento que acompanha as reinserações pura e uniforme.

c. Reinservação uniforme

Na reinservação uniforme, o número de indivíduos da geração anterior é mantido a partir da seleção de T_p cromossomos, dentre pais e filhos, a partir de qualquer um dos métodos de seleção já apresentados.

d. Reinservação baseada na aptidão

Na reinservação baseada na aptidão, todos os indivíduos (pais e filhos) são ordenados conforme seus valores de aptidão e os T_p melhores são selecionados para a próxima geração.

4.3.9 Condições de Parada

Em se tratando de problemas de otimização, o ideal é que o algoritmo genético pare assim que a solução ótima é descoberta. No entanto, em problemas práticos, muitas são as ocasiões onde não se pode afirmar que o AG encontra a solução ótima em tempo viável. Como consequência, alguns outros critérios de parada são aplicados, como:

- Excedido o número máximo de gerações (iterações) pré-estabelecido;
- Excedido o tempo máximo de processamento pré-estabelecido;
- Encontrado um indivíduo com aptidão maior ou igual a um limiar pré-definido; e
- Estagnação da população ou do melhor indivíduo após um determinado número de gerações.

Nesses casos, o algoritmo genético é encerrado e o melhor indivíduo é apresentado como a melhor solução encontrada para o problema em questão. Observa-se que uma próxima execução do mesmo AG para a solução do mesmo problema pode resultar em uma solução mais ou menos adequada, devido ao caráter aleatório dessa técnica de inteligência artificial. Entretanto, é esperado que o algoritmo genético, se bem implementado, apresente ao menos uma solução razoável para o problema tratado.

4.3.10 Parâmetros de Controle

Os algoritmos genéticos têm seu funcionamento baseado em uma série de parâmetros que exercem grande influência sobre seu comportamento e desempenho. São eles:

- Tamanho da população, T_p ;
- Método de seleção;

- Probabilidade de cruzamento, p_c ;
- Probabilidade de mutação, p_m ;
- Reinserção; e
- Critério de parada.

A intuição e bom senso são normalmente utilizados na escolha desses parâmetros. Alguns autores aconselham que estes devem ser determinados com base em experimentações enquanto outros, motivados pela análise de problemas de convergência, acreditam que a variação dinâmica desses parâmetros faz com que os AGs apresentem melhor desempenho. Nada obstante, a consideração das seguintes observações faz-se fundamental para uma melhor implementação de um algoritmo genético.

- Uma população muito pequena implica em pouca cobertura do espaço de busca e conseqüentemente, maiores probabilidades de convergência prematura, como apontado na seção 4.4 deste capítulo. Em contrapartida, uma população muito grande possibilita uma ampla cobertura do espaço de busca, exigindo elevada carga computacional e possível lentidão;
- Quanto maior a probabilidade de cruzamento, mais rapidamente novas estruturas serão introduzidas na população. Caso essa for muito alta, estruturas com boas aptidões poderão ser deformadas ou até destruídas mais rapidamente do que a capacidade da seleção em mantê-las. Assim, normalmente são utilizadas operações de reinserção por elitismo para garantir que os melhores indivíduos não sejam destruídos pelas operações de *crossover*; e

- Mutações previnem que a busca fique estagnada em regiões pouco promissoras do espaço de busca (máximos locais), possibilitando que qualquer ponto do espaço tenha uma probabilidade não nula de ser atingido. Entretanto, uma taxa muito alta de mutação a torna uma busca essencialmente aleatória.

4.4 Problemas Práticos

Dentre os problemas geralmente encontrados por pesquisadores da área de AGs, três são amplamente conhecidos – a convergência prematura, a convergência lenta e o tempo de processamento elevado.

A **convergência prematura**, também conhecida como decepção, ocorre quando genes de alguns poucos indivíduos com relativo alto valor de aptidão rapidamente dominam a população causando a convergência para um máximo local.

Uma vez que a população tenha convergido, a habilidade do AG de continuar pesquisando por melhores soluções fica bastante restrita. Esse tipo de convergência normalmente ocorre devido à populações reduzidas ou má distribuídas no espaço de busca.

Já a **convergência lenta**, é o problema oposto. Depois de uma série de gerações, a população demonstra certa convergência, porém, uma solução razoável não fora encontrada. Embora a aptidão média possa ser relativamente alta e a diferença entre o melhor e o pior indivíduo ser pequena, não há variância significativa para que uma melhor solução seja localizada.

Grande parte das aplicações de algoritmos genéticos absorve um **elevado tempo de processamento** e muitas vezes o problema a ser solucionado é complexo, tornando a avaliação de todos os indivíduos da população uma tarefa delongada. Em outras ocasiões, o

espaço de busca tem dimensões elevadas, exigindo maiores populações e conseqüentes extensões no tempo de resposta e convergência.

Uma excelente solução para esse problema é a utilização de processamento paralelo na execução dos AGs, devido à sua característica intrinsecamente paralela, tornando essa abordagem bastante promissora. Cantú-Paz et al. [55], dentre outros autores, propõem a implementação de AGs paralelos no intuito de explorar essas propriedades.

4.5 AGs e Realimentação de Relevância

A principal função de um mecanismo de realimentação de relevância em um sistema de recuperação da informação é capturar os requerimentos do usuário a fim de realizar uma nova busca que melhor satisfaça suas necessidades, sendo que esses mecanismos podem atuar, fundamentalmente, de duas maneiras distintas – modificando a representação da consulta do usuário ou alterando a medida de similaridade utilizada para ordenar documentos da coleção de referência.

Existem diversas abordagens e ferramentas empregadas para a tarefa de ajuste de pesos – abordagens estatísticas (como correlação e desvio padrão), redes neurais artificiais, redes bayesianas e computação evolutiva, entre outras.

Nesta seção é apresentada a abordagem proposta por Stejić et al. [56], utilizando-se de computação evolutiva por meio de algoritmos genéticos para inferir, a partir da realimentação de relevância do usuário, pesos que representam a significância de características como cor, textura e forma, além da importância das regiões das imagens, na tentativa de ajustar um modelo de similaridade de imagens às necessidades momentâneas de busca do usuário, de forma dinâmica.

Os trabalhos de Stejić et al. são relacionados a recuperação de imagens a partir do modelo de similaridade de padrões locais (*Local Similarity Pattern – LSP*), que se baseia na comparação de regiões correspondentes para obter o grau de similaridade final entre imagens. As imagens são divididas em regiões retangulares uniformes.

Supondo que as figuras em um banco de imagens sejam divididas em 9 regiões retangulares uniformes, conforme ilustrado na Figura 4.10, a similaridade entre as imagens A e B será obtida por:

$$Sim(A, B) = \frac{\sum_{i=1}^9 Sim(r_i(A), r_i(B))}{9} \quad (4.5)$$



Figura 4.10. Uma imagem dividida em nove regiões uniformes.

A partir de 2002, Stejić e colaboradores começaram a desenvolver sistemas de recuperação de imagens utilizando múltiplas características e regiões de imagens segundo o modelo LSP, além da aplicação da realimentação de relevância via algoritmos genéticos [56].

O primeiro sistema criado tinha como meta utilizar a realimentação do usuário para obter ótimas combinações de características para regiões de imagem. A similaridade de uma dada região era comparada utilizando características de cor, ou forma, ou textura, ou uma combinação de duas destas características.

A combinação de características utilizadas no modelo de similaridade entre imagens era inferida por um AG de forma que a precisão de recuperação fosse maximizada.

Em 2003, eles propuseram o modelo de similaridade *Weighted Local Similarity Pattern*, ou *WLSP* [57], como uma extensão do modelo *LSP*. No *WLSP*, além da combinação de características e da discretização dos pesos, também são considerados os pesos de regiões.

No mesmo ano, Stejić et al. desenvolveram duas novas extensões, uma com pesos contínuos não-negativos denominada *LSP-C+/WLSP-C+* e outra com pesos contínuos positivos e negativos denominada *LSP-C±/WLSP-C±*.

Um total de 67 operadores de agregação para a composição da medida de similaridade foram testados por Stejić et al. [58], com o desenvolvimento de um AG para a escolha dos operadores de agregação que melhor ajustassem o modelo de similaridade ao critério do usuário, manifestado durante a realimentação.

Entretanto, apesar de concluírem que a escolha de operadores de agregação tem grande importância na modelagem de similaridade conforme os critérios do usuário, a tentativa de adaptação dos operadores de agregação utilizando AG obteve resultados inferiores àqueles das extensões dos modelos *LSP* e *WLSP*.

Como os modelos *LSP-C±/WLSP-C±* foram os que obtiveram melhores resultados, maior enfoque será dado a essa abordagem, descrevendo sua técnica com maior detalhamento.

O modelo de similaridade de imagens utilizado é apresentado a seguir, pelas Equações (4.6) a (4.9):

$$S_I(q, i) = \sum_{r \in R} \omega_R(q, i, r) S_R(q, i, r) \quad (4.6)$$

$$\text{tal que} \quad \sum_{r \in R} |\omega_R(q, i, r)| = 1 \quad (4.7)$$

onde,

$$S_R(q, i, r) = \sum_{f \in F} \omega_F(q, i, r, f) S_F(q, i, r, f) \quad (4.8)$$

$$\text{tal que} \quad \sum_{r \in R} |\omega_F(q, i, r, f)| = 1 \quad (4.9)$$

A Equação (4.6) representa uma soma ponderada de similaridades de regiões sendo estas obtidas por uma soma ponderada das similaridades de características como cor, forma e textura, conforme a Equação (4.8).

Nessas equações q denota a imagem de consulta, i , uma determinada imagem do banco de imagens I, R , o conjunto de regiões r consideradas e $F = \{cor, forma, textura\}$.

Nos modelos de similaridade propostos por Stejić e colaboradores, as funções pesos, tanto de características, quanto de regiões têm intervalos contínuos onde,

$$\omega_F^{LSP-C\pm/WLSP-C\pm} \rightarrow [-1, 1] \quad (4.10)$$

A introdução de pesos negativos permite ao modelo expressar o fato de certa similaridade ser mais ou menos indesejável, combinado ao nível de relevância expresso por pesos positivos e a irrelevância expressa por um peso nulo.

O seguinte esquema de normalização é adotado para a obtenção de valores de similaridade de imagem no intervalo $[0, 1]$:

$$S_I(q, i) = \frac{1}{2} (S_I^*(q, i) + 1) \quad e \quad (4.11)$$

$$S_R(q, i, r) = \frac{1}{2} (S_R^*(q, i, r) + 1) \quad (4.12)$$

Nessa abordagem, as similaridades de características dependem exclusivamente da consulta e da coleção de referência. Contudo, as importâncias das características podem variar com a necessidade de busca do usuário.

A inferência de pesos fora desenvolvida a partir da aplicação de um algoritmo genético com as seguintes qualidades:

- **Codificação de indivíduos** – cada cromossomo representa a área da imagem partida uniformemente em n_R regiões. Cada gene de um indivíduo corresponde a uma região, sendo mapeada por um conjunto de quatro pesos: três referentes às suas características (cor, forma e textura) e o quarto representando a importância da região [57];
- **Medida de aptidão** – é obtida pela medida de precisão- R de recuperação. Para facilidade de nomenclatura, esta função será denominada como $F_{Stejić et al.}$.
- **Operadores genéticos** – foram experimentados os operadores de *crossover* uniforme, simples e múltiplo (de dois pontos) além de um total de sete operadores de mutação.

Os descritores de características de cor, forma e textura utilizados por Stejić et al. foram momentos de cor, histograma de direção de bordas e vizinhança de textura, respectivamente. As distâncias entre pares de vetores de características são computados pela distância Euclidiana para momentos de cor e distância em quadras para histograma de direção de bordas e vizinhança de textura.

As similaridades de características S_F são obtidas pela Equação (4.13).

$$S_F = 1 - D_F \quad (4.13)$$

onde

$D_F \in [0,1]$; e

D_F é a distância normalizada entre um par de descritores de características.

Considere uma seqüência de valores de similaridade entre a imagem consulta q e todas as imagens do banco de dados i_K ($K = 1, 2, \dots, n_l$) com respeito a região r e a característica f – $S_F(q,r,f) = [S_F(q,i_1,r,f), S_F(q,i_2,r,f), \dots, S_F(q,i_{n_l},r,f)]$ – e tendo computado a média $\mu(q,r,f)$ e o desvio padrão $\sigma(q,r,f)$ desta seqüência, a seqüência normalizada é calculada transformando cada elemento conforme as Equações (4.14) e (4.15), abaixo.

$$S'_F(q, i_k, r, f) = \frac{S_F(q, i_k, r, f) - \mu(q, r, f)}{4 \sigma(q, r, f)} \quad (4.14)$$

$$S''_F(q, i_k, r, f) = \frac{S'_F(q, i_k, r, f) + 1}{2} \quad (4.15)$$

Possíveis valores de S''_F fora do intervalo $[0, 1]$ são simplesmente mapeados para o limite do intervalo mais próximo.

Os modelos propostos por Stejić e colaboradores foram experimentalmente avaliados em seus trabalhos, diante de cinco bancos de imagens: *Vistex-60*, *Vistex-167*, *Brodatz-208*, *Corel-1000-A* e *Corel-1000-B*. Como medida de desempenho foi utilizada a precisão de recuperação, sendo que o sistema hardware-software utilizado não fora mencionado.

A Tabela 4.1 mostra a precisão de recuperação média obtida pelos métodos *LSP-C±* e *WLSP-C±* [57].

Tabela 4.1. Precisão de recuperação dos melhores métodos de Stejić et al.: *LSP-C±* e *WLSP-C±*.

Modelo	Precisão média para os bancos de imagens				
	V-60	V-167	B-208	C-1000-A	C-1000-B
<i>LSP-C±</i>	98	86	96	76	68
<i>WLSP-C±</i>	95	84	95	75	66

4.6 Considerações Finais deste Capítulo

A utilização de algoritmos genéticos no contexto de sistemas CBIR de recuperação de imagens por conteúdo visual é recente, porém demonstra um potencial de aplicação interessante, como abordado neste capítulo.

Além de introduzir os conceitos e fundamentos de AGs, o capítulo descreveu em detalhes os componentes primordiais para o desenvolvimento de um AG, apontando os principais problemas práticos encontrados.

Os trabalhos de Stejić e colaboradores foram apresentados e analisados, introduzindo uma utilização de AGs no contexto de recuperação de imagens por conteúdo a partir do desenvolvimento de uma técnica de realimentação de relevância.

O capítulo seguinte realiza um estudo detalhado sobre o sistema de recuperação de imagens por conteúdo denominado *VIPIGraf CBIR* e descreve as contribuições deste trabalho.

VIPIGRAF CBIR

5.1 Introdução

Este capítulo apresenta o *VIPIGraf CBIR*, sistema de recuperação de imagens por conteúdo inspirado nos trabalhos de Stejić e desenvolvido pelo grupo de estudo *VIPIGraf* da Universidade Federal de Uberlândia, o qual o autor é membro e colaborador. Uma visão geral é apresentada, descrevendo os componentes principais e funcionamento do sistema. Os descritores de características, as medidas e o modelo de similaridade, a realimentação de relevância e a utilização do algoritmo genético são descritos na seqüência.

Em seguida, é apresentada a nova abordagem de modelagem de similaridade e realimentação de relevância, multidimensional, utilizada no *VIPIGraf CBIR*. Algumas considerações finais são desenvolvidas ao final do capítulo.

5.2 Estrutura Geral

Primeiramente, no sistema *VIPIGraf CBIR*, todas as imagens, fotografias, desenhos, figuras, ilustrações e rascunhos arquivados em um banco de imagens digitais são analisados e

as características de cada uma dessas imagens, extraídas, produzindo como saída, estruturas contendo o código de identificação e os vetores de características das mesmas. Esses dados são armazenados no banco de dados de características, um arquivo texto contendo todas as informações. O processo de caracterização das imagens do banco é realizado a priori, antes da utilização do sistema pelos usuários. A Figura 5.1 apresenta o fluxograma básico de funcionamento do sistema *VIPIGraf CBIR*.

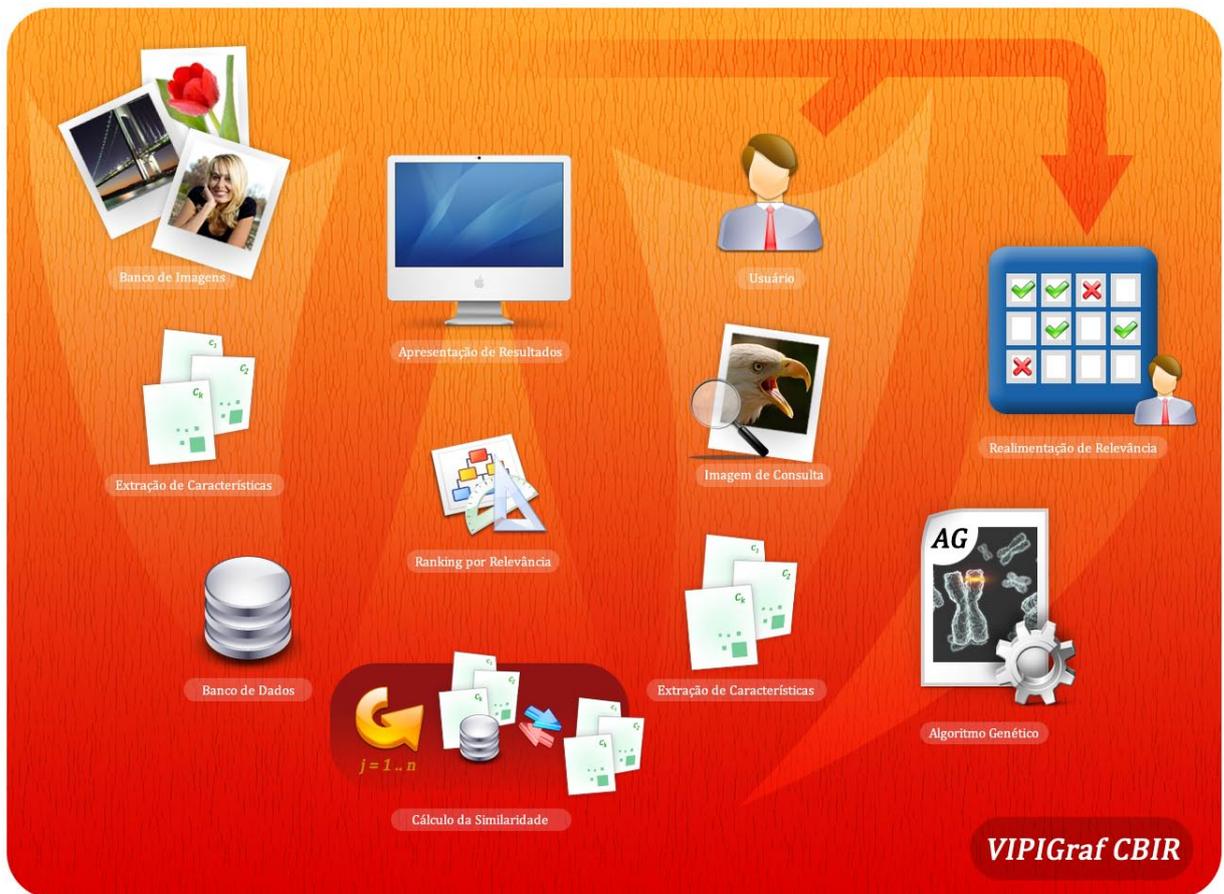


Figura 5.1. Fluxograma básico de funcionamento do CBIR.

Quando o usuário realiza uma busca ao sistema ao inserir uma imagem de consulta, vetores de características são extraídos dessa imagem pelo módulo extrator de características. Posteriormente, esses vetores da imagem de consulta são comparados às características contidas no banco de dados, em relação às medidas de similaridade e distância.

O módulo de similaridade retorna um valor de similaridade para cada imagem do banco em relação à imagem de consulta e de acordo com esses valores, um *ranking* é formatado em ordem decrescente de similaridade e então apresentado ao usuário como a primeira resposta do sistema.

Caso insatisfeito com o resultado da busca, o usuário pode realimentar o sistema indicando imagens relevantes à sua necessidade de informação, conforme seus requisitos. Assim, baseado nesta realimentação, o mecanismo de *feedback* de relevância realiza um ajuste à medida de similaridade, adequando-a aos critérios do usuário a partir da ponderação das regiões e características da imagem, possível graças à um algoritmo genético de código real, projetado sob-medida para o problema de inferência de pesos, visando maximizar a eficiência da recuperação, conforme os desejos manifestados pelo usuário.

5.3 Descritores de Características

Três características são aplicadas ao sistema *VIPIGraf CBIR* – cor, forma e textura – sendo que os descritores selecionados são os mesmos utilizados por Stejić et al. [56], ou seja, momentos de cores, histograma de direção de bordas e vizinhança de textura, respectivamente. Esses descritores de características foram descritos em detalhe no Capítulo 3 deste trabalho.

Os **momentos de cor** caracterizam as imagens em termos das suas distribuições de cores. Nesta aplicação, as imagens são convertidas do espaço de cor *RGB* (*Red, Green, Blue*) padrão das imagens digitais para o espaço de cor *HSI* (*Hue, Saturation, Intensity*), onde cada canal de cor é interpretado por três medidas estatísticas – *média* (E), *desvio padrão* (σ) e *inclinação* (S) – obtidas pelas Equações (3.1) a (3.3) do Capítulo 3 deste trabalho.

O descritor de cor final será composto por 9 valores reais, formatados da seguinte maneira – $(E_H, \sigma_H, S_H, E_S, \sigma_S, S_S, E_I, \sigma_I, S_I)$ – e a distância entre dois descritores é computada a partir da distância euclidiana, obtida pela Equação (3.38), também apresentada no Capítulo 3.

O **histograma de direção de bordas** modela a similaridade entre imagens em termos de suas distribuições das direções de bordas. Novamente, as imagens são convertidas para o espaço de cor *HSI* onde somente os canais *S* e *I* são utilizados na identificação das bordas. O canal *H* pode negligenciar bordas relevantes em regiões ou objetos com a mesma tonalidade de cor.

Operadores de Sobel de oito direções são aplicados sobre os dois canais da imagem e um descritor é definido representando as oito direções. Uma aplicação dos operadores de Sobel sobre uma imagem é apresentada na Figura 5.2. Cada pixel de cada canal de cor é analisado segundo as 8 máscaras de Sobel e o maior valor obtido define a direção da fronteira naquele ponto quando esse ultrapassar 35% para *S* ou 15% para *I* em relação à resposta máxima das máscaras nos respectivos canais.

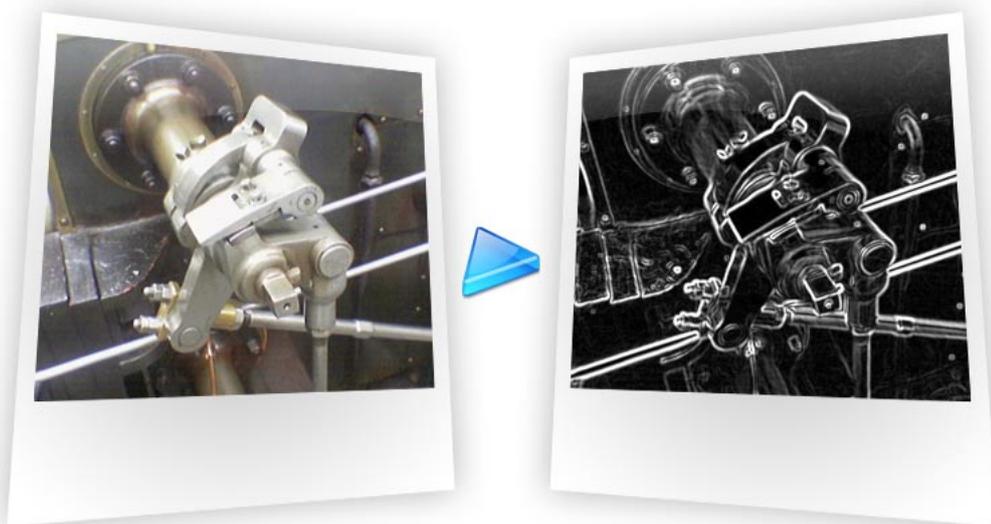


Figura 5.2. Aplicação dos operadores de Sobel.

A cada ocorrência das condições descritas acima, o contador correspondente àquela direção é incrementado e ao final da varredura da imagem, o histograma de direção de bordas é normalizado pelo total de pixels de borda.

O descritor de forma final é composto por 8 valores reais, compreendidos no intervalo $[0, 1]$, e formatado da seguinte maneira – $(H_{0^\circ}, H_{45^\circ}, H_{90^\circ}, H_{135^\circ}, H_{180^\circ}, H_{225^\circ}, H_{270^\circ}, H_{315^\circ})$ – sendo a distância entre dois descritores calculada a partir da distância em quadras, obtida pela Equação (3.39) (Capítulo 3).

No método de **vizinhança de textura** a imagem é representada em níveis de cinza, correspondente ao canal I do espaço de cor HSI . Quando o valor de um pixel na vizinhança de oito for maior que o pixel central, então o contador direcional indicando a respectiva direção é incrementado. Após varridos todos os pixels, cada um dos oito contadores é normalizado em relação ao total de pixels da imagem.

O descritor final de textura é também composto por 8 valores reais, compreendidos no intervalo $[0, 1]$ e com formatação – $(V_{0^\circ}, V_{45^\circ}, V_{90^\circ}, V_{135^\circ}, V_{180^\circ}, V_{225^\circ}, V_{270^\circ}, V_{315^\circ})$ – sendo a distância entre dois descritores calculada a partir da distância em quadras.

5.4 Modelo de Similaridade Escalar

O modelo escalar de similaridade entre imagens é baseado na proposta de padrões locais conforme a abordagem de Stejić et al. [56] de forma que as imagens são divididas uniformemente em 16 regiões, conforme demonstram a Equação (5.1) e a Figura 5.3 a seguir.



Figura 5.3. Separação da imagem em regiões.

$$i = \bigcup r_j, \quad j = 1, 2, \dots, 16. \quad (5.1)$$

Considere q uma imagem de consulta, i uma determinada imagem do banco de imagens, r uma região da imagem, f uma característica da imagem, I o banco de imagens, R o conjunto de regiões da imagem e $F = \{cor, forma, textura\}$.

A função similaridade de características mede o grau de similaridade entre um par de imagens q e i , com relação à característica f na região r e é obtida pela Equação (5.2).

$$S_F(q, i, r, f) = 1 - D_F(q, i, r, f) \quad (5.2)$$

onde $D_F(q, i, r, f) \in [0,1]$ é a distância normalizada entre um par de vetores de características.

A função peso de característica $\omega_F(r, f)$ pondera com valores reais no intervalo $[-1,1]$ a importância da característica f na região r , conforme os critérios do usuário definido pela imagem de consulta e o conjunto de imagens marcadas na realimentação de relevância.

Com base nas funções de similaridade de características e no peso de características, a função similaridade de região pode ser obtida pela Equação (5.3). O resultado dessa equação representa o grau de similaridade entre as imagens q e i em relação à região r .

$$S_R(q, i, r) = \sum_{f \in F} \omega_F(r, f) S_F(q, i, r, f) \quad (5.3)$$

Cada uma das regiões tem graus de relevância diferentes para o usuário. Geralmente, o foco de uma imagem ou fotografia é o objeto importante retratado, posicionado nas regiões centrais da imagem. Entretanto, em alguns momentos, o desejo do usuário pode estar representado em um objeto retratado ao fundo, ou localizado no canto da imagem, fazendo com que qualquer estimativa a priori dos critérios do usuário seja meramente especulativo.

A função peso de região $\omega_R(r)$ pondera com valores reais no intervalo $[-1,1]$ a importância da região r , em conformidade com os requisitos do usuário. Assim, a partir do somatório ponderado dos valores de similaridade de cada uma das regiões, tem-se a função similaridade de imagem, que mede o grau de similaridade entre duas imagens q e i , conforme a Equação (5.4).

$$S_I(q, i) = \sum_{r \in R} \left(\omega_R(r) \cdot \sum_{f \in F} (\omega_F(r, f) S_F(q, i, r, f)) \right) \quad (5.4)$$

Diferentemente do modelo de Stejić e colaboradores, o modelo *VIPIGraf CBIR* não realiza as restrições

$$\sum_{f \in F} |\omega_F(r, f)| = 1 \quad (5.5)$$

$$\sum_{r \in R} |\omega_R(r)| = 1 \quad (5.6)$$

Portanto, essa simplificação elimina a necessidade de um operador de normalização no algoritmo genético. Observa-se, no entanto, que os valores das funções peso $\omega_F(r, f)$ e $\omega_R(r)$ devem estar compreendidos no intervalo $[-1, 1]$.

As funções pesos de características e regiões fazem o modelo flexível às necessidades de informação do usuário, pois se adequam de forma a aumentar a eficiência da recuperação. A introdução de pesos contínuos positivos e negativos por Stejić e outros permite ao sistema simular os conceitos de relevância, irrelevância e indesejabilidade expressos pelo usuário, em vários graus.

Embora apresente ser simples de calcular, o modelo *VIPIGraf CBIR* apresenta uma complexidade intrínseca de sistemas de recuperação de imagens por conteúdo, que é o salto semântico no estabelecimento de correspondência entre as características de baixo nível da imagem e os conceitos de alto nível dos usuários.

Buscando minimizar os efeitos do salto semântico, aplica-se a realimentação de relevância com a utilização de algoritmo genético.

5.5 Algoritmo Genético

O algoritmo genético projetado foi desenhado especificamente sob-medida para o problema de realimentação de relevância e readequação de pesos de características e de regiões no modelo de similaridade tal que otimizem a recuperação de imagens, condizentes aos critérios e necessidades momentâneas dos usuários, conforme Figura 5.4.

Um cromossomo representa a área de uma imagem partida uniformemente em n_R regiões, sendo que cada região possui um peso que representa sua importância em relação à imagem de consulta e outros três pesos, um para cada descritor de característica – cor, forma e

textura – representando suas respectivas relevâncias àquela região. A Equação (5.7) representa matematicamente cada gene do cromossomo:

$$G_j \equiv (\omega_F(r_j, cor), \omega_F(r_j, forma), \omega_F(r_j, textura), \omega_R(r_j)) \quad (5.7)$$

Na implementação *VIPIGraf CBIR*, 16 regiões definem a imagem, fazendo com que cada indivíduo contenha 16 genes, totalizando 64 pesos, todos eles valores reais pertencentes ao intervalo $[-1,1]$.

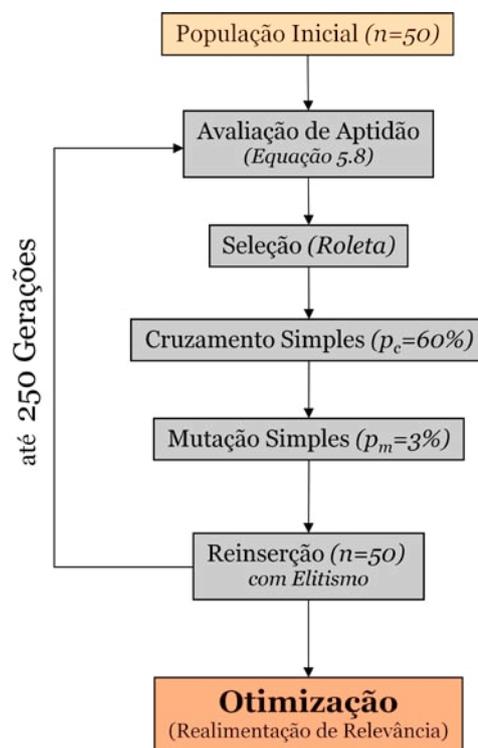


Figura 5.4. Fluxograma do algoritmo genético desenvolvido no *VIPIGraf CBIR*.

A seguir os operadores genéticos utilizados e decisões de projeto são apresentados. A seleção de indivíduos para a operação de reprodução aplicada foi a seleção proporcional pelo método da roleta, como descrito na Subseção 4.2.5, item a. Para reinscrição, os 50 melhores indivíduos, de acordo com o valor de aptidão entre pais e filhos, são selecionados para a próxima geração. O operador genético de cruzamento utilizado foi o *crossover* uniforme, apresentado na Subseção 4.2.6, item c, com taxa de cruzamento $p_c = 60,0\%$. A mutação

uniforme, onde cada peso selecionado para mutação tem seu valor substituído por outro valor real no intervalo $[-1,1]$ sorteado ao acaso, é aplicado ao algoritmo genético com probabilidade de mutação $p_m = 3,0\%$. Como critério de parada, o AG evolui por 250 gerações ou até que algum indivíduo de aptidão máxima seja encontrado durante o processo evolutivo.

A função aptidão (*fitness*) utilizada (Equação (5.8)) foi baseada no trabalho de Silva e colaboradores [54], que testaram a aplicação de dez diferentes funções de aptidão para experimentação, sendo estas classificadas em dois tipos:

- Funções de aptidão **não baseadas** em ordem; e
- Funções de aptidão **baseadas** em ordem

A principal diferença entre as duas classes de funções de aptidão é a consideração ou não da ordem em que a imagem relevante encontra-se no ranking. A Figura 5.4 ilustra essa diferença – nela uma função de aptidão não baseada em ordem não difere quantitativamente duas consultas onde o mesmo número de imagens é considerado relevante em relação a um total de 20 imagens recuperadas, mesmo que o primeiro caso tenha as duas imagens relevantes nas **primeiras** posições do ranking e o segundo caso tenha as duas imagens relevantes nas **últimas** posições do ranking.



Figura 5.5. Dois cenários avaliados da mesma forma por uma função de aptidão não baseada em ordem.

Em contrapartida, funções de aptidão baseadas em ordem levam o posicionamento do ranking em consideração, tornando o valor de utilidade de um documento inversamente proporcional à sua posição no ranking, como mostrado na Equação (5.8).

$$Utilidade(q, C) = \sum_{\forall i \in D} \frac{1}{\text{posição de } i \text{ no } ranking(q, C)} \quad (5.8)$$

onde

D é o conjunto de documentos conhecidos como relevantes para a consulta q ; e $ranking(q, C)$ representa o ranking de imagens recuperadas para a consulta q usando o indivíduo C .

5.6 Complexidade Computacional

A análise de algoritmos tem como objetivo determinar os recursos computacionais necessários para a execução de determinado algoritmo, comparando a complexidade temporal (eficiência) de algoritmos alternativos.

Em geral, a eficiência ou complexidade de um algoritmo é obtida por uma função cujo crescimento é proporcional ao tamanho da entrada n . Assim, têm-se algoritmos de complexidade constante (c), linear (n), logarítmica ($\log n$), polinomial (n^c), exponencial (c^n), fatorial ($n!$) ou compostas como $n \log n$ e $n^c \log n$, onde c é uma constante.

Os algoritmos do *VIPIGraf CBIR* podem ser divididos em duas partes – a busca inicial, que consiste no cálculo da similaridade considerando todos os pesos iguais a 1, e a realimentação de relevância, desempenhada pelo AG. Ambos têm complexidade $n \log n$, como demonstrado por Silva [59].

5.7 Modelo de Similaridade Vetorial

Diferentemente do método usual e escalar encontrado na literatura, a abordagem proposta neste trabalho se refere a uma modelagem de similaridade, técnica de visualização e realimentação de relevância multidimensionais cujo objetivo é aprimorar a efetividade do *feedback* do usuário ao permitir que ele personalize a consulta de acordo com suas preferências.

Assim, nota-se um decréscimo no problema do *gap*-semântico, intrínseco de sistemas CBIR, criando um sistema útil e favorável a vários propósitos de consulta, adequado a inúmeras aplicações como pesquisa por arte e cultura, sistemas de segurança e de proteção a direitos autorais, além de aplicativos *web* para busca de imagens.

Comumente, na literatura, utiliza-se o termo vetor de características (*image features array*) para uma descrição matemática de características como cor, forma, textura e etc. Como citado nos capítulos anteriores deste trabalho, um vetor de características tem por finalidade representar os aspectos visuais significativos de uma imagem.

Um ponto chave a ser observado na terminologia é a tradução do termo computacional “array” por vetor. Um *array* de características é um conjunto de números que representam a imagem e não necessariamente possui uma representação no espaço vetorial, podendo assim ser apenas uma simples tabela de dados.

Em uma abordagem realmente vetorial, os descritores de características são vetores que podem ser representados por módulo e ângulo. O termo **multidimensional** se refere à utilização de k vetores de características, representados no espaço $\mathbb{R}^{S_1} \times \mathbb{R}^{S_2} \times \dots \times \mathbb{R}^{S_k}$, para $k \geq 2$, ou seja, utilizando-se de no mínimo dois descritores de características representados individualmente no espaço \mathbb{R}^n e \mathbb{R}^s respectivamente.

Com a proposta da abordagem multidimensional, as aplicações e técnicas tradicionais de sistemas CBIR podem ser classificadas como “escalares”. A diferença entre os modelos de similaridade escalar e vetorial pode ser notada fazendo-se uma analogia entre grandezas escalares e vetoriais, conforme ilustra a Figura 5.5.



Figura 5.6. Informações escalares e vetoriais.

Também implementada no *VIPIGraf CBIR*, a modelagem de similaridade escalar, que representa o grau de similaridade entre a imagem de consulta q e determinada imagem i da coleção, é um número real proveniente do somatório algébrico ponderado da multiplicação de medidas de similaridades regionais, como já mostrado nas Equações (5.2) a (5.4) neste capítulo. Nestas equações, é possível identificar que valores de similaridade de certa característica, como cor, são simplesmente somados com os demais, como forma e textura, resultando em uma perda de informação característica da simplificação da representação dimensional, significando que grandezas não colineares são somadas de uma forma algébrica, acarretando em uma perda de dimensionalidade da medida de similaridade.

Na nova abordagem multidimensional, considerando um conjunto de k características que representam as imagens em um banco, a medida de similaridade entre uma imagem consulta e outra imagem da coleção é analisada em k componentes e as imagens são

ordenadas em um ranking, em um espaço k -dimensional, de acordo com suas projeções sobre os eixos x_n , onde $n = 1, 2, \dots, k$. As Equações (5.9) e (5.10) ilustram o modelo multidimensional de similaridade:

$$S_I(q, i) = (S_I^1(q, i), S_I^2(q, i), \dots, S_I^k(q, i)) \in \mathbb{R}^k \quad (5.9)$$

onde

$$S_I^n(q, i) = \sum_{r \in R} (\omega_R(r) \omega_n(r) S_n(q, i, r)) \quad (5.10)$$

Aparentemente, o modelo multidimensional pode aparentar uma versão um pouco mais complexa da modelagem de similaridade escalar já apresentado, porém uma série de benefícios pode advir de sua aplicação:

- O usuário pode desabilitar a apresentação de resultados segundo determinada característica k que por algum motivo não apresente resultados satisfatórios, concentrando-se nos descritores que estão trabalhando em conformidade com os seus requisitos de informação;
- Mais do que um descritor para a mesma característica, como momentos de cores, correlogramas de cores e histogramas de cores, para cor, por exemplo, podem ser implementados trabalhando paralelamente como excelente mecanismo para experimentação;
- O usuário assume maior poder sobre os mecanismos de *feedback* de relevância e apresentação de resultados, participando ativamente do processo de pesquisa e interagindo a cada resposta do sistema, se co-responsabilizando pelos resultados recuperados;

- Por se tratar de uma abordagem multidimensional, um mesmo documento relevante pode ser recuperado por quaisquer dos k descritores de características utilizados, de forma a reforçar sua seleção por parte do usuário;
- Mesmo utilizando a modelagem multidimensional, o usuário ainda tem a opção de visualizar uma combinação escalar dos resultados, sem trazer ônus significativo ao desempenho do sistema ou tempo de resposta; e
- Geralmente, sua aplicação não requer a realização de drásticas mudanças e sua implementação é realizada a partir de algumas adaptações ao sistema corrente, facilitando o controle e a experimentação para a comparação de resultados.

Um grande problema das abordagens escalares de modelagem de similaridade ocorre pelo fato de que o resultado final é uma soma ponderada das similaridades regionais de cada uma das características utilizadas, de maneira que a ocorrência de um efeito colateral não pode ser evitada, pois, devido aos pesos serem positivos e negativos, uma imagem muito similar de acordo com determinada característica, mas extremamente diferente quanto à outra pode ter um valor total de similaridade que não a figure entre as melhores posições do ranking total ponderado, mesmo que essa imagem seja relevante ao usuário. Na abordagem multidimensional vetorial de similaridade, esse comportamento indesejado não ocorre.

5.8 Considerações Finais deste Capítulo

O sistema *VIPIGraf CBIR* foi apresentado em detalhe ao longo deste capítulo, que delineou sua estrutura geral, ilustrando seu fluxograma básico de funcionamento e apresentou os descritores de características empregados e o primeiro modelo de similaridade adotado, o

qual classifica-se como “escalar”, pelo motivo de considerar os valores de similaridade como um somatório algébrico das similaridades regionais e de características de forma ponderada.

Detalhes sobre a construção do algoritmo genético utilizado para a realimentação de relevância foram apresentados, bem como a medida de aptidão de um documento, por meio do cálculo de sua utilidade. A complexidade computacional do sistema também foi avaliada, mostrando que o crescimento da base de dados não influencia em um acréscimo significativo no tempo de execução da pesquisa.

Além de apresentar especificidades do projeto do *VIPIGraf CBIR*, o Capítulo 5 propõe uma abordagem multidimensional de modelagem de similaridade e de realimentação de relevância ainda não encontrado na literatura. Essa contribuição apresenta considerável melhoria na experiência de utilização de um sistema CBIR por parte dos usuários, tornando-os parte fundamental e co-responsável pela qualidade dos resultados das pesquisas.

O próximo capítulo apresenta os resultados obtidos experimentalmente e o aprimoramento no desempenho do sistema *VIPIGraf CBIR* ao utilizar a abordagem multidimensional proposta, em contrapartida ao modelo escalar tradicional, avaliando alguns experimentos realizados e consultas exemplos.

RESULTADOS EXPERIMENTAIS

6.1 Introdução

Com o objetivo de avaliar a nova estratégia de modelagem de similaridade e realimentação de relevância conforme a abordagem multidimensional proposta neste trabalho, 50 consultas, selecionadas arbitrariamente e realizadas no banco de imagens BD-20000, foram conduzidas e as avaliações mais relevantes são apresentadas neste capítulo.

Para fins de comparação, os resultados dos experimentos são apresentados em contraposição àqueles obtidos pelos trabalhos de Stejić e colaboradores [56] e de Silva e colaboradores [59], com a utilização da mesma medida de desempenho empregada nos referidos trabalhos, ou seja, a Precisão-*R*.

Os bancos de imagens de teste são empregados sob a nova abordagem proposta neste trabalho, similarmente àqueles que a este estão sendo comparados. Outro banco de imagens, contendo 10.000 imagens, é também utilizado.

Os resultados experimentais são abordados em uma ordem adequada à clara compreensão dos objetivos deste trabalho. Primeiramente, são apresentados os bancos de imagens empregados na experimentação. Então, alguns resultados práticos da implementação

da abordagem multidimensional são ilustrados e observações advindas de sua utilização, elucidadas. Por fim, comparações são realizadas aos trabalhos anteriormente citados a fim de qualificar a aplicação da abordagem proposta quanto ao seu desempenho.

6.2 Bancos de Imagens Empregados

Para a consecução dos experimentos, seis bancos de imagens de teste foram utilizados, descritos a seguir.

a. Vistex-60

O Vistex-60, codinome para *Vision Texture Database* [60], é um banco com 60 imagens coloridas, com resolução de 128 x 128 pixels, de cenas do mundo real. O banco é composto por 10 categorias, cada qual contendo de 2 a 12 imagens. A fonte é encontrada sob o diretório *pub/FLAT/scene128x128/*. O *Massachusetts Institute of Technology (MIT)* é o mantenedor deste banco. A Figura 6.1 ilustra uma amostra desse banco de dados.

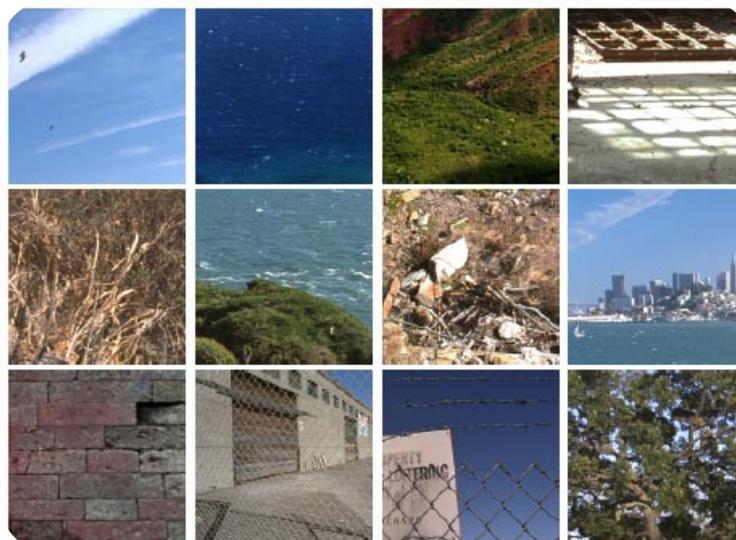


Figura 6.1. Banco de imagens Vistex-60.

b. Vistex-167

O Vistex-167, também pertencente ao *Vistex Database*, é um banco com 167 imagens coloridas, com resolução de 128 x 128 pixels, de texturas homogêneas. O banco é composto por 19 categorias, cada qual contendo de 3 a 20 imagens. A fonte é encontrada sob o diretório *pub/FLAT/128x128/*. A Figura 6.2 apresenta uma amostra do Vistex-167.



Figura 6.2. Banco de imagens Vistex-167.

c. Brodatz-208

O Brodatz-208 [61] é um banco de dados que contém 208 imagens em tons de cinza, com resolução de 128 x 128 pixels, representativas de texturas homogêneas. O banco é dividido em 13 categorias, cada uma contendo 16 imagens. Cada categoria do Brodatz corresponde a uma imagem de 512 x 512 pixels repartida em 16 sub-imagens não sobrepostas de 128 x 128 pixels. A Figura 6.3 ilustra uma amostra desse banco de imagens.

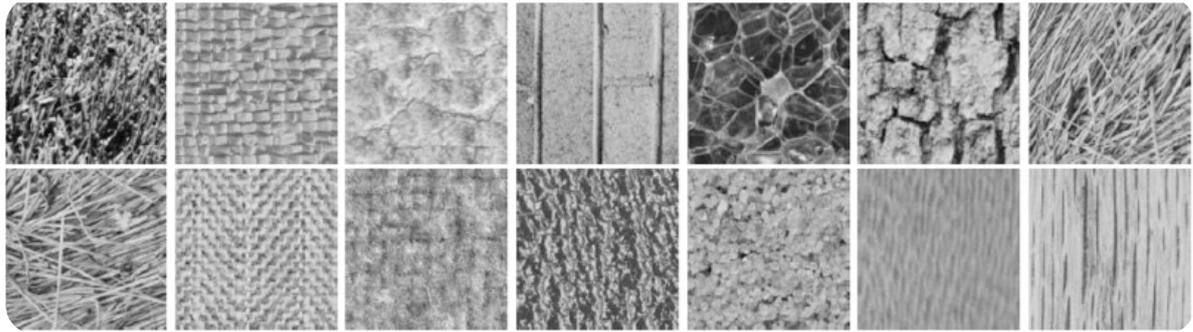


Figura 6.3. Banco de imagens Brodatz-208.

d. Corel-1000

O Corel-1000 [62] é um subconjunto do banco de imagens da Corel, contendo 1.000 imagens com resolução de 384 x 256 (ou 256 x 384) *pixels*. O banco é composto por dez categorias, sendo que cada uma contém 100 imagens. A Figura 6.4 apresenta uma amostra desse banco.

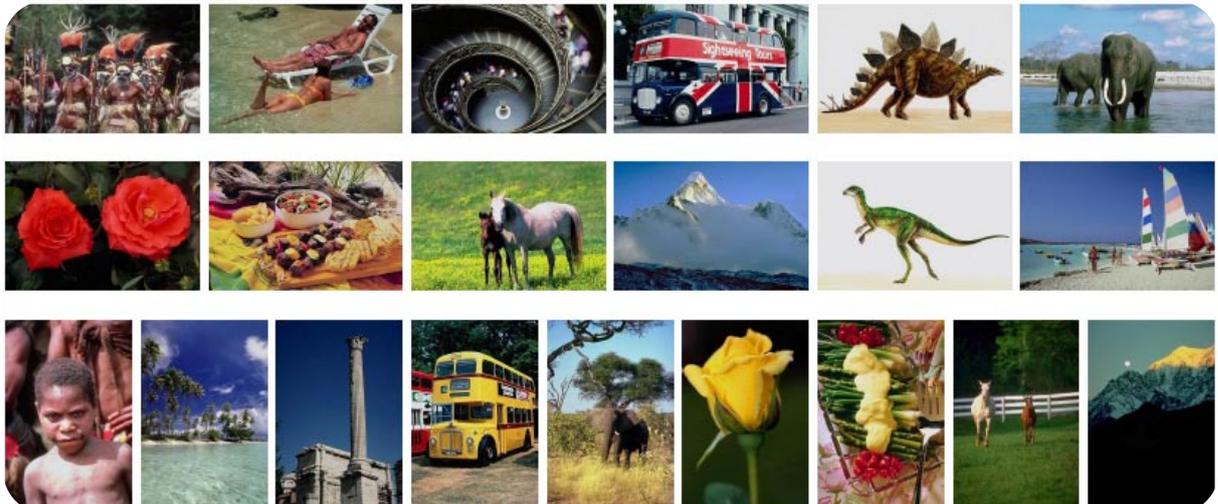


Figura 6.4. Banco de imagens Corel-1000.

e. BD-10000

O BD-10000 é um banco de imagens que contém 10.000 imagens reais, cobrindo uma ampla variedade de categorias semânticas como textura uniforme, pôr-do-sol, nuvens, oceanos, mamíferos, peixes, pássaros, répteis, plantas, flores, paisagens naturais, edifícios, motocicletas, carros e aviões, entre outros. Do total de imagens, 1.000 delas são provenientes do banco Corel-1000 e as demais foram coletadas de bases de imagens públicas, disponíveis na internet, principalmente de CalPhotos [63], e pertencentes a diferentes categorias daquelas da coleção Corel-1000. A Figura 6.5 ilustra uma amostra desse banco de imagens.



Figura 6.5. Banco de imagens BD-10000.

f. BD-20000

Representado pela Figura 6.6 a partir de uma pequena amostra, o BD-20000 contém imagens reais de diversas categorias semânticas, contudo, nenhuma delas é pré-classificada ou agrupada, de forma a simular um banco de imagens Web real, pelo qual os usuários cadastrados inserem suas imagens e mantêm álbuns, como o serviço *Flickr*, além de um robô de busca que acrescenta elementos à coleção periodicamente.



Figura 6.6. Banco de imagens BD-20000.

6.3 Experimentos

Várias consultas práticas, presentes no dia-a-dia, foram executadas aplicando a abordagem proposta neste trabalho [64]. De forma a ilustrar o desempenho desse sistema, dois exemplos utilizando três descritores de características ($k = 3$; cor, textura e forma) são apresentados na seqüência. Na seção seguinte, os resultados obtidos são comparados à abordagem tradicional e escalar para apresentação e discussão.

A Precisão- R foi a medida de precisão utilizada. Para a realização dos testes, o percentual de imagens relevantes encontradas nas primeiras 18 imagens do ranking foi examinado, ou seja, $R = 18$. O banco de imagens utilizado para a realização dos experimentos exemplos foi o BD-20000.

No primeiro experimento, o objetivo é recuperar imagens de carros do banco de dados, considerando uma imagem exemplo encontrada no disco rígido do computador. Em contrapartida, no segundo teste, a intenção é recuperar uma coleção de figuras de flores distintas, baseada em uma imagem exemplo de uma plantação de girassóis encontrada na Internet. As imagens exemplo inseridas ao sistema de consulta são apresentadas na Figura 6.7.



Figura 6.7. Imagens de consulta para os experimentos-exemplo.

Com a finalidade de avaliar individualmente cada descritor de característica utilizado e identificar o potencial de avanço e aprimoramento de desempenho da abordagem proposta em relação à implementação escalar convencional, foram comparadas durante os experimentos, as 18 primeiras imagens para:

- Apenas cor;
- Apenas forma;
- Apenas textura;
- CBIR tradicional (escalar, convencional); e
- CBIR multidimensional.

Uma importante observação é que no modelo tradicional, a medida de similaridade é um somatório algébrico ponderado das similaridades regionais para cada uma das características, enquanto no modelo multidimensional, obtém-se medidas de similaridade para cada uma das k características descritas.

No caso da utilização de três descritores ($k = 3$), pode-se observar pelos exemplos de consulta apresentados a seguir nesta seção, o motivo de se adotar o valor da Precisão- R com $R = 18$, ou seja, um número divisível pelo número de características k .

Este fato é considerado pelo motivo de que os resultados apresentados pela abordagem multidimensional, proposta neste trabalho, são ordenados não somente por um simples *ranking*, mas por k rankings, de forma a apresentar as R/k imagens com melhores valores de similaridade em relação à imagem de consulta.

Portanto, adotando as primeiras 18 imagens recuperadas na primeira consulta ao sistema para (a) somente-cor, (b) somente-forma, (c) somente-textura, (d) pelo modelo tradicional escalar e (e) pelo modelo multidimensional proposto, tem-se representadas apenas as 6 primeiras imagens para cada um dos eixos (cor, forma e textura) nessa última representação.

6.3.1 Consulta Exemplo – Carros

A partir da utilização da imagem de um carro encontrada no disco rígido do computador pelo qual os experimentos foram realizados, os seguintes resultados, apresentados na Figura 6.8, foram obtidos após a primeira consulta ao sistema. As imagens recuperadas e consideradas relevantes são marcadas com uma borda vermelha e contabilizadas à direita dos quadros entre parênteses, também em vermelho.

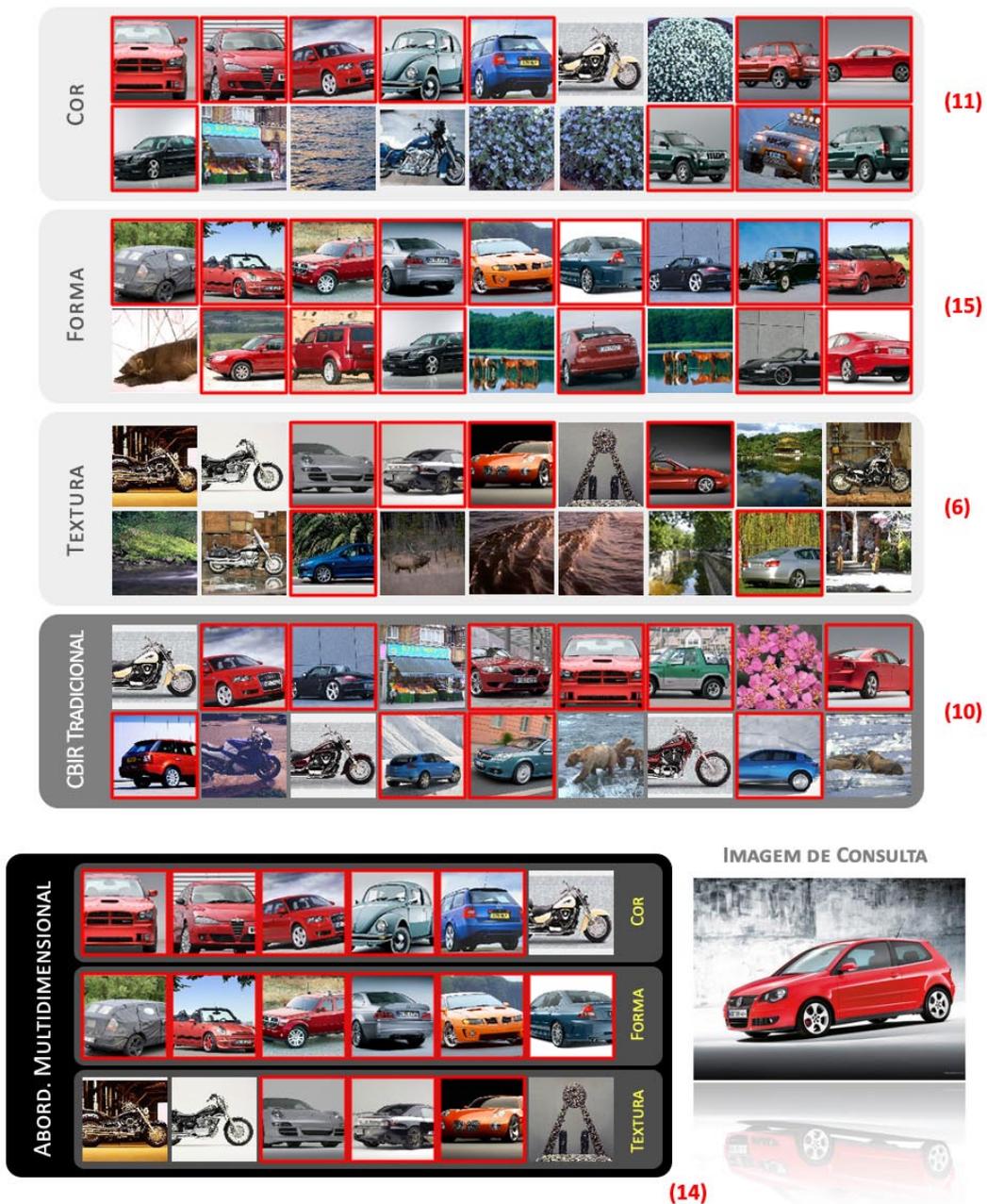


Figura 6.8. Imagem consulta e miniaturas dos primeiros resultados produzidos pelo experimento Carros.

Após a apresentação e análise dos resultados obtidos, pode-se observar do experimento *Carros* que a abordagem multidimensional sugerida mantém intactos os resultados para as primeiras imagens do ranking de cor-, forma- e textura- apenas enquanto o modelo tradicional escalar apresenta uma ordem completamente diferente.

Esse comportamento ocorre por causa da somatória direta dos valores de similaridade das características provenientes dos descritores no modelo tradicional causando um efeito colateral inesperado.

Em suma, valores de similaridade de cor são adicionados aos valores de similaridade de forma e textura, causando uma perda de informação, já que no valor final de similaridade uma imagem bastante relevante sob aspectos de textura, mas totalmente irrelevante quanto à cor e fracamente relevante à forma, provavelmente será, no somatório, mal posicionada no *ranking*, sendo que poderia representar um desejo inicial de informação do usuário.

Após a primeira busca ao sistema, o primeiro teste apresentou melhores resultados para forma- apenas e cor- apenas em detrimento ao resultado escalar combinado no modelo tradicional, sendo que forma- apenas recuperou imagens relevantes com uma eficácia 50% maior do que no modelo escalar.

6.3.2 Consulta Exemplo – *Coleção de Flores*

A partir da utilização da imagem de uma plantação de girassóis encontrada na Internet, os seguintes resultados, apresentados na Figura 6.9 foram obtidos após a primeira consulta ao sistema. Assim como no primeiro exemplo, as imagens recuperadas e consideradas relevantes são marcadas com uma borda vermelha e contabilizadas à direita dos quadros entre parênteses, também em vermelho.



Figura 6.9. Imagem consulta e miniaturas dos primeiros resultados produzidos pelo experimento *Coleção de Flores*.

O segundo teste apresentou melhores resultados quantitativos para cor- apenas do que o resultado combinado no modelo tradicional, atingindo uma precisão de 94,44%. Os resultados de forma- apenas trouxeram diversidade, um fator qualitativo, à coleção, ao produzirem uma variedade de tipos de flores.

Experimentação e vários trabalhos encontrados na literatura apontam que um único descritor de característica não é o mais apropriado para todos os tipos de consultas, e este fato pode ser observado nos testes realizados neste capítulo.

Além disso, a modelagem de similaridade escalar em aplicações CBIR introduz um efeito colateral que deteriora a efetividade de seus resultados, de sua apresentação e da realimentação de relevância, em cadeia, por desprezar informações importantes de correlação entre as características. Esse problema pode ser apontado em ambos os experimentos, mais evidenciado no exemplo *Carros*, como notado anteriormente.

Um aspecto interessante da abordagem multidimensional proposta é a organização e paginação automática de resultados no caso da exclusão de determinado eixo de característica, como pode ser observado no experimento *Coleção de Flores*, onde resultados de textura não apresentavam relevância ao usuário, não contribuindo às necessidades de informação como esperado, sendo conseqüentemente desconsiderados para que mais imagens de cor e forma fossem evidenciadas sem a necessidade de navegação, respeitando o número total de 18 imagens apresentadas.

6.4 Conclusões

Pode-se observar que a abordagem multidimensional recuperou 4 imagens a mais do que o modelo tradicional escalar para ambos os testes realizados.

Em todas as consultas executadas durante a fase de experimentação, o modelo proposto apresentou um acréscimo, ou pelo menos a manutenção, no número de imagens relevantes apresentadas após a primeira consulta ao sistema de recuperação, conseqüentemente aprimorando a precisão do mesmo.

A Tabela 6.1 apresenta o número de imagens relevantes e os valores de precisão ($R = 18$) para os experimentos exemplos e alguns outros, totalizando 50 consultas, selecionadas arbitrariamente e realizadas no banco de imagens BD-20000.

A modelagem multidimensional proposta apresentou uma melhoria na métrica de desempenho precisão-*R* na ordem de 13,33% quando comparado à abordagem escalar tradicional.

Mesmo que ambos os modelos permitam a navegação pelos resultados por parte do usuário na intenção de encontrar documentos relevantes à sua necessidade de informação, um fator importante a ser considerado é a expectativa evidenciada pela grande maioria dos usuários de sistemas CBIR que consiste em encontrar imagens relevantes logo na primeira página do *ranking*, agilizando o processo de busca e tornando o sistema mais amigável ao usuário e eficiente.

Tabela 6.1. Resultados experimentais – número de imagens relevantes e precisão-*R* para 50 consultas teste.

Experimento	Cor Apenas	Forma Apenas	Textura Apenas	Tradicional (Escalar)	Abordagem Proposta
<i>1. Carros</i>	11 (61,11%)	15 (83,33%)	6 (33,33%)	10 (55,56%)	14 (77,78%)
<i>2. Flores</i>	17 (94,44%)	9 (50,00%)	4 (22,22%)	9 (50,00%)	13* (72,22%)
<i>Resultado Médio</i>	10,2 (56,67%)	8,6 (47,78%)	2,4 (13,33%)	7,2 (40,00%)	9,6 (53,33%)

* Desabilitando o eixo de textura

6.5 Comparação ao Estado da Arte

Toda nova abordagem ou método sugerido para a solução de problemas deve ser comparado e referenciado a outras abordagens ou mesmo ao estado da arte a fim de qualificar

a aplicação dessa abordagem quanto ao seu desempenho, tanto quantitativamente quanto qualitativamente. A esse processo de referenciação, dá-se o nome de *benchmarking*.

Além da comparação entre as abordagens escalar e multidimensional, este trabalho realizou consultas nos mesmos bancos de imagens de teste utilizados por outros pesquisadores da área. Esses bancos de imagens, apresentados no início deste capítulo, servem como base para comparação entre diversos trabalhos, de forma a referenciá-los frente ao estado da arte.

Em [56], Stejić et al. apresentam uma tabela comparativa entre diversos sistemas CBIR. Silva propõe, em [59], a utilização de diferentes funções aptidão para o AG. Os principais resultados desses trabalhos e os obtidos pelo *VIPIGraf CBIR* em ambas as modelagens de similaridade (escalar e vetorial) são apresentados na Tabela 6.2, ordenados conforme seu desempenho percentual geral (medida: *Precisão-R*), em ordem crescente.

Tabela 6.2. Comparação entre o desempenho do modelo multidimensional e o estado da arte.

Banco de Imagens	Vistex-60	Vistex-167	Brodatz-208	Corel-1000	BD-10000*	
Método	Precisão-R (%)					Total
<i>Cor-Textura</i>	48	41	-	43	-	44,00
<i>SIMPLIcity</i>	-	-	-	46	-	46,00
<i>P-F1 (Silva et. al.) [59]</i>	90	76	98	80	58,60	86,00
<i>WLSP-C± (Stejić et al.) [56]</i>	95	84	95	75	-	87,25
<i>LSP-C± (Stejić et al.) [56]</i>	98	86	96	76	-	89,00
<i>P-F2 (Silva et. al.) [59]</i>	98	91	99	78	56,40	91,50
<i>VIPIGraf Escalar</i>	98	91	99	78	56,40	91,50
<i>VIPIGraf Multidimensional</i>	98	92	99	81	60,40	92,50

* Não foi levado em consideração para o cálculo do desempenho geral.

CONCLUSÕES, CONTRIBUIÇÕES E TRABALHOS FUTUROS

7.1 Conclusões

O corrente trabalho obteve resultados significativos na apresentação de uma abordagem multidimensional de modelagem de similaridade e técnica de *feedback* de relevância, minimizando o bem conhecido problema do *gap*-semântico, intrínseco de sistemas de recuperação de imagens baseados em conteúdo, ao permitir aos usuários maior poder e flexibilidade na personalização de suas consultas de acordo com suas preferências.

A interface sugerida é altamente intuitiva e flexível [64], permitindo desenvolvimentos concorrentes de descritores de características de imagens de maiores dimensões, para experimentação e aprimoramento de resultados.

Com a aplicação da estratégia multidimensional de visualização e realimentação de relevância, incluindo a capacidade de habilitação e desabilitação de eixos de características, foi obtido um incremento maior que 10% na precisão para as consultas, aperfeiçoando à

excelência a eficiência do sistema CBIR. Em comparação com outros sistemas, apresentou igual ou melhor desempenho quando testado sob cinco diferentes bancos de imagens de teste.

Os testes realizados comprovaram que a abordagem apresentada neste trabalho pode aprimorar, garantidamente, tanto quantitativa quanto qualitativamente, o resultado e a experiência dos usuários em sistemas de recuperação de imagens.

Com sucintas alterações, dependendo da maneira com que foram desenvolvidos, outros sistemas CBIR tradicionais podem adotar a abordagem multidimensional, conseqüentemente elevando seus níveis de precisão.

7.2 Contribuições

O principal objetivo deste trabalho foi elucidar os conceitos e metodologias aplicados no desenvolvimento de um sistema de recuperação de imagens baseado em conteúdo com realimentação de relevância utilizando um algoritmo genético e apresentar uma nova estratégia, multidimensional, para a modelagem de similaridade e apresentação de resultados, delineando as idéias, boas práticas e tendências neste campo da ciência.

A princípio, a maneira de pesquisar essas coleções de imagens seria por indexação de palavras-chave e categorias ou pela simples procura, imagem a imagem, tornando o processo insatisfatório ou mesmo inviável considerando o atual estado da arte tecnológico.

Em um sistema CBIR, imagens são comparadas com base na similaridade entre a requisição do usuário (*query*) e as características extraídas da coleção de imagens.

O desenvolvimento deste trabalho trouxe contribuições no contexto de sistemas de recuperação de imagens por aplicar uma estratégia ainda não observada na literatura. Diferentemente do método usual e escalar, a abordagem proposta se refere a uma modelagem

de similaridade, técnica de visualização e realimentação de relevância multidimensionais com o objetivo de aprimorar a efetividade de sistemas CBIR.

Deste modo, há uma atenuação no problema do salto semântico, intrínseco desses sistemas. Além disso, cria-se uma solução útil e favorável a vários propósitos de consulta, adequada a inúmeras aplicações para busca de imagens, de caráter acadêmico e comercial.

Na nova abordagem multidimensional, considerando um conjunto de k características que representam as imagens em um banco, a medida de similaridade entre uma imagem consulta e outra imagem da coleção é analisada em k componentes e as imagens são ordenadas em um ranking, em um espaço k -dimensional, de acordo com suas projeções sobre os eixos x_n , onde $n = 1, 2, \dots, k$, como ilustrado pelas Equações (5.9) e (5.10).

Além do desenvolvimento do *VIPIGraf CBIR Multidimensional* que apresentou resultados superiores quando comparado a outros sistemas de recuperação de imagens, o presente estudo teve caráter multidisciplinar por trabalhar com a aplicação de descritores de características, métricas, modelos e estratégias de similaridade entre imagens, desenvolvimento de algoritmos genéticos para a aplicação da realimentação de relevância, além de estudos de complexidade computacional, que mostraram que o crescimento da base de dados não influencia em um acréscimo significativo no tempo de execução da pesquisa.

O algoritmo genético foi projetado sob-medida para o problema de realimentação de relevância e readequação de pesos de características e de regiões no modelo de similaridade tal que otimizem a recuperação de imagens, condizentes aos critérios e necessidades momentâneas dos usuários. Sua principal diferença em relação ao AG proposto por Stejić é sua abordagem multidimensional e a ausência da necessidade de um operador de normalização pela não utilização das restrições apresentadas pelas Equações (5.5) e (5.6).

No projeto do AG, um cromossomo representa a área de uma imagem partida uniformemente em 16 regiões, sendo que cada região possui um peso que representa sua

importância em relação à imagem de consulta e outros três pesos, um para cada descritor de característica – cor, forma e textura – representando suas respectivas relevâncias àquela região.

Os operadores genéticos utilizados e decisões de projeto foram baseados em um estudo aprofundado destes sistemas. A seleção de indivíduos para a operação de reprodução aplicada foi a seleção proporcional pelo método da roleta. Para reinserção, os 50 melhores indivíduos são selecionados para a próxima geração. O operador genético de cruzamento utilizado foi o *crossover* uniforme com taxa de cruzamento $p_c = 60,0\%$. A mutação é aplicada ao algoritmo genético com probabilidade $p_m = 3,0\%$. Como critério de parada, o AG evolui até o limite de 250 gerações.

Um grande problema das abordagens escalares de modelagem de similaridade ocorre pelo fato de que o resultado final é uma soma ponderada das similaridades regionais de cada uma das características utilizadas, de maneira que a ocorrência de um efeito colateral não pode ser evitada, pois, devido aos pesos serem positivos e negativos, uma imagem muito similar de acordo com determinada característica, mas extremamente diferente quanto à outra pode ter um valor total de similaridade que não a figure entre as melhores posições do ranking total ponderado, mesmo que essa imagem seja relevante ao usuário. Na abordagem multidimensional vetorial de similaridade, esse comportamento indesejado não ocorre.

Aparentemente, o modelo multidimensional pode aparentar uma versão um pouco mais complexa da modelagem de similaridade escalar, porém uma série de benefícios pode advir de sua aplicação. Dentre elas:

- O usuário pode desabilitar a apresentação de resultados segundo determinada característica k que por algum motivo não apresente resultados satisfatórios, concentrando-se nos descritores que estão trabalhando em conformidade com os seus requisitos de informação;

- Mais do que um descritor para a mesma característica, como momentos de cores, correlogramas de cores e histogramas de cores, para cor, por exemplo, podem ser implementados trabalhando paralelamente como excelente mecanismo para experimentação;
- O usuário assume maior poder sobre os mecanismos de *feedback* de relevância e apresentação de resultados, participando ativamente do processo de pesquisa e interagindo a cada resposta do sistema, se co-responsabilizando pelos resultados recuperados;
- Por se tratar de uma abordagem multidimensional, um mesmo documento relevante pode ser recuperado por quaisquer dos k descritores de características utilizados, de forma a reforçar sua seleção por parte do usuário;
- Mesmo utilizando a modelagem multidimensional, o usuário ainda tem a opção de visualizar uma combinação escalar dos resultados, sem trazer ônus significativo ao desempenho do sistema ou tempo de resposta; e
- Geralmente, sua aplicação não requer a realização de drásticas mudanças e sua implementação é realizada a partir de algumas adaptações ao sistema corrente, facilitando o controle e a experimentação para a comparação de resultados.

Além dessas contribuições, uma rica pesquisa bibliográfica e a abordagem de tópicos essenciais para o desenvolvimento de um sistema CBIR como a realimentação de relevância, estratégias de visualização e apresentação de resultados foram realizados ao longo deste trabalho, bem como a apresentação e análise de 14 exemplos práticos de implementações de sistemas de recuperação de imagens baseado em conteúdo de cunho acadêmico e comercial.

7.3 Trabalhos Futuros

A partir deste trabalho, pode-se visualizar o desenvolvimento de um novo sistema CBIR, arquitetado para a Web 2.0, que contenha um cadastro de usuários e que permita que estes insiram suas imagens, ilustrações, figuras e desenhos, entre outros, que por sua vez, possa pesquisar a Internet, por meio de um robô de busca, as diversas imagens encontradas na grande rede, armazenando apenas suas informações descritivas (vetores de características), seu endereço eletrônico (uri) e uma cópia em miniatura para fins de apresentação e visualização rápida.

Para a consecução desse novo sistema, é fundamental o estudo, análise e desenvolvimento de alguns tópicos, descritos a seguir. Esses tópicos são objetos de estudo e propostas de trabalhos futuros advindos desta dissertação.

a. Robô de Busca por Imagens na Internet

Um robô de busca, associado ao sistema CBIR, irá permitir com que este atenda necessidades diversas por documentos, desde pesquisas por elementos básicos como carros, demonstradas neste trabalho, como consultas mais específicas por modelos tridimensionais de determinada peça do motor de um automóvel, por exemplo.

b. Implementação de vários descritores de características

Com a implementação de diversos descritores de características para cor, textura, forma e posição, o sistema conterà um banco de dados altamente descritivo, permitindo ao usuário controle integral sobre sua consulta, além de tornar o sistema auto-ajustável para determinadas classes de consultas.

Essa característica de auto-ajuste pode ser considerada após a intensiva realização de testes aplicando um novo algoritmo genético, uma rede neural artificial ou outra abordagem inteligente que consiga classificar determinados tipos de consulta, visando o aprimoramento de resultados.

Por exemplo, observado que a consulta por carros geralmente apresenta melhores resultados para o descritor de característica de forma-somente por meio de histogramas de direção de bordas, o sistema automaticamente se ajustaria para utilizar somente este descritor, caso o usuário demonstrasse este tipo de interesse em sua pesquisa.

c. Técnicas de Indexação

O estudo e análise de técnicas de indexação podem aprimorar o desenvolvimento das consultas do sistema, aprimorando sua eficácia e eficiência.

d. Algoritmo genético executado concorrentemente

O algoritmo genético tem na sua concepção a avaliação do valor de aptidão, indivíduo por indivíduo, geração a geração, o que pode torná-lo dispendioso em relação ao seu custo computacional.

A implementação desses algoritmos utilizando conceitos da programação paralela por intermédio de *threads* em conjunto à sua instalação em uma “*Web-Farm*”, que é uma série de servidores trabalhando paralelamente (*cloud-computing*), pode tornar o desempenho da consulta incomparavelmente superior, principalmente se tratando de uma aplicação web, onde dezenas ou até milhares de usuários podem acessar o sistema simultaneamente.

e. Designação de uma base de características padrão

Com o estudo da correlação entre cada descritor de característica, espera-se estabelecer uma base multidimensional que apresente ortogonalidade entre seus eixos de forma que se possa delinear cada descritor de característica existente a partir de projeções aos eixos dessa base padrão.

f. Algoritmo de identificação por verossimilhança

A mesma imagem é encontrada na Internet inúmeras vezes, sendo que em alguns cenários é levemente alterada em relação à sua escala, coloração ou sofre alguma transformação de rotação de forma que o banco de dados não necessariamente precise indexá-la.

A utilização de um algoritmo que identifique cópias, mesmo que levemente alteradas, se torna uma ferramenta importante a esses sistemas web em relação à dimensionalidade ou custo computacional para armazenamento de dados, de maneira a minimizar, ou pelo menos, diminuir, os recursos utilizados que não sejam efetivamente necessários.

g. Métricas de desempenho e avaliação qualitativa pelos usuários

A implantação de outras métricas para o acompanhamento do desempenho do sistema como tempo de resposta, tempo total de consulta, número de documentos relevantes baixados podem ser mecanismos de retro-alimentação para a melhoria contínua do sistema, bem como a avaliação qualitativa pelos usuários ao final da consulta, gerando um índice de satisfação.

Referências Bibliográficas

- [1] LIUA, Y.; ZHANGA, D.; LUA, G. AND MA, W. **A survey of content based image retrieval with high-level semantics.** Pattern Recognition, (40):262-282, 2007.
- [2] RUI, Y.; HUANG, T. S. AND CHANG, S. **Image Retrieval: Current techniques, promising directions, and open issues.** Journal of Visual Communication and Image Representation, (10):39-62. 1999.
- [3] BERNERS-LEE, TIMOTHY. **Information Management: A Proposal.** World Wide Web Consortium. 1989. <http://www.w3.org/History/1989/proposal.html>.
- [4] RICHARDSON, E. C. **The beginnings of libraries.** Princeton University Press. 1914.
- [5] BERNERS-LEE, TIM; CAILLIAU, R.; GROFF, J-F AND POLLERMANN, B. **World-Wide Web: The information universe.** Electronic Networking: Research, Applications and Policy 1(2):74–82. 1992. <http://citeseer.ist.psu.edu/article/bernerslee92worldwide.html>.
- [6] MANNING, C. D.; RAGHAVAN, P. AND SCHÜTE, H. **Introduction to Information Retrieval.** Cambridge University Press. 2008.
- [7] BAEZA-YATES, R. AND RIBEIRO-NETO, B. **Modern Information Retrieval.** Addison-Wesley, Essex, UK. 1999.
- [8] HUIJSMANS, D. P. AND SEBE, N. **How to complete performance graphs in content-based image retrieval: Add generality and normalize scope.** IEEE Trans. Pattern Anal. Mach. Intell. 27, 2, 245–251. 2005.
- [9] MERCHEY, JASON. **Values of the Wise: Humanity's Highest Aspirations.** p.63. Haverford, PA: Infinity Publishing. 2003.
- [10] VELTKAMP, R. AND TANASE, M. **Content-Based Image Retrieval Systems: A Survey.** Revised and Extended version of Technical Report UU-CS-2000-34, Department of Computing Science, Utrecht University. 2002.

- [11] DATTA, R., JOSHI, D., LI, J., AND WANG, J. Z. **Image retrieval: Ideas, influences, and trends of the new age.** ACM Computing Survey 40, 2, Article 5. 2008.
<http://doi.acm.org/10.1145/1348246.1348248>
- [12] CHANG, S. K. AND HSU, A. **Image information systems: where do we go from here?** IEEE Transactions on Knowledge and Data Engineering, 5(5):431-442. 1992.
- [13] H. TAMURA, S. MORI, AND T. YAMAWAKI. **Texture features corresponding to visual perception.** IEEE Transactions on Systems, Man and Cybernetics, 8(6):460–473, 1978.
- [14] CNN. **Computer decodes Mona Lisa's smile.**
<http://www.cnn.com/2005/TECH/12/16/mona.lisa.smile/index.html>. 2005.
- [15] FLICKR. **The flickr homepage.** <http://www.flickr.com>. 2002.
- [16] SMEULDERS, A. W.; WORRING, M.; SANTINI, S.; GUPTA, A.; AND JAIN, R. **Content-based image retrieval at the end of the early years.** IEEE Transactions on Pattern Anal. Mach. Intell. 22, 12, 1349–1380. 2000.
- [17] AIRLINERS.NET. **Airliners.net homepage.** <http://www.airliners.net>. 2005.
- [18] GONZALEZ, R. C. AND WOODS, R. E. **Digital Image Processing.** Prentice-Hall, 2nd edition. 2002.
- [19] RUI, Y.; T.S.HUANG AND MEHROTRA, S. **Content-based image retrieval with relevance feedback in MARS.** Proceedings of International Conference on Image Processing, volume 2, pages 815-818. 1997.
- [20] SIGGELKOW, S. AND BURKHARDT, H. **Fast invariant feature extraction for image retrieval.** State-of-the-Art in Content-Based Image and Video Retrieval. Kluwer, 2001.
- [21] RUI, Y.; HUANG, T. AND CHANG, S-F. **Image Retrieval: Past, Present, and Future.** In Journal of Visual Communication and Image Representation. 1997.
- [22] WIKIPEDIA. **Thomas Young.** [http://en.wikipedia.org/wiki/Thomas_Young_\(scientist\)](http://en.wikipedia.org/wiki/Thomas_Young_(scientist)).
- [23] STRICKER, M. AND ORENGO, M. **Similarity of Color Images.** In Proceedings of IS&T and SPIE Storage and Retrieval of Image and Video Databases III, San Jose, USA. 1995.

- [24] PASS, G. AND ZABIH, R. **Histogram refinement for content-based image retrieval.** In IEEE Workshop on Applications of Computer Vision, pages 96-102. 1996.
- [25] PASS, G.; ZABIH, R. AND MILLER, J. **Comparing Images Using Color Coherence Vectors.** ACM Conference on Multimedia, Boston, Massachusetts, pages 65-74. 1996.
- [26] HUANG, J., KUMAR, S. R., MITRA, M., ZHU, W., AND ZABIH, R. **Image indexing using color correlogram.** In IEEE International Conference on Computer Vision and Evolutionary Computation and Pattern Recognition, Puerto Rico. 1997.
- [27] LONG, F., ZHANG, H., AND FENG, D. **Fundamentals of Content-Based Image Retrieval.** Multimedia information retrieval and management – technological fundamentals and applications, Springer-Verlag, pages 1-26. 2003.
- [28] GEVERS, T. AND SMEULDERS, A. W. M. **Pictoseek: Combining color and shape invariant features for image retrieval.** IEEE Trans. on image processing, 9(1):102-119. 2000.
- [29] GIMEL FARB, G. **CBIR: Texture Features I. COMPSCI.708.S1.C Lectures.** University of Auckland. 2006. <http://www.cs.auckland.ac.nz/compsci708s1c/lectures/Glect-pdf/2006-CS708GG-07.pdf>
- [30] VORONOI, G. **Nouvelles applications des paramètres continus à la théorie des formes quadratiques.** Journal für die Reine und Angewandte Mathematik, 133:97-178, 1907.
- [31] HARALICK, R. M. **Statistical and Structural Approaches to Texture.** In Proc. IEEE, volume 67, pages 786-804. 1979.
- [32] LAAKSONEN, J. **Analyzing low-level visual features using content-based image retrieval.** In Proceedings of the 7th International Conference on Neural Information Processing, Teajon, Korea. 2000.
- [33] TAMURA, H., MORI, S., AND YAMAWAKI, T. **Textural Features Corresponding to Visual Perception.** IEEE Transactions on Systems, Man and Cybernetics, SMC-8(6):460-473. 1978.
- [34] MALLAT, S. G. **A theory for multi-resolution signal decomposition: the wavelet representation.** IEEE Trans. Pattern Analysis and Machine Intelligence, 11:674-693. 1992.

- [35] LIU, F. AND PICARD, R. W. **Periodicity, directionality, and randomness: Wold features for image modeling and retrieval.** IEEE Trans. on Pattern Analysis and Machine Learning, 18(7):184-189. 1996.
- [36] FU, K. S. **Syntactic Methods in Pattern Recognition.** Academic Press, N.Y. 1974.
- [37] ZHANG, D. **Image Retrieval Based on Shape.** Phd thesis, Faculty of Information Technology, Monash University. 2002.
- [38] LU, G. **Chain code-based shape representation and similarity measure.** Visual Information Systems. Lecture Notes in Computer Science. Springer Berlin / Heidelberg. p. 135-150. 1997.
- [39] ZHANG, D. AND LU, G. **Review of shape representation and description techniques.** Pattern Recognition, 37:1-19. 2004.
- [40] TORRES, R. S. AND FALCÃO, A. X. **Contour Saliency Descriptors for Effective Image Retrieval and Analysis.** Image and Vision Computing. 2006.
- [41] SALTON, G. AND BUCKLEY, C. **Improving retrieval performance by relevance feedback.** Journal of the American Society for Information Science, 41(4):288-297. 1990.
- [42] DOULAMIS, N. AND DOULAMIS, A. **Evaluation of relevance feedback schemes in content based image retrieval systems.** Signal Processing: Image Communication, 21:334-357. 2006.
- [43] CIOCCA, G. AND SCETTINI, R. **A relevance feedback mechanism for content-based image retrieval.** Information Processing and Management, 35:605-632. 1999.
- [44] FRENCH, J. AND JIN, X. **An empirical investigation of the scalability of a multiple viewpoint CBIR system.** In Proceedings of the CIVR, Ireland, pages 252-260. 2004.
- [45] NAKAZATO, M., MANOLA, L., AND HUANG, T. S. **ImageGrouper: Search, Annotate and Organize Images by Groups.** In 5th International Conference VISUAL, pg 129-142, Taiwan. 2002.
- [46] GOLDBERG, D. E. **Genetic algorithms in search, optimization and machine learning.** Addison Wesley. 1989.

- [47] WOOD, J. G. **Nature's teachings: human invention anticipated by nature.** London, W. Glasher. University of California Libraries. 1907.
- [48] MICHALEWICZ, Z. AND FOGEL, D. B. **How to solve it: modern heuristics.** Springer. Verlag. 2000
- [49] HOLLAND, J. H. **Adaptation in natural and artificial systems.** Michigan: MIT Press. 1975.
- [50] OLIVEIRA, G. M. B. **Dinâmica e Evolução de Autômatos Celulares Unidimensionais.** Tese de Doutorado, Instituto Tecnológico de Aeronáutica, Curso de Engenharia Eletrônica e Computação na Área de Informática. 1999.
- [51] SILVA, A. P. A. **Tutorial: Genetic Algorithms. Learning and Nonlinear Models.** Revista da Sociedade Brasileira de Redes Neurais, 1(1):45-60. 2002.
- [52] MICHALEWICZ, Z. **Genetic Algorithms + Data Structures = Evolution Programs.** Springer Verlag. 1992.
- [53] HAUPT, R. L. AND HAUPT, S. E. **Practical Genetic Algorithms.** Wiley-Interscience. 1998.
- [54] SILVA, S. F.; BARCELOS, C. A. Z. AND BATISTA, M. A. **The effects of fitness functions on genetic algorithm applied to relevance feedback in image retrieval.** In: International Conference on Systems, Signals and Image Processing and Semantic Multimodal Analysis of Digital Media, 2006, Budapeste.
- [55] CANTÚ-PAZ, E. **Efficient and Accurate Parallel Genetic Algorithms.** Kluwer Academic Publishers. 2001.
- [56] STEJIĆ, Z., TAKAMA, Y., AND HIROTA, K. **Genetic algorithms for a family of image similarity models incorporated in the relevance feedback mechanism.** Applied Soft Computing, 2:306-327. 2003.
- [57] STEJIĆ, Z., TAKAMA, Y., AND HIROTA, K. **Weighted Local Similarity Pattern as image similarity model incorporated in GA-based relevance feedback mechanism.** Intelligent Data Analysis, 7:443-467. 2003.

- [58] STEJIĆ, Z., TAKAMA, Y., AND HIROTA, K. **Mathematical aggregation operators in image retrieval: effect on retrieval performance and role in relevance feedback.** *Signal Processing*, 85:297-324. 2005.
- [59] SILVA, S. F. **Realimentação de Relevância via Algoritmos Genéticos aplicada na Recuperação de Imagens.** Dissertação de Mestrado, Faculdade de Computação, Universidade Federal de Uberlândia. 2007.
- [60] VISTEX DATABASE. **Vision Texture Database.** Massachusetts Institute of Technology. Media Laboratory. Available in <ftp://whitechapel.media.mit.edu/pub/VisTex/>.
- [61] BRODATZ DATABASE. **USC-SIPI Image Database.** University of Southern California, Signal and Image Processing Institute. Available in <http://sipi.usc.edu/services/database/Databases.html>.
- [62] COREL DATABASE. **Corel Database.** Corel Corporation, Corel Gallery 3.0. Available in James Z. Wang's Research Group: <http://wang.ist.psu.edu/jwang/test1.tar>.
- [63] CALPHOTOS. **CalPhotos Database.** University of California, Berkeley. Available: <http://calphotos.berkeley.edu/>.
- [64] BARCELOS, E. Z.; FLORES, E. L.; BARCELOS, C. A. Z.; SILVA, S. F. AND BATISTA, M. A. **A Multi-Dimensional Similarity Modeling and Relevance Feedback Approach for Content-Based Image Retrieval.** In: International Conference on Systems, Signals and Image Processing, IWSSIP. Chalkida, Greece. 2009.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)