

SIMONE VIEIRA PEREIRA

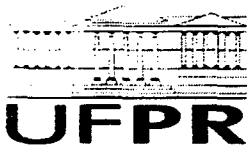
CLASSIFICAÇÃO E AVALIAÇÃO DE SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Informática, Curso de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná.

Orientadora: Prof.^a Olga R. P. Bellon

CURITIBA

2001



Ministério da Educação
Universidade Federal do Paraná
Mestrado em Informática

PARECER

Nós, abaixo assinados, membros da Banca Examinadora da defesa de Dissertação de Mestrado em Informática da aluna *Simone Vieira Pereira*, avaliamos o trabalho intitulado "*Avaliação de Sistemas de Recuperação de Imagens Baseada em Conteúdo: Um Estudo de Caso na Área Médica*", cuja defesa foi realizada no dia 07 de dezembro de 2001, às quatorze horas, no anfiteatro B do Setor de Ciências Exatas da Universidade Federal do Paraná. Após a avaliação, decidimos pela aprovação da candidata.

Curitiba, 07 de dezembro de 2001.

Prof.^a Dra. Olga Regina Pereira Bellon
DINF/UFPR - Orientadora

Prof. Dr. Luis Augusto Consularo
DIN/UEM

Prof. Dr. André Luiz Pires Guedes
DINF/UFPR

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus por estar sempre ao meu lado e me guiar durante todo o período necessário para a realização deste curso, em especial na construção deste projeto. Agradeço aos meus pais e irmãos, que muito me apoiaram e incentivaram durante o período de desenvolvimento deste trabalho. Devo destacar também o apoio, colaboração e incentivo que recebi dos meus colegas de curso e de trabalho.

Devo agradecer também à professora Olga Regina Pereira Bellon pela dedicação, atenção e orientação na elaboração deste trabalho.

SUMÁRIO

1. INTRODUÇÃO	1
1.1. OBJETIVOS	5
1.2. ORGANIZAÇÃO	6
2. EXTRAÇÃO DE CARACTERÍSTICAS.....	7
2.1. COR.....	8
2.2. TEXTURA	9
2.3. FORMA	10
2.4. ESTRUTURA DA COR (<i>LAYOUT</i>).....	11
2.5. CONCLUSÃO	12
3. INDEXAÇÃO MULTIDIMENSIONAL.....	14
3.1. REDUÇÃO DE DIMENSÃO	15
3.2. TÉCNICAS DE INDEXAÇÃO MULTIDIMENSIONAL	16
3.2.1. <i>Índices pictoriais</i>	18
3.2.2. <i>Hot Spots</i>	19
3.2.3. <i>Técnicas textuais</i>	20
3.3. CONCLUSÃO	21
4. MÉTODOS DE CONSULTA.....	22
4.1. BROWSING	22
4.2. CONSULTA POR IMAGEM EXEMPLO	23
4.3. CONSULTA POR ESBOÇO.....	25
4.4. CONSULTA POR TEXTO.....	25
4.5. CONSULTA POR CARACTERÍSTICAS	27
4.6. REALIMENTAÇÃO DE RELEVÂNCIA	28
4.7. CONCLUSÃO	29
5. AVALIAÇÃO	31

5.1. PROBLEMAS BÁSICOS DA AVALIAÇÃO DE SISTEMAS DE RECUPERAÇÃO POR CONTEÚDO.....	38
5.1.1. <i>Definição de um banco de dados de imagens padrão</i>	38
5.1.2. <i>Obtendo julgamentos de relevância</i>	39
5.2. OUTRAS MEDIDAS DE AVALIAÇÃO DE DESEMPENHO.....	41
5.3. REPRESENTAÇÕES GRÁFICAS.....	42
5.4. PROPOSTAS.....	43
6. SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO.....	45
6.1. QBIC.....	45
6.2. VIRAGE.....	46
6.3. PHOTOBOOK.....	48
6.4. VISUALSEEK.....	50
6.5. NETRA.....	52
6.6. MARS.....	53
6.7. SURFIMAGE.....	55
6.8. SAMMATCH.....	56
6.9. EPIC.....	57
6.10. MOSAIC.....	59
6.11. WAVEGUIDE.....	60
6.12. SOM-AIR.....	61
6.13. MARCO.....	63
6.14. PICHUNTER.....	65
6.15. FOCUS.....	67
6.16. FOOTSCAN.....	68
7. CLASSIFICAÇÃO DE SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO.....	70
8. CONCLUSÕES E TRABALHOS FUTUROS.....	82
9. REFERÊNCIAS BIBLIOGRÁFICAS.....	87

LISTA DE FIGURAS

FIG.1: GRÁFICOS PR PARA QUATRO DIFERENTES CONSULTAS: (A) SEM REALIMENTAÇÃO DE RELEVÂNCIA E (B) COM REALIMENTAÇÃO DE RELEVÂNCIA [20].....	34
FIG.2: GRÁFICO PR DOS RESULTADOS DE 100 CONSULTAS AO BANCO DE DADOS DE IMAGENS SINTÉTICAS PRODUZIDOS PELO SISTEMA VISUALSEEK.....	34
FIG.3: GRÁFICO PR DOS RESULTADOS DE TESTES FEITOS COM O SISTEMA MARS.....	35
FIG.4: GRÁFICO PR DOS RESULTADOS DE CONSULTAS AO BANCO DE DADOS VISTEX, PRODUZIDOS PELO SISTEMA SURFIMAGE.....	36
FIG.5: (A) GRÁFICO RECHAMADA VERSUS NÚMERO DE IMAGENS E (B) GRÁFICO PRECISÃO VERSUS NÚMERO DE IMAGENS [20].....	36
FIG 6.: CONSULTA POR FORMA NO SISTEMA QBIC. O USUÁRIO ESCOLHE A FORMA NA TELA À ESQUERDA E OS RESULTADOS APARECEM À DIREITA.....	46
FIG 7.: EXEMPLO DE INTERFACE DO SISTEMA VIRAGE. O RESULTADO DA BUSCA POR PAISAGENS DO PÔR-DO-SOL, ENFATIZANDO SIMILARIDADE POR COR.	48
FIG 8.: EXEMPLO DE INTERFACE DO SISTEMA PHOTOBOOK.	50
FIG 9.: EXEMPLO DE INTERFACE DO SISTEMA VISUALSEEK.	51
FIG 10.: EXEMPLO DE INTERFACE DO SISTEMA SOM-AIR. QUADRO DE CONSULTA.....	63
FIG 11.: EXEMPLO DE INTERFACE DO SISTEMA PICHUNTER.....	67
FIG.12: PRINCIPAIS PECULIARIDADES DOS SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO	79

LISTA DE TABELAS

TAB1.: COMPARAÇÃO ENTRE RECUPERAÇÃO DE DADOS E RECUPERAÇÃO DE INFORMAÇÃO.	24
TAB2.: COLEÇÃO DE SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO	74
TAB3: FORMA SINTÉTICA DA TABELA DE CARACTERÍSTICAS DA COLEÇÃO DE SISTEMAS	81

RESUMO

O aumento das coleções de imagens digitais acarretou um crescimento no uso e desenvolvimento de sistemas de recuperação de imagens por conteúdo (SRIC), bem como nas pesquisas sobre esse assunto. A avaliação de desempenho é um problema essencial da recuperação de imagens baseada em conteúdo. Muitos métodos diferentes para medir o desempenho desses sistemas foram adaptados da recuperação de informação e vêm sendo usados atualmente por muitos pesquisadores. Porém, em muitos casos, esses métodos não são adequados para recuperação de imagens. Este trabalho apresenta algumas contribuições para a evolução da pesquisa sobre os SRIC. Inicialmente, apresenta-se um estudo sobre a base da recuperação de imagens baseada em conteúdo: extração de características, indexação multidimensional e projeto do método de consulta. Em seguida, apresenta-se uma proposta original para classificação de SRIC. Essa classificação é baseada no estudo de uma coleção de sistemas desta natureza publicados na literatura. Finalmente, apresenta-se uma análise sobre os métodos existentes para avaliação de SRIC.

ABSTRACT

Increase of digital image collections lead to enhancement of use and development of content-based image retrieval systems (CBIRS), as well as research about this subject. Evaluation of retrieval performance is a crucial problem in CBIRS. Many different methods for measuring performance of CBIRS have been adapted from information retrieval, and are currently used by researchers. However, in many cases, these methods are not suitable for image retrieval. This work presents some contributions to evolve research about CBIRS. First of all, the guidelines of content-based image retrieval are discussed: features extraction, multidimensional indexing and search methods. Then, an original proposal for classification of CBIRS is presented. This classification is based on analysis of many systems of this nature published in the literature. Finally, an analysis of the evaluation methods for CBRIS is presented.

1. INTRODUÇÃO

O termo Recuperação de Informação, ou IR (*Information Retrieval*), descreve o processo através do qual um usuário converte uma consulta em uma coleção útil de referências [4]. O autor desta definição, Calvin Moores, refere-se à informação textual. Em um sistema de IR, um documento é representado como uma coleção de características, como palavras-chaves, citações, referências bibliográficas [37]. O usuário especifica a informação de que precisa através de uma consulta. Dada a consulta do usuário e uma coleção de documentos, o sistema estima a probabilidade de cada documento da coleção ser relevante à consulta.

Na verdade, este conceito pode ser estendido para recuperação de informações visuais. O interesse em adaptar este conceito a imagens surgiu devido à velocidade com que os avanços na área computacional ocorrem. A facilidade de captura e compressão de imagens digitais tem produzido uma quantidade gigantesca de informação visual on-line [26] e, conseqüentemente, aumentado em muito a popularidade dos bancos de dados de imagens. Com isso, nos últimos anos observou-se um rápido aumento no tamanho das coleções de imagens digitais. Tradicionalmente, a indexação destas imagens tem sido feita de forma manual. Atualmente, a análise do conteúdo multimídia e a indexação baseada no conteúdo visual têm sido áreas de muitos estudos.

Com todos estes avanços, é evidente a necessidade de indexar, detectar e processar toda esta informação de tal maneira que se consiga sua consulta e recuperação eficientes [26]. Com este objetivo, foram criados os sistemas de recuperação de imagens (SRI). O problema da recuperação de imagens é um caso particular da área de manipulação e recuperação de informação.

Uma abordagem muito utilizada para recuperação de imagens foi baseada em texto. A idéia era fazer anotações em forma de texto sobre as imagens e usar gerenciadores de bancos de dados convencionais para fazer a recuperação. Esta

primeira solução deve-se, em parte, à evolução da recuperação no campo textual, com o desenvolvimento de sistemas capazes de recuperar a informação de forma eficiente, como o Lycos, AltaVista e Yahoo [26].

Logo, duas dificuldades se apresentaram: a necessidade de muito esforço manual para representar imagens em forma de texto; e outra que resulta do rico conteúdo nas imagens e da subjetividade da percepção humana [5]. Isto é, pessoas diferentes podem perceber características diferentes pertencentes ao conteúdo de uma mesma imagem. A subjetividade da percepção e a imprecisão da anotação podem causar falhas graves nos processos de recuperação.

Antes dos anos 90, por causa da emergência em grande escala das coleções de imagens, as duas dificuldades citadas tornaram-se cada vez mais agudas. A área de Recuperação de Imagens por Conteúdo surgiu para tentar superar estas dificuldades [5]. Isto é, ao invés de serem anotadas por palavras-chaves, as imagens deveriam ser indexadas pelo seu próprio conteúdo visual, tais como cores, textura e outras características. Assim, o problema da recuperação de imagens é um aspecto particular do tratamento geral de recuperação de informação [9]. Desde então, muitas técnicas nesta direção foram desenvolvidas e muitos sistemas de Recuperação de Imagens foram construídos.

Sistemas de bancos de dados convencionais oferecem um método de recuperação de imagens chamado Consulta Direta (*Direct Consult*). Neste método uma consulta sobre imagens consiste apenas de informações textuais sobre as imagens já que dados pictoriais não podem ser processados por sistemas de bancos de dados convencionais por duas razões:

- **Estrutura de Dados:** os campos usados na consulta devem ser do mesmo tipo dos dados armazenados. Conseqüentemente, o usuário deveria fazer a pergunta usando a própria imagem que está procurando.
- **Combinações exatas:** uma consulta direta divide uma coleção de itens de dados em dois grupos. Um contém o dado que obedece às condições da busca e outro com

os dados que não obedecem. Entretanto, não existe um intervalo definido de imagens, tal como 1 a 5 para números e *a* a *z* para alfanuméricos. Isto faz com que a resposta à consulta seja do tipo discreto, ou seja, o sistema responderá apenas se o banco de dados possui ou não a imagem procurada.

Ainda há a necessidade de se representar imagens como abstrações do mundo real, mas a conversão dos objetos do mundo real para uma imagem digital é um processo de baixo nível, que não inclui a semântica da imagem.

A percepção de imagens é um processo muito complexo. Mas, foi descoberto [1] que células nervosas na retina imediatamente separam imagens em componentes, como contornos, texturas e cores. Estes fragmentos de informação são reunidos no cérebro em uma única imagem, de onde a informação sobre seus conteúdos é extraída.

Os computadores têm a capacidade de decompor imagens em componentes de textura, formas, contornos e cores. Porém, são muito limitados em sua capacidade de reintegrar estas características em uma imagem útil ao usuário. Sem o entendimento da imagem, um computador ou sistema automático não pode apresentar o nível de organização de imagem que é natural ao ser humano.

Uma interpretação automática de imagens pelo computador seria a solução ideal para o problema, apesar de ser unânime a idéia de que o conhecimento total sobre o assunto está longe de ser obtido. Na ausência de tal entendimento, a extração automática de algumas características, como a cor por exemplo, pode se tornar capaz de fazer com que o armazenamento de imagens em sistemas de recuperação seja cada vez mais adequado a este tipo de informação. Muitas abordagens são constantemente avaliadas de forma a usar a informação cor extraída automaticamente para filtrar coleções de imagens coloridas em bancos de dados de imagens [1].

Novas ferramentas e técnicas são necessárias para a eficiência da busca *on-line* e a recuperação de imagens e vídeo. Indexação por cores é um processo pelo qual as imagens e vídeos no banco de dados são recuperados com base nas cores que possuem. Um sistema de indexação de cores requer que muitos objetivos importantes

sejam satisfeitos, como: extração automática de cores, indexação e recuperação eficientes. Na prática, é difícil conseguir projetar um sistema que atinja a todos estes objetivos. Alguns sistemas [2] já foram propostos para indexação por cores com extração automática de regiões coloridas, indexação e consulta com desempenho satisfatório ao usuário.

Em uma recuperação típica de imagens por conteúdo, o usuário tem uma imagem em que tem interesse e quer encontrar imagens similares no banco de dados. O cenário aqui envolve uma consulta com uma imagem de exemplo. O usuário entra com uma imagem ou um pedaço de uma imagem e quer recuperar imagens do banco de dados que tenham alguma semelhança com a imagem de entrada. O primeiro problema é encontrar características adequadas para a representação da imagem. Então, deve-se usar uma medida eficiente para estabelecer similaridade entre duas imagens.

Algumas técnicas têm sido utilizadas para auxiliar o processo de busca, como realimentação de relevância [8], que é o processo pelo qual o usuário interage com o sistema, refinando sua busca.

Muitos pesquisadores [20] desta área estão voltando seus estudos para a avaliação dos sistemas existentes, já que trabalhos neste sentido podem direcionar o desenvolvimento de novos sistemas [5]. Atualmente, a avaliação é feita com base em precisão e rechamada, que são critérios definidos para recuperação de informação e adaptados para recuperação de imagens. Rechamada é a proporção de imagens relevantes no banco de dados que são recuperadas em resposta à consulta. Precisão é a proporção das imagens recuperadas que são relevantes à consulta [5].

Existem três bases fundamentais para Recuperação de Imagens por Conteúdo: Extração de Características, Indexação Multidimensional e Projeto de Sistemas de Recuperação. Assim, várias características visuais e suas representações, e ainda técnicas relacionadas foram estudadas. Para facilitar a busca rápida em coleções de imagens de grande escala, técnicas de indexação eficientes são necessárias.

1.1. OBJETIVOS

Um dos objetivos deste trabalho é estudar sistemas de recuperação de imagens por conteúdo, destacando peculiaridades importantes, como as características escolhidas para representar a imagem, os métodos de consulta suportados pelo sistema, a estrutura usada para armazenar a representação das imagens, os métodos de indexação e de avaliação aplicados, o uso da realimentação de relevância e ainda as categorias das imagens dos bancos de dados usados para testes. Este estudo permitirá a elaboração de uma classificação dos sistemas estudados. Além disso, será possível identificar caminhos adequados para se desenvolver sistemas de recuperação de imagens por conteúdo, dependendo da classificação em que se enquadra.

Outro objetivo deste trabalho é o levantamento de critérios de avaliação de sistemas de recuperação de imagens por conteúdo existentes e a elaboração de um conjunto de critérios que seja eficiente para avaliar o sistema FootScan, que está sendo desenvolvido por uma equipe de pesquisa da Universidade Federal do Paraná – UFPR, incluindo profissionais do Departamento de Informática – DINF – e do Hospital de Clínicas – HC. Este conjunto de critérios deve orientar as pesquisas realizadas nesta área para uma direção que leve ao melhoramento do sistema. Com base na avaliação deste sistema, poderão ser construídos outros sistemas mais eficientes e práticos.

Para que estes objetivos sejam alcançados, será realizado um estudo sobre os métodos de recuperação de imagens existentes, incluindo conceitos envolvidos em extração de características, indexação multidimensional e projeto de sistemas de recuperação. A intenção é direcionar as pesquisas ao desenvolvimento de novos sistemas, mais eficientes. Também serão coletados e estudados critérios de avaliação já existentes para sistemas de recuperação de texto ou outros tipos de recuperação de informação. Por fim, será estudada a forma de adaptação destes critérios a sistemas que lidem especificamente com informação visual.

1.2. ORGANIZAÇÃO

Os próximos capítulos deste trabalho apresentarão os três conceitos fundamentais para Recuperação de Imagens por Conteúdo: Extração de Características, Indexação Multidimensional e Projeto de Sistemas de Recuperação. Estes capítulos apresentam trabalhos relacionados a alguns itens de projeto necessários ao desenvolvimento de um sistema de recuperação de imagens por conteúdo. O capítulo 2 descreve as características gerais usadas em sistemas desta área: cor, forma e textura. O capítulo 3 diz respeito à indexação que deve auxiliar no processo de recuperação e apresenta técnicas para lidar com multidimensionalidade. O capítulo 4 apresenta as várias maneiras de se projetar a interface de um sistema e destaca as várias formas de consulta quando o tipo de informação desejada pelo usuário é visual.

Na seqüência, o capítulo 5 apresenta um estudo sobre os métodos de avaliação dos sistemas existentes e ainda um levantamento dos problemas encontrados nesta área. Uma coleção de sistemas de recuperação de imagens por conteúdo desenvolvidos comercial e cientificamente e suas principais peculiaridades são destacados no capítulo 6. O capítulo 7 apresenta uma proposta de classificação para os sistemas estudados e apresentados anteriormente.

O capítulo 8 apresenta um estudo sobre sistemas de recuperação de imagens médicas. Este estudo se fez necessário por ser essa a natureza do sistema FootScan, que é objeto de estudo deste trabalho. No capítulo 9, é feita uma apresentação e uma proposta de avaliação para o sistema FootScan, com base nos estudos do capítulo 5. Com base na situação corrente e nas necessidades atuais de aplicações para o mundo real, direções a serem tomadas no futuro e abordagens sugeridas são apresentadas no capítulo 10.

2. EXTRAÇÃO DE CARACTERÍSTICAS

Em recuperação por conteúdo, a primeira dificuldade é encontrar as características mais adequadas para a representação da imagem. Depois, deve-se usar uma medida que seja capaz de estabelecer similaridade entre duas imagens. As características e a medida de similaridade devem ter a capacidade de identificar imagens similares de forma satisfatória e ainda ser hábeis para discriminar as não similares.

Extração de características é o principal fundamento da recuperação de imagens por conteúdo. Em um sentido amplo, características podem incluir as que são baseadas em texto (palavras-chave, anotações, ...) e as visuais (cor, textura, forma, faces, ...). Mas, já existe uma ampla literatura em extração de características baseadas em texto nas áreas de gerenciadores de bancos de dados. Por esse motivo, a atenção agora está voltada às técnicas de extração de características visuais.

Dentro do escopo de características visuais, as características podem ser classificadas como características gerais e de domínio específico. A primeira inclui cor, textura e forma, enquanto a segunda é dependente da aplicação e pode incluir, por exemplo, faces humanas e impressões digitais. As características específicas do domínio são mais bem estudadas nas literaturas de reconhecimento de padrões. Assim, o enfoque aqui será dado às características gerais que podem ser usadas na maioria das aplicações. O principal motivo que levou a esta escolha é o interesse em avaliar sistemas de propósitos gerais, para que pudesse ser elaborada uma classificação destes sistemas.

Por causa da subjetividade da percepção, não existe uma única e melhor representação para uma certa característica. Podem existir muitas representações geradas a partir das diferentes perspectivas que sejam adequadas à mesma característica escolhida.

2.1. COR

Várias características visuais são usadas em técnicas de recuperação de imagens [1], mas cor é talvez a mais intuitiva e mais simples para o usuário. Acredita-se [1] que estudos sobre esta característica ainda podem provocar muitas mudanças neste cenário. Isto porque a extração de cores automatizada pelo computador ainda não é capaz de fazer referências ao contexto [26]. O que pessoas naturalmente vêem e classificam como cores específicas, ou como cores padrão, usualmente aparece distorcido para o computador devido a fatores como sombra de outros objetos e iluminação. Outros fatores que podem alterar a aparência da cor do mundo real são a textura da superfície e as condições de observação e captura [26].

Outros fatores fazem com que a cor seja uma das características visuais mais usadas em recuperação de imagens [5]. É relativamente robusta ao problema de *background*, ou seja, à distinção entre os objetos da imagem e o pano de fundo da mesma. Além disso, o estudo das cores de uma imagem não depende do tamanho e da rotação da mesma.

Em recuperação de imagens, histograma de cores é a representação mais usada desta característica. Estatisticamente, denota a probabilidade da intensidade dos três canais de cores. O histograma de cores parte do princípio que uma região de uma imagem não tem muitas cores diferentes em destaque [31]. Esta abordagem é muito utilizada porque o processo de extração do histograma de uma imagem não é complexo e o cálculo da similaridade é rápido [37].

A idéia básica de se usar um histograma é reduzir o espaço de cores possíveis de modo que as características possam ser organizadas adequadamente para permitir que consultas sejam elaboradas futuramente [34].

Enquanto o computador pode gravar precisamente a informação cor em valores do pixel, a percepção humana de cor não é um tipo de processamento pontual. O sistema visual humano automaticamente coleta pontos de cor para formar áreas, superfícies e objetos. Além disso, existem evidências de que a percepção de cor é

influenciada pelas cores que a circundam [2]. Para que a extração seja aceitável, o computador deve combinar entidades pixels de maneira a obter e representar regiões e padrões de cores satisfatórios ao homem. Vale salientar que o método de extração adequado depende do contexto da aplicação.

Para este objetivo foram propostas várias técnicas [1] pelas quais a cor das imagens e vídeos é extraída automaticamente para formar uma classe de metadados que é facilmente indexada. A utilidade destes sistemas pode aumentar em muito se eles puderem prover a habilidade de busca diretamente em dados não textuais, em vez de buscar somente a informação textual associada.

2.2. TEXTURA

Textura refere-se aos padrões visuais que têm propriedades de homogeneidade que não resultam da presença de apenas uma única cor ou intensidade [5]. A textura é um elemento importante na visão humana, provendo em uma cena a profundidade e orientação da superfície [26]. É uma propriedade inata de todas as superfícies, incluindo nuvens, árvores, tijolos, etc. Contém informações importantes sobre a disposição estrutural de superfícies e seus relacionamentos com o ambiente ao redor. Por causa de sua importância e utilidade para o Reconhecimento de Padrões e Visão Computacional, foram realizadas pesquisas com resultados importantes nas últimas três décadas. Agora, ela tem sido mais estudada e enfatizada na Recuperação de Imagens.

A textura é também uma característica muito importante para a classificação. Em muitos estudos é considerada como a distribuição espacial de níveis de cinza na vizinhança [3]. Dependências espaciais de níveis de cinza trazem mais informação sobre textura e são mais gerais que outros métodos e mais adequadas a várias aplicações.

Para representar textura, foi proposta uma abordagem que usa matriz de co-ocorrência [3], que explora a dependência do nível espacial de cinza. Primeiro foi

construída a matriz de co-ocorrência baseada na orientação e na distância entre os pixels da imagem e então extraídas estatísticas significativas da matriz como a representação da textura.

Para extrair esta característica de imagens existem duas abordagens principais: a estrutural, que usa a idéia de que textura é composta por primitivas com propriedades diferentes aparecendo em planos particulares, e a abordagem estatística, que considera distribuições estatísticas no domínio espacial. Outra possibilidade é usar uma combinação das duas, considerando que textura é a distribuição estatística das propriedades de diferentes primitivas que ocorrem em diferentes relacionamentos espaciais [3].

Baseado em estudos psicológicos da percepção visual humana, outro pesquisador [6] desenvolveu aproximações computacionais para as propriedades visuais da textura. Algumas das propriedades foram: suavidade, contraste, direcionalidade, regularidade e aspereza. A aspereza mede a escala de uma textura (se é formada por pequenos pedregulhos ou grandes rochas), o contraste descreve a quantidade de brilho, e a direcionalidade descreve se ela tem uma direção predominante (como o capim) ou não (como um objeto liso) [33].

Todas as propriedades da textura nesta representação são visualmente significativas, o que não ocorre necessariamente na matriz de co-ocorrência e a torna muito atrativa para a Recuperação de Imagens.

Nos anos 90, muitos pesquisadores [38], [39] começaram a estudar o uso da transformada de Wavelet na representação de textura. Esta abordagem foi combinada com outras várias vezes para se conseguir melhor aproveitamento da técnica e melhores resultados.

2.3. FORMA

Quando são usadas técnicas de consulta visual, ao contrário de especificação textual, as formas precisam ser desenhadas pelo usuário de maneira que sejam

semelhantes às armazenadas no banco de dados. Assim, forma é um critério importante para identificação de objetos com base em seu perfil e estrutura física [26]. Algumas aplicações ainda requerem que a representação da forma seja invariante para translação, rotação e escalamento [5].

Em geral, as representações de forma podem ser divididas em duas categorias, baseada em bordas e baseada em regiões [5] e [48]. A primeira lida somente com a borda externa da forma e representa o contorno dos objetos com base nas mudanças elementares que sofre. Neste caso, as características podem ser armazenadas em uma estrutura fixa, com um número fixo de vértices representando cada objeto e organizadas em uma árvore, onde cada nó é inserido seguindo diferentes critérios de similaridade. Esta árvore age como um índice, no qual as folhas são as próprias características [9].

A outra abordagem, baseada em regiões, usa o interior de cada região. Na construção do banco de dados, para cada imagem são armazenadas áreas retangulares correspondendo a objetos na imagem e os relacionamentos espaciais entre elas [9]. Na fase da consulta, os desenhos das formas na tela e seus relacionamentos espaciais são gravados. Uma imagem candidata é recuperada se tem duas ou mais áreas retangulares no mesmo relacionamento que as formas da consulta.

2.4. ESTRUTURA DA COR (*LAYOUT*)

Embora a característica cor global seja simples para calcular e tenha um poder de discriminação razoável em Recuperação de Imagens, ela tende a fornecer muitas imagens não similares como resposta no caso de coleções de imagens muito grandes. Muitas pesquisas sugerem que usar estrutura de cor (característica cor e seus relacionamentos espaciais) é uma solução melhor para recuperação de imagens [1]. Para estender uma característica cor global para uma local, uma abordagem natural é dividir a imagem em sub-blocos e extrair a cor de cada sub-bloco. Embora conceitualmente simples, esta abordagem pode não fornecer informação precisa sobre

cor local e é muito custosa em termos de processamento e armazenamento. Uma abordagem mais sofisticada é segmentar a imagem em regiões com características salientes de cor pela retro-projeção de conjuntos de cores e armazenar a posição e a característica conjunto de cor de cada região para suportar consultas posteriores. A vantagem é o aumento da precisão e a desvantagem é a necessidade de algoritmos de segmentação, o que ainda é um problema da área de Processamento de Imagens.

2.5. CONCLUSÃO

Depois de extraídas as características gerais de cada imagem a ser armazenada no banco de dados, são criadas representações para estas imagens. Esta representação pode ser chamada de imagem simbólica [41]. Mas, o termo imagem lógica foi estabelecido como um termo padrão para a informação simbólica que é extraída do dado imagem crua e informação textual auxiliar [24]. Conseqüentemente, o termo banco de dados lógico refere-se ao banco de dados que gerencia imagens lógicas. E ainda o termo imagem física refere-se ao arquivo gerado através do método de aquisição da imagem propriamente dita.

A idéia de representar imagens físicas por imagens lógicas ajuda a obter independência do domínio e é similar à representação de documentos por palavras-chaves em sistemas de informação bibliográfica [41].

Várias abordagens foram exploradas para derivar imagens lógicas de imagens físicas e organizá-las em bancos de dados lógicos. Chang [25] propôs o uso de strings de duas dimensões para a indexação de imagens baseada em relacionamentos espaciais e atributos dos objetos na imagem. Este método foi estendido por outros pesquisadores. Em todos estes métodos, primeiro a imagem é segmentada para destacar o contorno dos objetos na imagem [24]. O usuário pode interagir para selecionar e modificar o contorno dos objetos dominantes. Atributos apropriados do objeto tais como forma, orientação, tipo de borda, área, localização, textura, etc, podem então ser calculados. Estes atributos, junto com a relação espacial

de objetos, são armazenados no banco de dados lógico.

Atualmente, estas representações são armazenadas em forma de vetores multidimensionais [5]. Para que o sistema de busca seja satisfatório, deve-se criar um índice usando estas representações.

3. INDEXAÇÃO MULTIDIMENSIONAL

Um sistema de armazenamento e recuperação de imagens suporta anotações textuais e indexação. Um índice é uma representação sintética dos documentos [9]. Normalmente, em sistemas de recuperação de imagens, a informação relevante em cada imagem é codificada na forma de um vetor de características multidimensional. Isso facilita as consultas dos usuários, mas a indexação desses vetores degrada o desempenho da busca quando o número de dimensões do vetor é alto.

Uma idéia muito promissora de indexação para fazer com que a busca fique mais rápida em bancos de dados tradicionais e multimídia é mapear objetos em pontos no espaço multidimensional, usando funções de extração de características, que dependem do domínio específico [23]. Assim, é possível usar bons métodos de acesso espacial (SAMs – *Spatial Access Methods*), para elaborar muitos tipos de consultas, incluindo a consulta por exemplo.

Em [23], os autores sugerem o uso de mineração de dados (*datamining*) como uma boa ferramenta para auxiliar o processo de indexação, já que os objetos podem agora ser vistos como pontos do espaço 2-D ou 3-D, revelando agrupamentos potenciais, correlações entre atributos e outras regularidades que a mineração de dados procura.

A criação e a manutenção destes índices para coleções grandes são muito custosas e demoradas. No caso de bancos de dados heterogêneos, considerando que heterogeneidade refere-se às diferenças no significado e no uso dos dados, o que torna difícil identificar os relacionamentos entre os objetos, não se pode permitir um alto grau de interação com o usuário ou que ele faça uma revisão manual da busca. Além disso, indexar texto para imagens somente provê uma resposta discreta à busca, ou seja, sucesso ou fracasso. Se o usuário não especificar de maneira correta as palavras-chave, as imagens desejadas podem nunca ser encontradas.

Felizmente, este problema pode ser melhorado permitindo que o computador

tenha suporte para extração automática de características visuais. Neste caso, o computador analisa cada imagem e extrai algum tipo de informação, tal como cores, texturas e formas. Embora o índice para informações do tipo texto seja criado por um especialista humano, o índice de cores, por exemplo, é gerado automaticamente pelo computador.

Para fazer com que a Recuperação de Imagens por conteúdo seja realmente aplicável para grandes coleções de imagens, é preciso explorar as técnicas de indexação multidimensional. Para resolver os problemas que envolvem este tema, uma abordagem promissora é primeiro reduzir a dimensão e então usar as técnicas de indexação multidimensional apropriadas.

3.1. REDUÇÃO DE DIMENSÃO

Embora a dimensão dos vetores de características em Recuperação de Imagens seja normalmente muito alta, a dimensão significativa é muito baixa. Antes de usar alguma técnica de indexação, é conveniente reduzir a dimensão. Existem duas abordagens bastante conhecidas na literatura: a transformada de Karhunen-Loève (KLT – *Karhunen-Loève Transform*) e agrupamento com relação a coluna.

A KLT, sua versão no reconhecimento de faces e sua versão na análise de informações foram estudadas por pesquisadores para realizar redução de dimensão [5]. Resultados destas e outras pesquisas mostraram que a maioria dos dados reais pode ser consideravelmente reduzida em dimensão sem degradação significativa na qualidade da recuperação [5]. Considerando que o sistema de recuperação de imagens é um sistema dinâmico e novas imagens são continuamente adicionadas à coleção, uma atualização dinâmica da estrutura de índices é necessária.

Agrupamento é outra ferramenta poderosa para redução de dimensão. Esta técnica é usada em várias disciplinas tais como reconhecimento de padrões, análise da fala, recuperação de informação, etc. Este tipo de agrupamento é chamado agrupamento com relação a linha. Mas, agrupamento também pode ser usado com

relação a coluna para reduzir a dimensionalidade do espaço de características. Experimentos mostram que esta é uma abordagem simples e eficiente [5].

Além dessas técnicas para redução de dimensão dos vetores, outras já foram testadas, como DFT (*Discret Fourier Transform*) e SVD (*Singular Value Decomposition*) [7]. O banco de dados usado para os testes é composto por imagens de textura. A formulação de uma consulta é feita usando-se uma imagem como exemplo. É montado um vetor de características para cada imagem, consistindo de uma tupla de valores. A imagem passa a ser um ponto no espaço de características e, se usada para uma consulta, a resposta deve apontar para seus vizinhos.

A partir dos testes, concluiu-se que apesar do SVD geralmente proporcionar melhor desempenho que a técnica da DFT, esta última tem um desempenho satisfatório em um intervalo de tempo razoável para navegar em bancos de dados de imagens grandes.

Outro trabalho que usou o SVD [23] concluiu que esta abordagem e também a KLT oferecem uma ótima solução para comparações envolvendo características das imagens determinadas previamente, mas não são aplicáveis para casos em que é considerada a medida de similaridade entre as imagens, de uma maneira mais geral.

O problema é que a redução da dimensão pode ser perigosa, já que informações podem ser perdidas. Para evitar que a redução seja cega e haja perda de informações, é necessário que se faça uma verificação depois do processo de redução.

3.2. TÉCNICAS DE INDEXAÇÃO MULTIDIMENSIONAL

Depois de identificar a dimensão embutida dos vetores de características, é preciso selecionar algoritmos adequados de indexação multidimensional para indexar os vetores, que continuam de alta dimensão. Três comunidades contribuem nesta área: geometria computacional, gerenciamento de bancos de dados e reconhecimento de padrões. Os índices devem ser significativos, discriminantes e utilizáveis [26].

As técnicas de indexação multidimensionais mais populares incluem o

algoritmo de Bucketing, árvore k-D, árvore k-D de prioridades, quadtree e outras árvores [5]. Além disso, agrupamentos e redes neurais são muito usados em reconhecimento de padrões e são técnicas muito promissoras.

A indexação consiste em associar uma descrição a cada imagem para facilitar sua recuperação. Normalmente, envolve três passos [9]:

1. Entidades ou características relevantes são extraídas das imagens e usadas como chaves para o índice;
2. Uma representação é escolhida para as chaves;
3. Eventualmente, um significado específico é associado às chaves.

Existem muitas maneiras diferentes de se representar as características extraídas de uma imagem. Uma representação pode variar desde anotações textuais até um ícone, por exemplo, produzido por um módulo de análise automático [9]. A escolha da melhor representação está associada ao contexto da aplicação, ou seja, vai depender do interesse do usuário. Conseqüentemente, o método de indexação vai variar também de acordo com o tipo de consultas que serão elaboradas. Em alguns casos, a indexação requer o uso de técnicas de reconhecimento de padrão [9]. O problema da indexação geral para imagens ainda não foi resolvido, ou seja, ainda não existe um método de indexação automática que possa ser aplicado a qualquer banco de dados de imagens.

A maioria dos índices de conteúdo de imagens usa representações textuais, características específicas dos objetos contidos nas imagens ou ainda características extraídas diretamente das imagens (semânticas ou sintáticas) [9]. Se o objetivo é permitir respostas em tempo real às consultas, as técnicas de indexação adotadas devem ser escolhidas com prudência considerando-se o tamanho do banco de dados.

Os índices baseados no conteúdo das imagens podem ser classificados em: *pictorial*, *hot spots* e *textual* [9]. Os índices pictoriais ainda podem ser classificados em sintáticos ou semânticos.

3.2.1. Índices pictoriais

Um dos principais problemas para montar índices para recuperação de imagens é a extração de chaves pictoriais, tanto na fase de inserção da imagem no banco de dados quanto na fase de consulta. Uma das tarefas mais difíceis durante a extração de chaves é identificar automaticamente objetos nas imagens, o que requer que o projeto dos algoritmos de segmentação seja bem elaborado. Por isso, alguns sistemas usam técnicas manuais ou semi-automáticas para extração de índices, baseados no conhecimento específico do domínio da aplicação [9].

Outro problema difícil é extrair descrições semânticas das imagens. É possível, automaticamente, encontrar imagens no banco de dados que tenham objetos com forma semelhante à da consulta. Mas, é difícil encontrar imagens que contenham algum tipo de objeto específico.

Índices pictoriais podem ser classificados como sintáticos ou semânticos, dependendo no tipo de informação contida na estrutura de dados [9].

Um bom exemplo de estrutura de dados hierárquica [9] que pode ser usada como um índice sintático é a quadtree. Esta estrutura resulta da decomposição recursiva de uma imagem em quadrantes, baseada em um critério que depende da aplicação e é muito usada em compressão de imagens em Sistemas de Informação Geográfica (GIS – *Geographic Information Systems*) e no cálculo de propriedades geométricas de regiões das imagens. Uma desvantagem desta abordagem é que não lida com informações sobre os objetos das imagens, tais como tamanho, localização e forma.

As características cor, textura e forma dos objetos podem ser usadas para a construção de índices sintáticos. O sistema QBIC [33] permite consultas por conteúdo usando estas características. O índice sintático usado para organizar o espaço multidimensional de características é baseado em outra estrutura, a árvore R^* .

As características citadas até aqui ignoram o significado do conteúdo das imagens, mas para uma consulta inteligente é necessária a interpretação semântica.

Alguns sistemas utilizam técnicas com regras de reescrita [9] para classificar objetos automaticamente em diferentes níveis. O princípio usado é o fato de que um conjunto de símbolos pode ser reescrito em outro símbolo se está embutido no mesmo contexto e se um certo predicado é satisfeito.

As semânticas contam com o fato de que as regras aplicadas para substituir estruturas de baixo nível por outras de alto nível, e calcular os atributos, dependem do contexto. Isto significa que para escolher um significado para ser atribuído a uma estrutura é preciso considerar o domínio da aplicação.

3.2.2. Hot Spots

Um refinamento de técnicas de indexação é representado pelo conceito de imagem inteligente (*smart*) [9], na qual a informação relevante deve ser inicialmente selecionada através da detecção de regiões de interesse dentro da imagem. Serão criadas ligações baseadas no conteúdo entre as regiões (*hot spots*).

Isso permite a criação de operações de indexação através de ícones e protocolos flexíveis para um acesso inteligente aos dados relacionados. Estes protocolos são baseados nas seqüências de operações normalmente realizadas por um usuário trabalhando com o banco de dados no contexto da aplicação, com a capacidade de ignorar ações que não são necessárias ao processo. De acordo com estes protocolos, é possível prever, por exemplo, quais imagens relacionadas o usuário poderia precisar e realizar uma pré-recuperação para reduzir o tempo de resposta.

Quando um *hot spot* é definido ou modificado em uma imagem, esta operação é refletida para todas as imagens conectadas a ela. O modelo computacional para este processo é a rede de Petri [9].

O ambiente inteligente permite a otimização de recursos informativos e oferece uma ferramenta flexível para consultar e gerenciar bancos de dados. Segundo [9], esta abordagem foi testada com imagens médicas, para detectar patologias em vasos sanguíneos.

3.2.3. Técnicas textuais

Em geral, é difícil construir um módulo automático para criar índices pictoriais para recuperação de imagens. Uma possibilidade é usar a informação contida na imagem, sem depender dos resultados do processamento das imagens. Existe um tipo de indexação simbólica [9], baseada na descrição do conteúdo das imagens que lida com um conjunto de atributos em forma de texto, que pode ser administrada por um sistema de gerenciamento de banco de dados convencional. No sistema QBIC [33] palavras-chave são associadas às imagens do banco de dados.

Os problemas desta abordagem estão relacionados à subjetividade das imagens. Normalmente, as imagens contêm muitos objetos diferentes e descrever uma imagem através de um número limitado de palavras-chave torna-se uma tarefa muito difícil. Além disso, pessoas diferentes podem associar palavras-chave diferentes à mesma imagem, o que poderia fazer com que uma consulta não tivesse a resposta esperada. Outro problema é representar certos tipos de características que poderiam ser importantes para determinadas aplicações, como textura ou forma de objetos. Existem trabalhos que comprovam a melhoria nos resultados de recuperação quando a indexação textual é combinada com uma indexação baseada em conteúdo [9].

Técnicas tradicionais vetoriais e booleanas para indexação, hábeis para indicar a presença ou ausência de um termo no banco de dados, não são suficientes quando é necessário preservar uma noção dos relacionamentos entre conceitos, especificamente relacionados ao domínio da aplicação. Neste caso, o problema envolve interpretação semântica de um texto. O conhecimento sobre o domínio da aplicação é representado por um grafo semântico, no qual os nós são classes de conceitos e os arcos são relacionamentos entre essas classes. O problema é comparar estes grafos no momento da consulta, o que pode causar uma sobrecarga computacional.

3.3. CONCLUSÃO

O principal objetivo das pesquisas sobre indexação é construir índices diretamente das características dos objetos presentes na imagem. O desafio é o desenvolvimento de mecanismos para a automação da indexação, que começaria com a extração automática de conteúdo e faria uso de possíveis regras do conhecimento junto com a informação contextual relevante, permitindo uma identificação automática ou inferência de conteúdo complexo [26]. A estrutura destes índices permitiria uma busca eficiente. Isto é um problema devido à estrutura complexa das características extraídas, o que torna difícil evitar que uma busca acabe percorrendo todo o banco de dados.

Depois de extraídas as características das imagens a serem armazenadas e de ser criada uma estrutura de índice para suportar uma busca pelo conteúdo destas imagens, é necessário criar mecanismos de busca para que o usuário tenha acesso ao banco de dados. Vários sistemas já foram criados e suportam diferentes tipos de ferramentas para interagir com o usuário. O próximo capítulo apresenta algumas maneiras de se projetar uma interface com o usuário.

4. MÉTODOS DE CONSULTA

Na última década, recuperação de imagens por conteúdo se tornou uma área muito ativa. Muitos sistemas de recuperação de imagens, tanto comerciais quanto científicos, foram criados [33], [4], [11], [31], [34], [10], [12], [13], [15], [16], [17], [21], [22], [32], [35]. A maioria destes sistemas suporta uma ou mais destas opções:

- browsing randômico,
- consulta por imagem exemplo,
- consulta por esboço,
- consulta por texto (incluindo palavras-chave ou fala),
- consulta por características.

4.1. BROWSING

Os métodos para especificar uma consulta a bancos de dados usam diferentes tipos de abstração, têm formas distintas de avaliar uma consulta e várias formas de entrada de dados. Mas, todos estes métodos têm o mesmo modo de navegação, isto é, a proposta dos métodos é localizar informação. O usuário define a informação necessária, de forma que o sistema de recuperação possa buscá-la. Este modo de navegação é chamado *browsing*, onde o usuário não procura por uma imagem específica.

Browsing é uma busca *ad hoc* de informação por usuários humanos. As três principais características deste tipo de navegação são:

- **É uma forma altamente interativa de recuperação de informação.** A regra que o usuário usa para realizar a busca é muito importante. Este tipo de recuperação de informação usa as capacidades cognitivas humanas para evitar os problemas de interação homem-máquina. Os humanos são melhores para reconhecer a informação desejada do que para descrevê-la.
- **O caminho de busca do usuário não é estruturado.** A ordem na qual os itens

em um banco de dados são avaliados depende das escolhas do usuário.

- **A formulação de uma consulta ao sistema não requer que o usuário tenha conhecimento de uma linguagem de consulta ou da arquitetura do banco de dados. O navegador possui apenas botões que indicam ao usuário o caminho a seguir.**

Este método é adequado quando a informação necessária está mal definida, por ajudar o usuário a determinar a informação necessária, ou quando o usuário quer ter uma impressão do conteúdo de uma coleção de dados, onde o navegador oferece uma visão geral dos dados da coleção.

Um exemplo de aplicação pode ser observado quando um pesquisador navega em um banco de dados de um museu para observar as obras disponíveis para estudo. A princípio, o estudante não tem em mente uma obra específica ou mesmo um artista preferido. O objetivo do usuário, neste caso, é navegar pelo banco de dados e apreciar o acervo do museu.

4.2. CONSULTA POR IMAGEM EXEMPLO

Em uma recuperação típica de imagens por conteúdo, o usuário tem uma imagem em que tem interesse e quer encontrar imagens similares do banco de dados. O cenário aqui envolve uma consulta com uma imagem de exemplo. O usuário entra com uma imagem ou um pedaço de uma imagem e quer recuperar imagens do banco de dados que tenham algum pedaço semelhante à imagem de entrada. Consulta por imagem exemplo é o método de recuperação no qual uma consulta por imagens consiste de uma ou mais imagens.

A maioria das técnicas de recuperação de imagens usa características de imagens apesar de não ser tão eficiente no caso de consulta direta [2].

Um dos maiores problemas da recuperação de imagens por conteúdo é que sistemas tradicionais de busca comparam apenas tipos de dados iguais, o que não seria útil no caso de busca por imagens. A abordagem consulta por imagem exemplo lida

com operadores de similaridade. A tabela abaixo compara recuperação de dados, que é equivalente a consulta direta, e recuperação de informação, que corresponde a consulta por imagem exemplo.

	Recuperação de Dados	Recuperação de Informação
A representação da informação armazenada	Tipos de objetos e fatos bem definidos	Informação não-estruturada
O método de resposta à consulta	Direto, através dos fatos	Informação que conterà o que o usuário quer
A relação entre a consulta formulada e a satisfação do usuário	Satisfação ou não satisfação (determinística)	Uma alta probabilidade de que o usuário fique satisfeito
A definição de um sistema de sucesso	A saída do sistema são os fatos requeridos?	O sistema satisfaz o usuário?

TAB1.: COMPARAÇÃO ENTRE RECUPERAÇÃO DE DADOS E RECUPERAÇÃO DE INFORMAÇÃO.

A recuperação de informação é baseada em probabilidades. Quando um usuário consulta uma coleção de documentos, o sistema de recuperação de informação retorna um número para cada documento na coleção, indicando a sua relevância, que está relacionada à probabilidade do documento responder à consulta. A lista de documentos com suas relevâncias ajudará o usuário a obter sua resposta.

Embora as características para computar o número de relevância sejam bem diferentes, a consulta por imagem exemplo usa os mesmos princípios da recuperação de informação em texto para satisfazer ao usuário. A proposta dos sistemas, que era dar respostas determinísticas, agora é minimizar a busca do usuário e avaliar esforços.

Esta abordagem é muito usada em sistemas que usam as características cor global e textura [36]. Por exemplo, um profissional da área de propaganda pode ter interesse em encontrar uma imagem de uma paisagem específica, como montanhas cobertas por neve ou praias em dia ensolarado. Como este tipo de imagem é muito comum em vários bancos de dados, é possível que o profissional tenha um exemplo disponível, mas esteja procurando uma outra com características semelhantes.

4.3. CONSULTA POR ESBOÇO

A abordagem de consultar o banco de dados usando um esboço assemelha-se à anterior, já que o usuário deve fornecer como entrada um esboço da informação de que precisa. O sistema pode oferecer uma ferramenta interativa que permita que o usuário faça um desenho sobre o assunto que procura, ou ainda oferece partes de imagens para que o usuário consiga compor um esboço a partir de fragmentos de imagens.

Esta abordagem interativa pode consistir de um espaço da tela onde o usuário pode escolher os ícones para representar os objetos de que precisa e das ferramentas necessárias para selecionar e posicionar os ícones para compor sua consulta. Esta tela é chamada de *sketch pad window* [41].

Da mesma maneira que na consulta por imagem exemplo, a resposta do sistema será uma lista com as imagens semelhantes ao esboço fornecido pelo usuário. Esta lista será ordenada de acordo com o grau de relevância da imagem para a consulta.

Esta abordagem é muito utilizada para recuperar imagens que contenham objetos com formas similares à do exemplo [36]. Um exemplo de aplicação é a busca de fotos de pessoas baseada em retratos falados. O usuário pode elaborar um esboço do rosto de uma pessoa desenhando o formato do mesmo ou ainda montar o mesmo esboço usando imagens de partes da fisionomia humana previamente criadas.

4.4. CONSULTA POR TEXTO

Uma maneira intuitiva de descrever uma imagem é usando palavras. Estas abstrações são chamadas descrições ou anotações. Consulta Direta em Descrições (*Direct Query on Descriptions*) é o método de recuperação no qual uma consulta por imagens consiste apenas de valores de características especificadas pelo usuário. Estes valores são associados às imagens através do conhecimento humano ou através de um

título dado à imagem no momento da sua inserção. Assim, a consulta se referirá às descrições e o resultado será um conjunto de imagens associado às descrições. E será necessário o uso de um método de indexação.

Já existem sistemas de recuperação de imagens automáticos [1] que utilizam a semântica da imagem e processamento de linguagem natural. A recuperação é feita totalmente com base nas descrições, as imagens não são comparadas com outras imagens. Estes métodos estão sendo avaliados e utilizados com parcimônia já que apresentam falhas e não são capazes de recuperar totalmente a semântica de uma imagem.

O trabalho de associar palavras-chaves a imagens pode ser superado, mas a ambigüidade das descrições continua a ser um problema. Técnicas de processamento de linguagem natural podem ser usadas para amenizar o problema. Mas, um outro problema pode continuar, quando uma imagem está associada a uma palavra-chave em um contexto e a outra num outro momento.

Para indexação de descrições usualmente uma equipe de humanos é contratada. Cada membro da equipe pode ter sua própria interpretação subjetiva da imagem. Entretanto, para recuperação de imagens não é importante o que o indexador pensa que a imagem representa, mas o que o usuário associa à imagem. A associação depende do objetivo do usuário.

Então, as duas maiores desvantagens de se permitir consultas feitas apenas com o uso de descrições são os esforços requeridos para a indexação e a ambigüidade destas descrições. A vantagem do método é que permite uma recuperação baseada na semântica das imagens. Esta abordagem será muito eficiente em um domínio onde cada objeto tem uma única descrição.

Um exemplo de aplicação seria um sistema para consultar figuras de animais. Neste caso, cada imagem teria uma única descrição e o indexador e o usuário teriam a mesma percepção sobre o objeto.

4.5. CONSULTA POR CARACTERÍSTICAS

Consulta Direta em Características da Imagem (*Direct Query on Image Features*) é o método de recuperação no qual uma consulta por imagens consiste de valores diretamente derivados da imagem. Uma característica é uma abstração de uma imagem a valores numéricos que um computador pode processar. É um atributo que não tem informação sobre a imagem [2].

Características podem variar de medidas simples, como o número de pixels vermelhos na imagem, a propriedades de objetos na imagem, como a forma de um objeto. Uma implementação do primeiro caso não oferece grandes problemas para os bancos de dados convencionais, que operam com facilidade com dados numéricos. O outro caso é mais complexo e requer tipos de dados específicos para armazenar dados geométricos no banco de dados. Mesmo assim, já existem sistemas que usam as formas dos objetos da imagem. O maior problema é que o usuário precisa definir limiares durante a procura. A principal vantagem do método é que possibilita a extração automática de características das imagens. Outra vantagem é que o uso de um modelo pré-definido garante uma abstração objetiva da imagem, já que o processo será privado da subjetividade inerente ao ser humano [1].

Assim, a vantagem do método Consulta Direta em Características da Imagem é que as abstrações podem ser derivadas de imagens automaticamente e objetivamente. A desvantagem é que a formulação de consultas é difícil para o usuário porque as condições têm que conter valores das características das imagens. Aplicações de sistemas de recuperação usando características das imagens são usualmente restritas a domínios específicos para reduzir a complexidade do modelo de extração de característica necessário [2].

Um exemplo de aplicação é o RUI (*Representations for Understanding Images*) [40], um sistema tutorial inteligente que oferece facilidades para estudantes de medicina estudarem grandes bancos de imagens radiológicas. Os conceitos visuais são representados pelas características forma, tamanho e localização de componentes

anatômicos.

4.6. REALIMENTAÇÃO DE RELEVÂNCIA

Em algumas abordagens, normalmente centradas no processamento automático, o usuário precisa decompor a informação de que precisa em valores de características e ainda associar a elas os pesos de acordo com a sua importância para a busca. O sistema tenta encontrar imagens similares à consulta do usuário. O desempenho destes sistemas não é satisfatório por causa da subjetividade da percepção humana e ainda da distância entre os conceitos de alto nível e as características de baixo nível.

As abordagens centradas no processamento automático supõem que o usuário terá facilidade para mapear os conceitos de alto nível para características de baixo nível. Em alguns casos, isso pode acontecer [8]. Por exemplo, no caso de mapear uma maçã em função das características forma e cor. Mas, no caso de um castelo, por exemplo, essa tarefa não será tão trivial.

Além disso, deve-se considerar a subjetividade da percepção humana [8]. Pessoas diferentes, ou ainda a mesma pessoa em momentos diferentes, podem ter uma percepção diferente sobre o mesmo conteúdo visual.

Abordagens mais recentes utilizam a técnica da realimentação de relevância, na qual há uma interação entre o usuário e o computador para refinar consultas de alto nível a representações baseadas em características de baixo nível. O usuário pode começar sua busca com uma consulta “grosseira” e continuamente refinar sua consulta, acrescentando detalhes sobre a informação de que precisa através de realimentação de relevância. Cada vez que o usuário interage com o sistema, ele informa quais imagens são realmente relevantes para a sua consulta, associando novos pesos a cada imagem retornada como resposta.

O objetivo da realimentação de relevância é encontrar os pesos adequados para modelar a informação de que o usuário precisa [8]. Os pesos embutidos no objeto

da consulta são dinamicamente atualizados, a cada interação com o usuário, para modelar os conceitos de alto nível e a subjetividade humana.

Uma arquitetura de realimentação de relevância foi implementada em uma das versões do sistema MARS (*Multimedia Analysis and Retrieval System*) [10]. Os testes foram feitos usando dois conjuntos de imagens. O primeiro contém artefatos antigos da África e o segundo, imagens da empresa *Corel Corporation* [8].

Foi apresentada uma outra técnica de realimentação de relevância [14] que tem uma estrutura teórica sólida, com fundamentação matemática, e se baseou no modelo que considera que imagens contêm múltiplos níveis de conteúdo. O objetivo fundamental desta técnica é estimar os parâmetros ideais de consulta.

As imagens são modeladas em dois níveis: a um alto nível, o usuário humano observa as cores, formas e texturas dos objetos da imagem; a um nível mais baixo, cada uma destas características pode ser representada por um vetor de características. Este modelo tem como origem a percepção visual humana, exige menos processamento e oferece melhor desempenho na recuperação.

Foi construído um sistema baseado no algoritmo de otimização descrito. Foram feitos testes usando dois conjuntos de imagens: um de imagens de mais de 500 categorias diferentes e outro com imagens de textura.

Os resultados experimentais indicaram que o novo sistema foi mais eficiente e mais rápido que os outros em todas as situações simuladas, em termos de chamada e precisão. Concluiu-se ainda que o novo sistema requer menos esforço computacional porque modela o conteúdo da imagem em dois níveis.

Uma das desvantagens do sistema é que não permite consultas em função de características semânticas da imagem. Uma expansão deste sistema pode prover ao usuário a possibilidade de associar à imagem também uma característica semântica.

4.7. CONCLUSÃO

A maioria dos sistemas atuais de recuperação de imagens por conteúdo

oferece como resposta às consultas uma lista de imagens, em ordem decrescente de probabilidade de relevância. Estes sistemas diferem em domínio de aplicação, escolha das características, técnicas para o processamento das características visuais e mecanismos de consulta. Por este motivo, tornou-se necessária uma avaliação destes sistemas para que possam ser construídos novos sistemas, cada vez mais satisfatórios às necessidades dos usuários.

5. AVALIAÇÃO

A avaliação foi reconhecida como um elemento significativo do processo através do qual sistemas de recuperação de informação são implementados no mundo real [15]. A direção dos estudos de uma área qualquer é determinada, e implicitamente validada, pela adoção de algum tipo de metodologia estruturada ou estrutura de avaliação. Isso serve para guiar os pesquisadores em decisões como quais critérios, medidas e métodos utilizar. Além disso, faz com que os pesquisadores indiquem com clareza e precisão as funções do sistema desenvolvido.

Müller et al [20] desenvolveu, recentemente, um trabalho sobre avaliação de sistemas de recuperação de imagens por conteúdo que destaca as principais medidas utilizadas pelos pesquisadores da área. Este capítulo baseou-se no estudo destes autores. A avaliação de sistemas desta natureza é muito difícil por causa do grande número de variáveis a serem consideradas e do alto grau de subjetividade envolvido no processo, já que as decisões a serem tomadas durante o desenvolvimento do sistema dependem em grande parte do contexto da aplicação.

As primeiras pesquisas (por exemplo, [33]) sobre o desempenho de sistemas de recuperação de imagens por conteúdo mostravam os resultados sobre poucos exemplos de consultas, o que não refletia com exatidão o desempenho do sistema. Os desenvolvedores selecionavam as consultas que ofereciam os melhores resultados e assim os adaptavam para que esta simulação oferecesse bons resultados a serem apresentados. Portanto, essa não é uma medida objetiva de desempenho e nem uma maneira adequada de se comparar sistemas diferentes.

Existem várias medidas de desempenho para recuperação por conteúdo, mas poucas foram usadas em grande escala. Além disso, muitas das medidas usadas em recuperação de imagens por conteúdo (tais como precisão, chamada e sua representação gráfica) foram usadas em recuperação de informação.

É claro que a avaliação do desempenho de sistemas é essencial para se

afirmar quais métodos são satisfatórios ou não. Mas, com tantas diferenças, a comparação destes sistemas torna-se muito difícil. A maioria das aplicações falha ao tentar desenvolver modelos automáticos orientados a pixels ou modelos totalmente assistidos por um usuário especialista [4].

A avaliação de recuperação de imagens é um tópico muito importante para o sucesso do desenvolvimento de sistemas e sua utilidade em aplicações práticas [19]. Por esse motivo, foram criados alguns critérios para se avaliar a resposta que o sistema oferece ao usuário. Entre eles, destacam-se:

- **O grau de satisfação do usuário.** A comparação de usuários é um método interativo. Os usuários julgam o sucesso de uma consulta diretamente depois da consulta. É difícil considerar um número grande de comparações de usuários pois consomem muito tempo. O usuário interage com o sistema e indica se a resposta do sistema condiz ou não com a sua expectativa [5]. Ainda é possível usar um método comparativo de respostas. Este é o método de teste mais fácil. São apresentados dois ou mais resultados a cada usuário, que deve escolher qual é o melhor resultado. Algumas pesquisas anteriores usam os termos sistemas totalmente automáticos e tentam encontrar uma característica única perfeita. Pesquisas mais recentes usam o termo sistemas interativos e humanos no processo. Alguns experimentos foram feitos nesta área [4] e concluíram que a maioria dos usuários usa dois tipos de julgamento: se a recuperação é “boa”, ou seja, se a resposta corresponde às suas expectativas; e se a recuperação é eficiente, no contexto específico da aplicação. Uma consideração importante é que um sistema cuja resposta tenha imagens relevantes com classificação inferior à de imagens não relevantes não apresenta bom desempenho. As imagens do topo da classificação devem ser mais relevantes que as de nível mais baixo. Em alguns estudos [6] foi necessário avaliar o desempenho do usuário, considerando-se o número de iterações necessárias para a convergência à imagem alvo.
- **O tempo de resposta.** Dependendo do campo de aplicação, o tempo que o

sistema leva para executar uma consulta deve ser um item de muita importância na avaliação. Portanto, é recomendável declarar o tempo de execução para cada consulta em conjunto com a velocidade da máquina. Assim, os sistemas podem ser comparados com base no desempenho da recuperação e também com base nas considerações sobre precisão e velocidade se estes métodos estiverem disponíveis. Para avaliar se o tempo de resposta de um sistema de recuperação de imagens por conteúdo é satisfatório, é necessário considerar o objetivo principal do sistema [5].

- **A relação entre as imagens relevantes e as não relevantes que pertencem ao conjunto de saída (rechamada e precisão).** As medidas de avaliação mais comuns usadas em recuperação de informação são *precisão* e *rechamada*. É possível conseguir um ótimo índice de rechamada recuperando todas as imagens. Da mesma forma, precisão pode ser alta se o resultado apresentar algumas poucas imagens. Assim, precisão e rechamada devem ser usadas em conjunto (por exemplo, precisão a 0.5 de rechamada), ou o número de imagens recuperadas ser especificado (por exemplo, rechamada depois de 1000 imagens ou precisão depois de 20 imagens recuperadas). Estas medidas muitas vezes são apresentadas como um gráfico *precisão versus rechamada* (gráfico PR – Fig.1). Pesquisadores estão familiarizados com este tipo de gráfico e conseguem extrair informação dele sem problemas de interpretação. Rechamada é a proporção de imagens relevantes no banco de dados que são recuperadas em resposta à consulta. Precisão é a proporção das imagens recuperadas que são relevantes à consulta. A maior desvantagem deste método é que sua utilização é limitada à fase de testes do sistema. O usuário só será capaz de determinar o número de imagens relevantes no banco de dados a cada consulta durante os testes, já que novas imagens poderão ser inseridas no banco de dados pelo usuário a cada execução do sistema.

$$\text{rechamada} = \frac{\text{n}^\circ \text{ de documentos relevantes recuperados}}{\text{n}^\circ \text{ total de documentos relevantes na coleção}}$$

$$\text{precisão} = \frac{\text{n}^\circ \text{ de documentos relevantes recuperados}}{\text{n}^\circ \text{ total de documentos recuperados}}$$

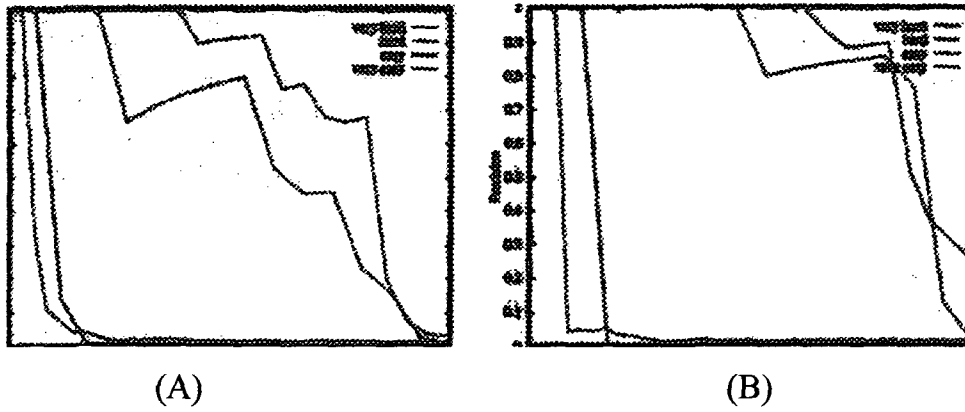


FIG.1: GRÁFICOS PR PARA QUATRO DIFERENTES CONSULTAS: (A) SEM REALIMENTAÇÃO DE RELEVÂNCIA E (B) COM REALIMENTAÇÃO DE RELEVÂNCIA [20].

Alguns dos sistemas estudados no capítulo 6 foram avaliados segundo o gráfico PR:

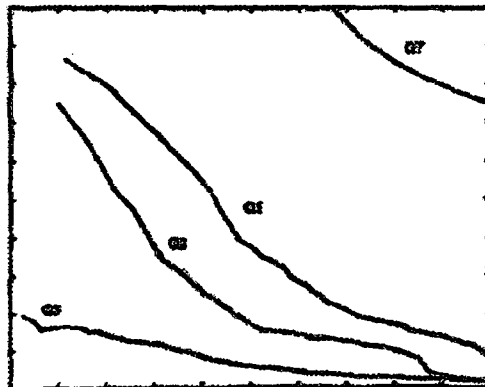


FIG.2: GRÁFICO PR DOS RESULTADOS DE 100 CONSULTAS AO BANCO DE DADOS DE IMAGENS SINTÉTICAS PRODUZIDOS PELO SISTEMA VISUALSEEK.

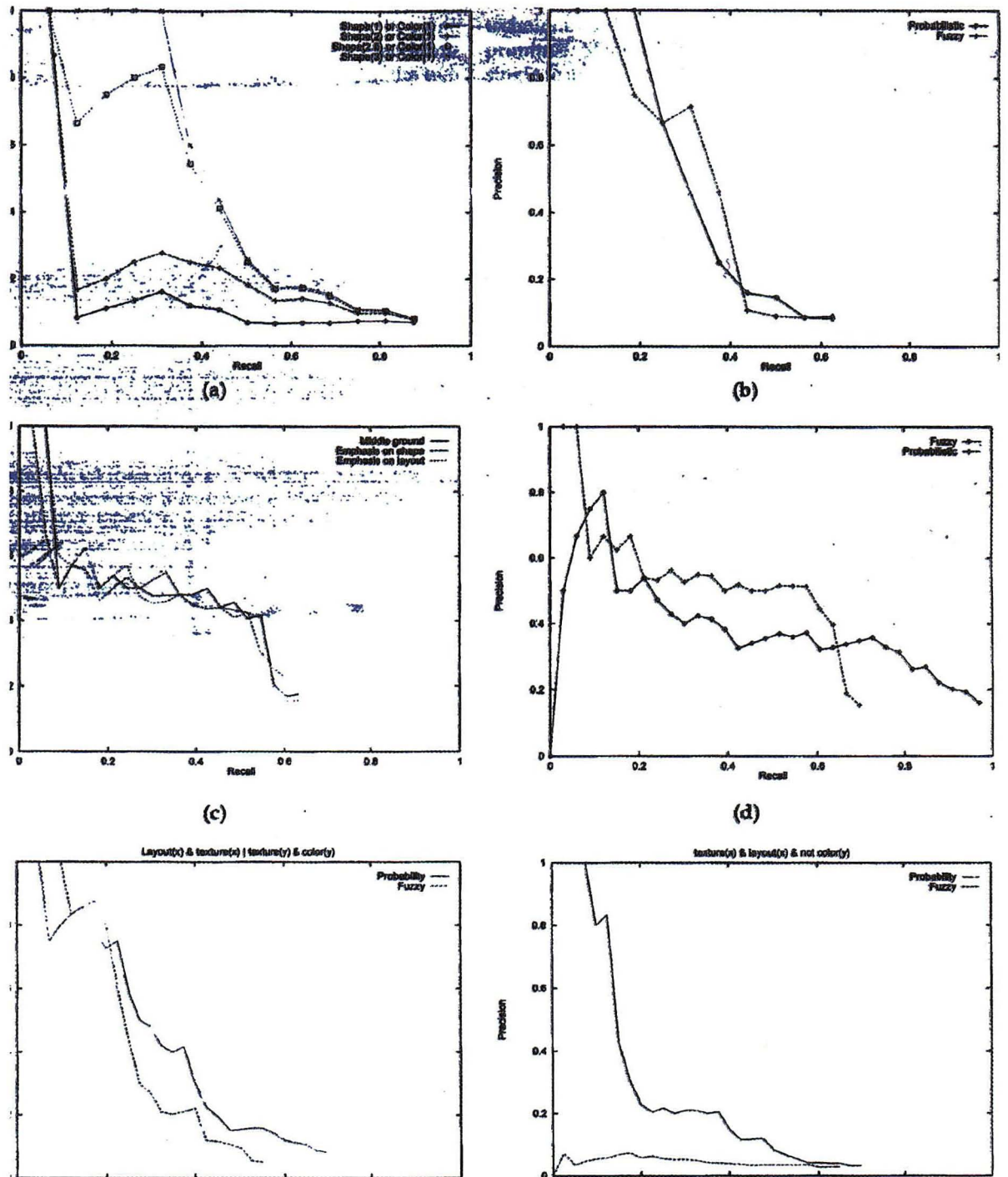


FIG.3: GRÁFICO PR DOS RESULTADOS DE TESTES FEITOS COM O SISTEMA MARS.

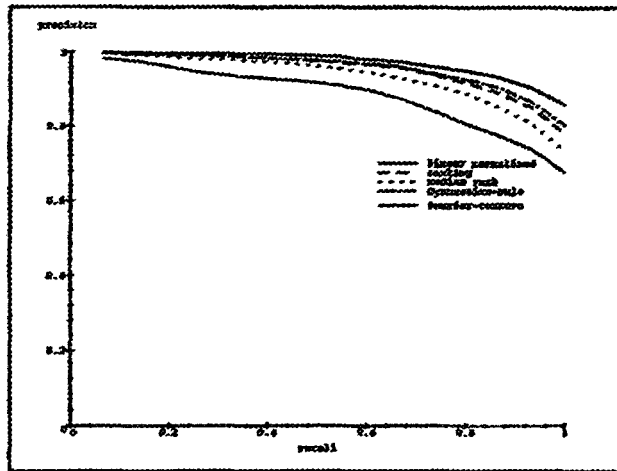


FIG.4: GRÁFICO PR DOS RESULTADOS DE CONSULTAS AO BANCO DE DADOS VISTEX, PRODUZIDOS PELO SISTEMA SURFIMAGE.

Como os gráficos podem não conter todas as informações desejadas, muitas outras medidas são usadas com frequência. Estas medidas são baseadas em precisão e rechamada:

- $P(10)$, $P(30)$, $P(N_R)$ – a precisão depois dos primeiros 10, 30, N_R documentos recuperados, onde N_R é o número de documentos relevantes para este tópico.
- *Mean Average Precision* – é a média simples da precisão.
- Rechamada a 0.5 precisão – rechamada na lista onde precisão está abaixo de 0.5.
- $R(1000)$ – rechamada depois de 1000 documentos recuperados (Fig.5).

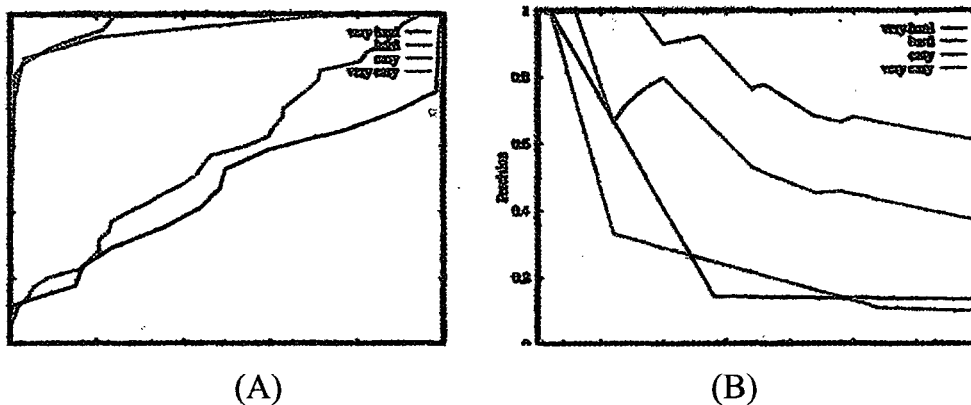


FIG.5: (A) GRÁFICO RECHAMADA VERSUS NÚMERO DE IMAGENS E (B) GRÁFICO PRECISÃO VERSUS NÚMERO DE IMAGENS [20]

- Primeira lista relevante – a lista dos documentos considerados mais relevantes.

Estes números chaves oferecem um conjunto de descritores de desempenho, permitindo que sistemas diferentes possam ser comparados de maneira objetiva.

A desvantagem destas métricas é a necessidade de se conhecer o grau de relevância das imagens armazenadas com relação à consulta antes do processamento da mesma [5].

As medidas precisão e chamada são ferramentas úteis em recuperação de informação. Mas, no contexto de bancos de dados de imagens elas não são suficientes por duas razões [19]. Primeiro, a seleção de um conjunto relevante em um banco de dados de imagens é muito mais problemática que em um banco de dados textual porque a definição do significado de uma imagem é um processo muito complexo.

A segunda razão é que sistemas de recuperação de imagens normalmente retornam uma lista de resultados ordenada por índice de relevância e o tamanho do conjunto resposta é reduzido a um número determinado de resultados relevantes. E a resposta esperada pelo usuário pode não pertencer a esse conjunto resposta.

Mesmo assim, precisão e chamada são medidas úteis em algumas circunstâncias, como por exemplo, quando o banco de dados de imagens lida com a semântica fornecida por uma descrição textual [19]. Neste caso, a busca usa a semântica em forma de texto e a relevância atribuída pelo usuário.

Também foram feitos experimentos psicofísicos que conduziram à construção do sistema PicHunter [6]. Foi estabelecida uma fundamentação para o banco de dados medindo-se o tempo requerido para encontrar uma imagem similar à imagem alvo quando as imagens são apresentadas numa ordem aleatória. O sistema construído envolve julgamento de similaridade de imagens por humanos. Além disso, guarda um histórico sobre as escolhas do usuário durante a busca, incorpora um modelo para interpretar a realimentação do usuário e tem um projeto flexível para permitir o desenvolvimento de diferentes versões. Algumas destas versões usam apenas características pictoriais, ou semânticas, ou uma combinação das duas.

PicHunter [6] é caracterizado como um sistema de busca ao alvo, que foi projetado para terminar sua busca quando uma imagem idêntica ao alvo é localizada. A maioria dos sistemas retorna um conjunto de imagens similares à imagem alvo. O

problema da última abordagem é que permite uma variedade de interpretações de diferentes usuários. Por outro lado, encontrar uma imagem exatamente igual à imagem alvo não tem muitas aplicações práticas.

5.1. PROBLEMAS BÁSICOS DA AVALIAÇÃO DE SISTEMAS DE RECUPERAÇÃO POR CONTEÚDO

Os avanços na área de avaliação de sistemas de recuperação de informação textual permitem que sistemas deste tipo sejam avaliados e comparados. Na área de recuperação de imagens por conteúdo, os avanços ainda não são satisfatórios e estão muito aquém dos anteriores. Existem vários grupos diferentes elaborando trabalhos nesta área, mas eles trabalham com muitos conjuntos diferentes de imagens especializadas. Não existe uma coleção comum de imagens. Além disso, não existe uma maneira comum para adquirir julgamentos de relevância e nem um esquema comum de avaliação.

5.1.1. Definição de um banco de dados de imagens padrão

Muitos problemas devem ser considerados para criar uma coleção de imagens comum. A coleção deve ser gratuita e sem restrições para cópias, de forma que as imagens possam ser copiadas da web e usadas em publicações. O maior problema é criar uma coleção com diversidade suficiente para servir a todos os domínios de aplicação em recuperação por conteúdo, como imagens médicas, de carros, reconhecimento de faces. Outro problema a ser ressaltado é o fato de que as imagens não podem estar degradadas por processos de compressão.

Uma maneira comum de construir uma coleção de imagens é usar os CDs de fotos da Corel, cada um contendo 100 imagens semelhantes. Infelizmente estas imagens não são gratuitas. Muitos grupos de pesquisa usam apenas um subconjunto desta coleção, e isso pode resultar em uma coleção consistindo de muitos grupos de

imagens muito diferentes. Isto pode dar uma falsa idéia sobre o grau de desempenho do sistema.

Outra coleção bastante usada é *VisTex*, que contém imagens de textura. Uma boa candidata para uma coleção padrão é a coleção de imagens e vídeos da MPEG-7. Infelizmente estas imagens não podem ser mostradas na web e o seu preço é muito alto.

Uma abordagem alternativa é para que os pesquisadores da área desenvolvam sua própria coleção. Existe um projeto em andamento na Universidade de Washington em Seattle. Esta coleção está disponível livremente sem restrições para cópia e oferece fotografias de diferentes regiões e sobre diferentes assuntos. É ainda pequena (mais ou menos 500 imagens), mas muitos grupos estão contribuindo para aumentar o conjunto. O tamanho da coleção deverá ser suficientemente alto para que permita que a precisão seja avaliada e não prejudique a velocidade dos testes.

5.1.2. Obtendo julgamentos de relevância

Em recuperação de imagens por conteúdo não existe ainda um meio comum de obter julgamentos de relevância para consultas. Da mesma forma a inclusão do usuário no processo de julgamento não está padronizada.

Uma técnica muito comum é usar bancos de dados de imagens padrão com conjuntos de tópicos diferentes tais como a coleção Corel. Julgamentos de relevância são dados pela coleção em si já que contém grupos diferentes de imagens. A escolha dos conjuntos pode influenciar os resultados, já que alguns conjuntos são visualmente bastante diferentes de outros ou muito semelhantes. O agrupamento nem sempre é baseado na similaridade visual global, mas nos objetos que a imagem contém. Em alguns estudos [27], imagens que são visualmente muito diferentes são excluídas da coleção, o que melhora os resultados.

Uma abordagem alternativa é agrupar as imagens de acordo com algum critério estabelecido por um especialista do domínio da aplicação. O conhecimento de

especialistas é muito usado em sistemas de recuperação de imagens médicas por conteúdo [28].

Alguns estudos [29] tentaram simular a opinião de um usuário, assumindo que o julgamento de similaridade do usuário é modelado por uma métrica usada em sistemas de recuperação por conteúdo. Mas, é muito difícil modelar usuários. Os estudos mostraram que os julgamentos humanos de similaridade não parecem obedecer aos requisitos de uma métrica e são certamente dependentes dos usuários e suas tarefas. Portanto, não há garantias de que as simulações sempre refletem um usuário humano.

Existem diferenças fundamentais entre estes métodos. A facilidade de obter julgamentos de relevância é uma vantagem de usar coleções com grupos pré-definidos de imagens similares. Julgamentos do usuário ainda podem ser feitos para uma coleção assim.

O conhecimento de um especialista do domínio deve ser usado quando está disponível, como na medicina e outros campos de aplicação. Para tarefas gerais de recuperação por conteúdo, acredita-se que o teste com usuários é essencial. Para uma avaliação completa, o usuário e suas expectativas são partes vitais do sistema.

O número de imagens que um usuário deve examinar pode ser reduzido usando-se métodos comuns como em recuperação de informação. Estes métodos não alteram significativamente os resultados de um sistema porque as primeiras n imagens relevantes de cada sistema estão no conjunto.

É essencial que o usuário examine uma fração um tanto grande do banco de dados, e que os julgamentos de relevância sejam feitos num curto período de tempo. Os usuários tendem a ficar satisfeitos facilmente [29], mesmo que existam poucas imagens relevantes à sua consulta. As características do grupo de usuários que fornecem os julgamentos de relevância também são importantes. Desenvolvedores de sistemas de recuperação por conteúdo têm noções diferentes de similaridade de imagens de usuários novatos.

5.2. OUTRAS MEDIDAS DE AVALIAÇÃO DE DESEMPENHO

Rank da melhor comparação. Foi feita uma pesquisa [30] em que foi observado se as imagens mais relevantes estavam nas primeiras 50 ou nas primeiras 500 imagens recuperadas. 50 representa o número de imagens retornadas na tela e 500 é uma estimativa do número máximo de imagens que um usuário espera encontrar.

Rank médio de imagens relevantes. Esta medida pode dar um bom indicativo do desempenho do sistema, embora contenha menos informação que um gráfico PR. É uma medida vulnerável, já que uma imagem com índice de relevância muito alto pode influenciar o resultado [20].

Teste do alvo. Esta abordagem é bastante diferente das outras medidas de desempenho. Os usuários conhecem a imagem alvo e a avaliação é feita com base no número de imagens que o usuário precisa examinar antes de encontrar a imagem procurada. Começando com imagens aleatórias, o usuário marca cada uma das imagens como relevante ou não relevante. Este tipo de avaliação foi usada nos experimentos que envolveram o sistema PicHunter [6].

Razão de erro. Esta medida é comum em reconhecimento de faces ou objetos. É de fato um valor de precisão, sendo importante conhecer onde o valor é medido [20].

$$\text{Razão de erro} = \frac{\text{N}^\circ \text{ de imagens não relevantes recuperadas}}{\text{N}^\circ \text{ total de imagens recuperadas}}$$

Eficiência da recuperação. Uma definição de eficiência da recuperação é dada pela equação abaixo. Se o número de imagens recuperadas é menor ou igual ao número de imagens relevantes, este valor é a precisão, caso contrário é a chamada de uma consulta [20].

$$\text{Eficiência da recuperação} = \begin{cases} \frac{\text{N}^\circ \text{ de imagens relevantes recuperadas}}{\text{N}^\circ \text{ total de imagens recuperadas}} & \text{se } \text{N}^\circ \text{ recuperadas} \leq \text{N}^\circ \text{ relevantes} \\ \frac{\text{N}^\circ \text{ de imagens relevantes recuperadas}}{\text{N}^\circ \text{ total de imagens relevantes}} & \text{caso contrário} \end{cases}$$

Detecção de imagens corretas e incorretas. Estas medidas foram usadas em um contexto de reconhecimento de objetos. Os números de classificações corretas e incorretas são contados. Quando divididos pelo número de imagens recuperadas, estas medidas são equivalentes à razão de erro e precisão [20].

Centrada no usuário. O sistema Epic [15] foi avaliado através de uma metodologia centrada no usuário. Alguns usuários testaram as duas versões do Epic (uma que permite apenas consultas visuais e outra que permite apenas consultas textuais) e responderam a questionários, ou seja, a avaliação foi baseada na satisfação do usuário. Foi constatada a preferência dos usuários pela versão que permite consultas visuais. A desvantagem é que existe o risco de que novas sessões de avaliação venham a produzir outros resultados. Além disso, o resultado também poderia ser diferente se as imagens pertencessem a outra área de aplicação e os usuários fossem especialistas da área. O ideal seria o desenvolvimento de novas medidas de desempenho para avaliar sistemas de recuperação de imagens de forma mais objetiva.

5.3. REPRESENTAÇÕES GRÁFICAS

Gráficos precisão versus chamada (Gráficos PR). Gráficos PR são um método de avaliação padrão em recuperação de informação e estão sendo usados cada vez mais na comunidade de recuperação de imagens por conteúdo. O gráfico PR contém muita informação, e é bastante conhecido pelos pesquisadores. É também comum apresentar um gráfico parcial, o que pode ser útil para mostrar uma região em mais detalhes, mas pode também ser enganoso já que áreas de desempenho ruim podem ser omitidas.

Os gráficos PR podem fazer uma boa distinção entre vários resultados encontrados. O problema é que eles dependem do número de imagens relevantes para a consulta. Informações práticas, tais como precisão e chamada depois de um certo número de imagens terem sido recuperadas, não podem ser obtidas [20].

Gráficos de precisão versus n° de imagens recuperadas e rechamada versus n° de imagens recuperadas. Tomados separadamente, estes gráficos contêm apenas alguma informação de um gráfico PR. Quando combinados, eles contêm mais informação e podem facilmente ser interpretados. O gráfico de rechamada parece mais positivo que o gráfico PR, especialmente quando poucas imagens relevantes são recuperadas. O gráfico de precisão é semelhante ao gráfico PR, mas dá uma melhor indicação do que deve ser um bom número de imagens a recuperar. É mais sensível ao número de imagens relevantes para uma determinada consulta.

Gráficos das imagens corretamente recuperadas versus todas as imagens recuperadas contêm a mesma informação que os gráficos de rechamada, mas em escala diferente. Os gráficos da fração correta versus o número de imagens recuperadas são equivalentes aos gráficos de precisão. Os gráficos da razão média de reconhecimento versus o número de imagens recuperadas mostram o percentual médio de imagens relevantes dentre as primeiras recuperadas. Isto é equivalente ao gráfico de rechamada.

5.4. PROPOSTAS

A princípio, muitas das técnicas apresentadas são equivalentes ou contêm a mesma informação. Seria benéfico para a comunidade de recuperação de imagens por conteúdo se somente nomes e definições padronizados fossem usados para medir desempenho.

Uma proposta é usar bancos de dados de imagens disponíveis gratuitamente ou fazer com que bancos de dados já avaliados em pesquisas estejam disponíveis. Isto tornaria possível comparar os resultados obtidos por vários sistemas.

Também foram propostas medidas de desempenho similares às aquelas usadas em TREC (*Text REtrieval Conference*) porque estas medidas podem ser interpretadas facilmente e elas contêm informação complementar. Este conjunto contém uma mistura de medidas baseadas em *rank*, medidas com um único valor e medidas

gráficas:

- $Rank_1$ e $Rank_2$: $rank$ no qual a primeira imagem relevante é recuperada e o $rank$ médio de imagens relevantes normalizado.
- $P(20)$, $P(50)$ e $P(N_R)$: precisão depois de 20, 50 e N_R imagens são recuperadas.
- $R_p(0.5)$ e $R(100)$: chamada a uma precisão de 0.5 e depois de 100 imagens serem recuperadas.
- Gráfico PR.

Um simples $rank$ médio é difícil de ser interpretado, já que depende do tamanho da coleção (N) e do número de imagens relevantes para uma determinada consulta (N_R). Consequentemente, foi proposto um $rank$ médio normalizado, o $Rank_2$:

$$Rank_2 = \frac{1}{NN_R} \left(\sum_{i=1}^{N_R} R_i - \frac{N_R (N_R - 1)}{2} \right)$$

onde R_i é o $rank$ no qual a i th imagem relevante é recuperada. Esta medida é 0 para desempenho perfeito e se aproxima de 1 nos piores desempenhos. Para uma recuperação aleatória o resultado seria 0.5.

Existe a necessidade de medidas de avaliação padronizadas, já que muitas medidas são variações da mesma definição. Com isso, a comparação objetiva do desempenho de sistemas torna-se difícil. Para superar este problema, um conjunto padrão de medidas de desempenho e um conjunto de imagens também padronizado são necessários. Um banco de dados de imagens que pudesse ser frequentemente atualizado seria de grande ajuda às aspirações da comunidade de pesquisa. Além disso, seria útil que outras comparações entre sistemas fossem feitas regularmente.

Outros trabalhos precisam ser feitos para integrar os usuários no processo de avaliação. O objetivo final de um projeto é medir a utilidade de um sistema para um usuário. Portanto, avaliações de desempenho interativas precisam ser desenvolvidas.

6. SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO

Observa-se uma grande quantidade de opções de busca hoje, mas ainda é necessário que sejam feitos estudos mais atuais, envolvendo usuários em aplicações práticas, para explorar os compromissos entre as diferentes opções mencionadas. Aqui, foram selecionados alguns sistemas representativos e destacadas as suas características.

6.1. QBIC

- *Query by Image Content* [33]
- O sistema QBIC foi desenvolvido para explorar métodos de recuperação por conteúdo.
- Foi o primeiro sistema comercial de recuperação de imagens por conteúdo.
- Os primeiros artigos a seu respeito foram publicados por volta de 1993.
- Duas propriedades interessantes do QBIC são a capacidade de lidar com o conteúdo de imagens e vídeos – usando características como cor, textura, forma, e movimentos de imagens, vídeos e seus objetos – e a linguagem gráfica de consulta.
- Para representar cores, foi usado um histograma de cores. Para textura é usada uma representação matemática de características como aspereza, contraste e direcionalidade. A forma é representada por características como área, circularidade, direção quanto ao eixo maior e um conjunto de ângulos tangentes ao perímetro do objeto.
- O QBIC permite consultas baseadas em imagens exemplos, esboços ou desenhos feitos pelo usuário e padrões de cores, texturas selecionados pelo usuário, movimentos de câmeras e objetos e outras informações gráficas.
- A indexação usa características de alta dimensão, usa redução de dimensão, através de KLT, e R*-tree é usada como estrutura de índice.
- Durante a inserção de uma imagem no banco de dados, esta imagem é reduzida a

um ícone de tamanho padrão e associada a uma informação textual. Além disso, os objetos podem ser identificados, o que pode ser feito automaticamente ou através da interação com o usuário. Internamente, cada objeto é representado como uma matriz binária.

- O banco de dados usado para testes consiste de 7.450 imagens da *Mediasource Series* de imagens e áudio da *Applied Optical Media Corp.*, 4.100 imagens do CD *PhotoDisc*, 950 imagens da coleção *Corel Professional Photo CD*, e 450 imagens da coleção IBM.
- A avaliação do sistema foi feita de forma subjetiva, em função de testes com alguns exemplos de consultas e os resultados foram considerados satisfatórios pelos pesquisadores.

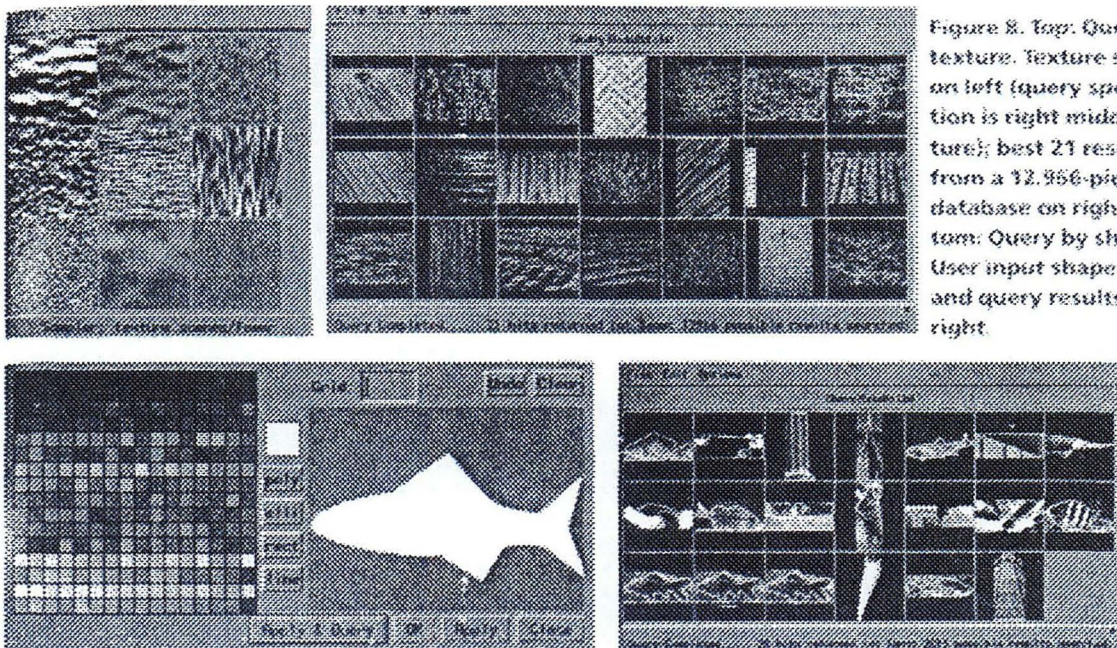


Figure 8. Top: Query texture. Texture is on left (query position is right middle); best 21 results from a 12,956-pic database on right. Bottom: Query by shape. User input shape and query results right.

FIG 6.: CONSULTA POR FORMA NO SISTEMA QBIC. O USUÁRIO ESCOLHE A FORMA NA TELA À ESQUERDA E OS RESULTADOS APARECEM À DIREITA.

6.2. VIRAGE

- Os primeiros artigos a seu respeito foram publicados por volta de 1997.
- Duas informações diferentes estão associadas ao objeto visual: informação sobre o objeto (metadado) e informação contida dentro da imagem (características).

- A extração de informação é feita por um ser humano especialista no conteúdo das imagens.
- As características extraídas automaticamente são cor e forma.
- Uma grande vantagem do sistema Virage [4] é a integração de características visuais e semânticas.
- Para indexar a característica cor foi usado um histograma de cores. Para a característica forma, foi usado um vetor com propriedades como cor, área, tamanho e centralidade.
- A principal abordagem para consulta é a consulta por exemplo, que pode envolver uma imagem exemplo ou valores de características visuais como exemplos.
- Suporta consultas visuais baseadas em cores, composição (layout de cores), textura e estrutura (informação sobre as bordas dos objetos).
- Também suporta combinações arbitrárias dos quatro tipos de consulta citados.
- Foi usado para testes um banco de dados de imagens aéreas, nas quais os objetos de interesse são edifícios, veículos terrestres, aeronaves, estradas.
- Outra contribuição deste trabalho é a observação de dois tipos de julgamentos para comparar sistemas de recuperação de imagens: a resposta do usuário, indicando se a recuperação foi satisfatória, e a eficiência do sistema, em função do domínio da aplicação.
- O usuário pode refinar a consulta através de realimentação de relevância.
- Um problema não resolvido é permitir o acesso remoto às informações visuais. Uma extensão para o sistema poderia envolver a permissão ao acesso remoto ao banco de dados em forma de consultas e ainda permitir que o usuário armazene resultados localmente.

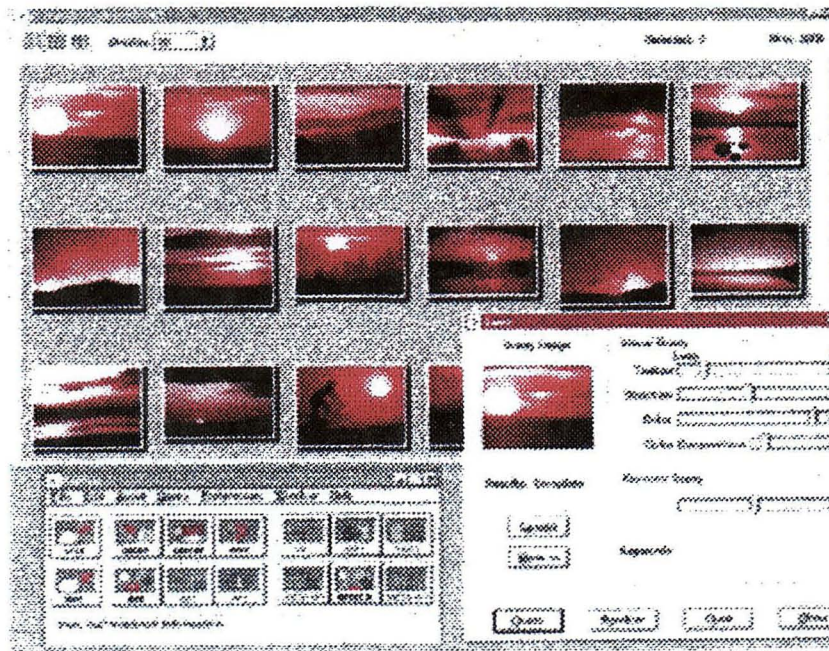


FIG 7.: EXEMPLO DE INTERFACE DO SISTEMA VIRAGE. O RESULTADO DA BUSCA POR PAISAGENS DO PÔR-DO-SOL, ENFATIZANDO SIMILARIDADE POR COR.

6.3. PHOTOBOK

- Os primeiros artigos a seu respeito foram publicados por volta de 1993.
- É um conjunto de ferramentas interativas para navegar e buscar imagens [11].
- Consiste de três módulos, um de forma, um de textura e um de faces, que constituem as características que o usuário pode buscar.
- Usa diretamente o conteúdo da imagem e anotações textuais em um banco de dados de imagens.
- Usa o FEM (*Finite Element Method*) para representar forma e derivação da KLT para faces.
- As representações utilizadas permitem a reconstrução das imagens a partir de seus coeficientes.
- Foram usadas estruturas do tipo árvore para armazenar os índices que permitem o processamento das consultas.
- Para formular uma consulta o usuário precisa escolher uma imagem do banco de dados ou uma categoria (através de uma imagem representante) e o sistema

retornará imagens semelhantes a ela no banco de dados.

- O sistema avalia vários tipos de descrições de imagens, como específicas da aparência (aplicadas a bancos de dados de faces humanas e *keyframes* de vídeos), descrições de textura (aplicadas a tipos de textura e também *keyframes*) e descrições de forma (aplicadas a bancos de dados de ferramentas manuais e peixes). E ainda usa uma combinação dessas descrições, usando dados 3-D de ventrículos do cérebro humano.
- Os testes foram realizados em bancos de dados de pessoas, quadros de vídeo, ferramentas manuais, peixes, imagens médicas 3-D.
- Sua versão mais recente se tornou mais interativa com o usuário, através de realimentação de relevância.
- O módulo de faces tem como principais aplicações sistemas de segurança e investigação criminal. Aplicações do módulo de forma incluem a busca em catálogos de produtos, como ferramentas manuais. Uma aplicação óbvia para o módulo de textura é auxiliar no projeto e decoração de indústrias.
- A avaliação do sistema foi feita em termos de precisão e chamada, para cada um dos módulos. Os resultados indicam que as expectativas foram satisfeitas.

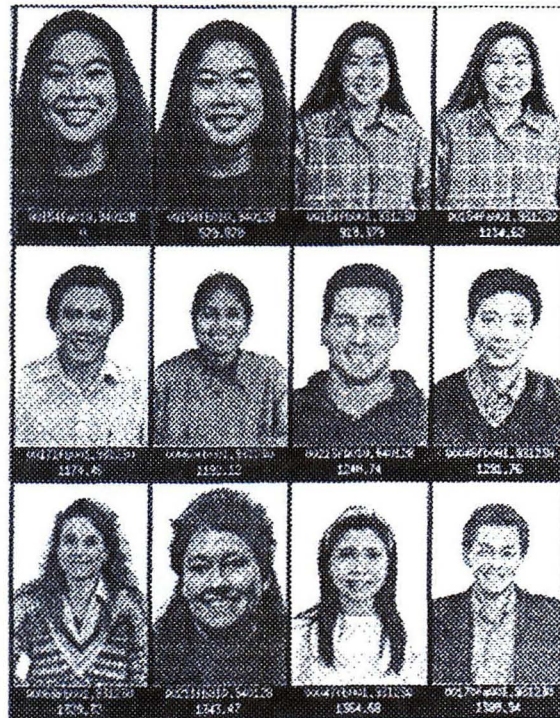


FIG 8.: EXEMPLO DE INTERFACE DO SISTEMA PHOTOBOOK.

6.4. VISUALSEEK

- O VisualSEEk [31] é um sistema de busca de características visuais, desenvolvido pela Universidade de Columbia.
- Os primeiros artigos a seu respeito foram publicados por volta de 1995.
- O aplicativo foi desenvolvido na linguagem Java por ser uma plataforma independente e oferecer facilidades para o acesso a *World Wide Web*.
- Suporta consultas baseadas em características visuais e seus relacionamentos espaciais.
- As características visuais são Conjuntos de Cores, a Transformada de Wavelet baseada em textura e localização espacial.
- Cada imagem é decomposta em regiões que têm propriedades de características, como cor, e propriedades espaciais, como tamanho, localização e relacionamentos com outras regiões. Assim, as imagens são comparadas através da comparação de suas regiões.

- O VisualSEEk é um sistema híbrido que integra indexação de imagens baseada em características e métodos de consulta espacial.
- Para acelerar o processo, a indexação usa uma estrutura de árvore binária para a característica cor. Para indexar a característica localização espacial, usa R-tree.
- O WebSEEk leva em conta o cálculo dos vetores de características dinâmicos baseado na realimentação do usuário. O usuário escolhe as regiões e os parâmetros de interesse para a sua consulta.
- A coleção de imagens usada na fase de teste do sistema tem 12.000 imagens. Mas, a aplicação é capaz de fazer busca em uma coleção de mais de um milhão de imagens da *World Wide Web*.
- Uma avaliação prévia foi feita, considerando alguns exemplos de consultas e uma amostra do banco de dados de teste. A avaliação consistiu de testes exaustivos, associando a cada imagem um valor relativo à sua relevância a cada consulta e comparando o resultado produzido pelo sistema com o esperado. O resultado da avaliação superou as expectativas e indicou novas direções a seguir, como por exemplo permitir uma maior integração entre as características cor, forma e textura.

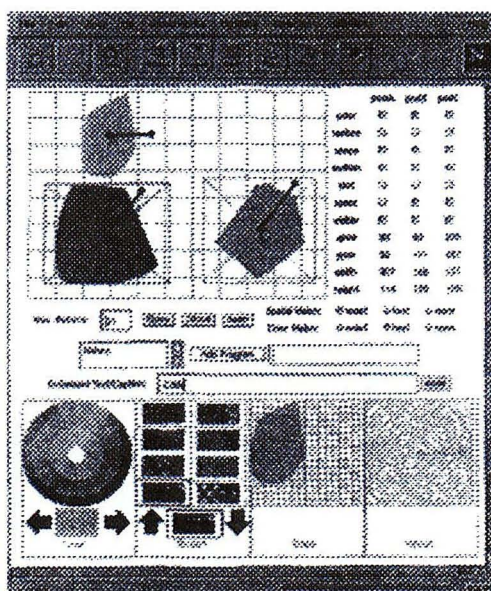


FIG 9.: EXEMPLO DE INTERFACE DO SISTEMA VISUALSEEK.

6.5. NETRA

- O sistema NeTra [34] foi implementado na *World Wide Web* usando a linguagem Java, por ser uma plataforma independente.
- Os primeiros artigos a seu respeito foram publicados por volta de 1997.
- Um aspecto que o distingue de outros sistemas é o uso de algoritmos de segmentação automática.
- Usa cor, textura, forma e localização espacial nas regiões da imagem segmentada para encontrar regiões similares no Banco de Dados.
- A característica cor foi representada através do histograma de cores, construído com um subconjunto de cores, dependente da aplicação. Isto porque a cor é representada para cada região, que normalmente contém um número pequeno de cores diferentes.
- A forma de um objeto ou região é descrita em função de seu contorno. Para representar o contorno de um objeto são usados alguns pontos da borda do objeto.
- A representação da textura é baseada na decomposição de Gabor e armazenada em um vetor de características.
- O índice para características textura e forma é representado usando uma estrutura de dados similar à SS-tree.
- Para indexar cor foi construída uma tabela binária $M \times N$, onde M é o número de cores na tabela e N é o número de regiões de imagens no banco de dados. Cada campo da tabela deve receber o valor 1 se a região possui aquela cor e 0, caso contrário. A imagem da consulta é representada como uma coluna da tabela e é feita uma interseção com a tabela para encontrar as imagens mais semelhantes.
- Cor, textura e forma de cada uma das regiões do banco de dados são indexadas separadamente. Para uma consulta consistindo de mais de uma destas características, a interseção dos resultados da busca usando uma característica individual pode ser calculada e ordenada com base em uma medida de similaridade.

- A informação sobre a localização espacial de um objeto ou característica dentro da imagem é representada usando a estrutura quadtree.
- Para reduzir o espaço de busca, o usuário também pode especificar outra característica relativa à posição espacial da ocorrência da característica dentro da imagem.
- Foram usadas imagens aéreas e de satélites para testes.
- Atualmente, o banco de dados contém 2.500 imagens coloridas da galeria de fotos da Corel. A coleção é organizada em 25 categorias diferentes com 100 imagens por categoria, incluindo jardins e montanhas.
- Cada imagem foi particionada em uma média de 10 regiões e o número total de regiões do banco de dados é de 26.000.
- O sistema ainda não usa realimentação do usuário, mas existe um interesse em explorar esta técnica futuramente.
- A avaliação foi feita com base em exemplos de consultas e nos resultados obtidos. Uma das conclusões obtidas é que acrescentando ao sistema uma capacidade de analisar e representar regiões de imagens de maneira individual, o desempenho do sistema aumenta de forma drástica.

6.6. MARS

- *Multimedia Analysis and Retrieval System* [10] [37].
- Os primeiros artigos a seu respeito foram publicados por volta de 1997.
- É uma integração entre DBMS (*DataBase Management System*) e IR (*Information Retrieval*), integração entre indexação e recuperação e integração entre computadores e seres humanos.
- O foco principal não é encontrar uma única representação que seja a melhor, mas como organizar várias características visuais em uma arquitetura de recuperação significativa que possa se adaptar dinamicamente a diferentes aplicações e diferentes usuários.

- As imagens são representadas por um vetor de características de baixo nível, como cor, textura, forma e layout de cor, e ainda uma anotação manual sobre o conteúdo das imagens.
- A característica cor é representada por um histograma de cores, no espaço de cores HSV (*hue, saturation, value*).
- A abordagem *wavelet* foi escolhida para representar a característica textura, usando um vetor de 10 dimensões. Representações usando matriz de co-ocorrência e Vetor de Características foram também testadas.
- A característica forma de um objeto é representada pelas bordas do objeto. A técnica escolhida para armazenar a forma foi o descritor de Fourier.
- A característica layout de cor é representada dividindo-se a imagem em blocos e extraindo o histograma de cor para cada bloco.
- Além das características visuais apresentadas, uma anotação textual sobre cada imagem é armazenada em um vetor de caracteres.
- O usuário escolhe algumas imagens da coleção e especifica as características de interesse.
- O sistema recupera imagens similares com base em cada característica escolhida na consulta e combina os resultados para produzir uma lista de imagens relevantes à consulta.
- O índice é montado utilizando-se uma estrutura de dados do tipo árvore. A estrutura é híbrida e incorpora as vantagens da k-d tree e da estrutura baseada nos limites das regiões.
- Propõe uma arquitetura de realimentação de relevância em Recuperação de Imagens e integra tal técnica a vários níveis durante a recuperação, incluindo refinamento de consultas, seleção automática de ferramenta de comparação e adaptação automática de características.
- Uma das propostas do experimento é comparar o desempenho da recuperação com e sem o uso de realimentação de relevância. Os experimentos testaram o tempo que

o algoritmo de recuperação leva para convergir para a resposta correta e se os resultados da recuperação foram satisfatórios, de maneira subjetiva. A conclusão foi que quanto maior o número de iterações de realimentação de relevância melhor será o desempenho da recuperação. Mas, os experimentos também mostraram que a maior diferença no desempenho acontece depois da primeira iteração.

- Dois bancos de dados foram usados para testar o sistema. O primeiro é uma coleção de imagens de artefatos antigos do *Fowler Museum of Cultural History*, que consiste de 286 imagens. O segundo conjunto de dados é a coleção Corel de imagens, com cerca de 70.000 imagens da natureza.
- O sistema foi avaliado em termos de precisão e chamada. Os resultados foram satisfatórios e comprovaram a robustez da abordagem.

6.7. SURFIMAGE

- Os primeiros artigos a seu respeito foram publicados por volta de 1997.
- O sistema tem por objetivo ser genérico e flexível [12].
- As categorias de classes disponibilizadas pelo Surfimage envolvem características de baixo nível, como cor, forma e textura, e características de alto nível, que consistem de modelos complexos do conteúdo da imagem e, algumas vezes, uma análise estatística do banco de dados.
- A característica cor é representada através de um histograma. A textura, através da transformada de *wavelet*. A forma, pela transformada de Fourier.
- Usa a abordagem consulta por imagem exemplo e integra características avançadas como a combinação de classes das imagens, consultas múltiplas e refinamento de consulta.
- Durante a consulta o usuário seleciona uma classe de imagens que deseja, deve escolher uma métrica de similaridade e ainda escolher uma imagem do banco de dados.
- Em termos estatísticos, cada imagem tem uma distribuição paramétrica que é usada

durante a busca para limitar o espaço de busca.

- Usa realimentação de relevância por parte do usuário, a cada iteração.
- Os bancos de dados Vistex, com 384 imagens de texturas homogêneas, Columbia, com 1440 imagens de 20 objetos 3D diferentes, o ORL, de faces, o MIT, com 7.652 imagens de faces, o BTphoto, de cidades e países, e outros foram usados para testes.
- Foi avaliado em função das medidas precisão e rechamada, através da construção do gráfico PR (precisão versus rechamada). Os resultados foram altamente satisfatórios, mas deve-se levar em consideração a escolha das consultas usadas no processo de avaliação.

6.8. SAMMATCH

- Os primeiros artigos a seu respeito foram publicados por volta de 1999.
- A técnica [13] é inspirada pela digitalização de som na qual as magnitudes de uma onda de som são testadas em algum intervalo de tempo fixo e armazenadas como números.
- Para cada região de teste, calcula-se a cor média dos pixels na região. Estas médias formam o vetor característica da imagem.
- Esta abordagem é boa para aplicações gerais porque não lida com a semântica das imagens. Outras vantagens são que a representação ocupa pouco espaço e a facilidade do processamento por lidar com números.
- A característica cor é representada usando-se um histograma de cores, que descreve a distribuição de cores em cada região da imagem.
- Para reduzir o alto custo da comparação entre histogramas, uma chave numérica é calculada para cada histograma local, facilitando a busca.
- A indexação é feita através de um filtro. O sistema calcula o vetor de características da imagem consulta e associa um peso para cada elemento do vetor baseado no seu valor.

- O sistema determina o filtro do vetor de característica da imagem consulta levando em consideração os pesos espaciais. O sistema descarta as imagens irrelevantes à consulta e classifica outras imagens com relação ao grau de relevância.
- O usuário deve escolher uma imagem exemplo, ou desenhar um esboço dela, e editar o cenário padrão, incluindo os pesos espaciais.
- O sistema calcula o vetor característica da imagem consulta, e associa um peso a cada elemento do vetor e usa um filtro para descartar imagens irrelevantes. Por fim é calculada a similaridade no conjunto candidato para atribuir um valor às imagens que sobraram e o usuário faz a seleção final.
- A principal desvantagem do sistema é que não lida com a semântica das imagens.
- Não usa realimentação de relevância, mas existe um projeto para uma nova versão com esta técnica implementada.
- Os testes foram feitos usando um banco de dados com 15.808 imagens, consistindo de uma variedade de categorias (rosas, pássaros, balões, prédios, céus, estátuas, comidas...).
- A avaliação do sistema foi feita com base na média da rechamada entre várias consultas e os resultados foram satisfatórios segundo os pesquisadores. Outra avaliação que apresentou resultados satisfatórios foi com relação ao tempo requerido por uma consulta e o espaço necessário para o armazenamento dos dados.

6.9. EPIC

- Os primeiros artigos a seu respeito foram publicados por volta de 1997.
- O sistema Epic [15] foi criado usando-se FLAIR (*FLexible Architecture for Information Retrieval*), um desenvolvimento de ECLAIR (*Extensible Class Library for Information Retrieval*), um banco de dados orientado a objetos e C++. A interface foi desenvolvida em Java.
- Cada registro do banco de dados contém: a imagem fotográfica digitalizada; dados

descritivos como o nome do fotógrafo ou a data da foto, armazenadas manualmente no momento do armazenamento da imagem; palavras-chave descrevendo o conteúdo da imagem, associadas à imagem no momento da criação da coleção de teste; e características espaciais, consistindo de um par ordenado com um rótulo representando um objeto em destaque na imagem e as suas coordenadas dentro da imagem.

- As características foram derivadas semi-automaticamente, ou seja, o sistema detectou parte da informação e o usuário interagiu com o sistema para produzir resultados mais satisfatórios.
- As características espaciais foram representadas por pares ordenados, consistindo de um rótulo representando um objeto da imagem e as coordenadas da posição do objeto dentro da imagem.
- Foram implementadas duas versões: uma que permite apenas consultas visuais e outra que permite apenas consultas textuais. Mas, a interface do sistema ainda oferece facilidades para a implementação de consultas baseadas em esboços do usuário e baseadas em valores de características.
- O sistema usa realimentação de relevância por parte do usuário.
- Foi criado um banco de dados para teste com 800 fotografias tiradas da *National Trust for Scotland*. Os principais temas das fotografias desta coleção são: exteriores de castelos e outras construções; algumas partes dessas construções, como jardins, lagos, ornamentos, arranha-céus e paisagens.
- Alguns usuários testaram as duas versões do Epic e responderam a questionários, ou seja, a avaliação foi centrada na satisfação do usuário. Foi constatada a preferência dos usuários pela versão que permite consultas visuais. O resultado mostrou a aceitabilidade do sistema por vários usuários, de níveis de conhecimento diferentes.
- Também foi usada a medida rechamada para avaliar o sistema. O teste foi feito por alguns usuários que responderam aos questionários. O resultado confirmou o

desempenho do sistema quando testado por usuários.

6.10. MOSAIC

- Os primeiros artigos a seu respeito foram publicados por volta de 1998 [16].
- Uma imagem é representada por um conjunto de agrupamentos, que são extraídos da imagem através de um processo de segmentação. Para cada agrupamento, são capturadas informações específicas como cor, tamanho e localização espacial. Todas estas características podem ser extraídas automaticamente das imagens, sem nenhuma intervenção humana.
- Para a representação da cor foi usado o espaço de cores CIEL $*u*v*$ porque a sua linearidade permite que as diferenças perceptíveis entre as cores possam ser determinadas pela sua distância no espaço 3D.
- A representação do agrupamento armazena o seu retângulo de contorno, que requer apenas dois pontos. Como os pontos indicam as coordenadas do agrupamento dentro da imagem, eles indicam sua posição e ainda a distribuição espacial de cores na imagem.
- O tamanho do agrupamento também é armazenado. As imagens são convertidas em uma mesma escala para que se possa manter a consistência do tamanho armazenado do agrupamento em cada imagem.
- Duas imagens parecem ser similares se ambas têm agrupamentos de cores similares na mesma localização nas imagens.
- O MOSAIC permite consultas baseadas em imagem exemplo e recuperações baseadas em texto.
- A estrutura de índice é multinível, cada nível lida com uma característica, permitindo descartar imagens diferentes da imagem da consulta logo no primeiro nível.
- Para os níveis que usam cor e localização espacial foi usada uma estrutura R-tree, para o tamanho do agrupamento, um *array* de uma dimensão. Os nós folhas dos

níveis superiores apontam para as raízes dos níveis inferiores. Com exceção do nível mais baixo que apontam para as imagens físicas. Assim, a estrutura proposta consiste de muitas árvores integradas de uma maneira hierárquica. Esta estrutura facilita a velocidade da busca por imagens.

- Não usa realimentação de relevância pelo usuário durante o processo de busca.
- Para os testes do sistema foi usado um banco de dados com 12.000 imagens divididas em 18 categorias. As maiores categorias incluíam animais, obras de arte, computadores, alimentos, natureza, viagens, símbolos, transporte.
- Para avaliar o sistema foram usadas as medidas precisão e chamada. Concluiu-se que variando a quantidade de espaço de busca e o número de agrupamentos de cada imagem, pode-se obter melhor eficiência ou eficácia, mas será difícil otimizar as duas medidas. O desempenho da recuperação também depende muito do tipo de imagens que se tem e do tipo de representação usada.

6.11. WAVEGUIDE

- Os primeiros artigos a seu respeito foram publicados por volta de 1997 [17].
- Imagens são comprimidas com a técnica de *wavelet* e indexadas com cor, textura e descritores das formas dos objetos gerados no domínio *wavelet* durante o processo. O processo de extração das características é totalmente automático.
- A representação de imagens baseada em *wavelet* oferece uma boa estrutura para codificação e descrição de conteúdo. O sistema permite indexação, busca, navegação e compressão de dados imagens.
- Coeficientes *wavelet* quantizados são usados para representar as imagens porque oferecem uma representação muito boa de frequências espaciais, enquanto a estrutura plana permite uma computação rápida do histograma dos coeficientes *wavelet*.
- Para representar textura, a abordagem baseada em *wavelet* integra a multi-resolução e as propriedades de espaço e frequência. A transformada *wavelet*

permite que cada padrão de textura tenha sua própria estrutura de decomposição.

- Para descrever a cor foi usado o sistema de coordenadas YUV, através do histograma de cores da imagem. Para comparar dois histogramas de cores foram usadas a média, a variância e a distorção de um histograma como métricas de similaridade.
- Uma imagem na forma binária deve ser gerada de uma imagem por um algoritmo automático ou manual antes de uma análise de forma. A abordagem usada é baseada em *wavelet*, que para fazer a análise da característica forma.
- O sistema suporta consultas baseadas em exemplos.
- Um índice foi formado usando-se uma estrutura de pacotes *wavelet*, em forma de pirâmide.
- O banco de dados usado para testes tem 2.127 imagens naturais, incluindo paisagens, animais, edifícios, pessoas, flores, plantas e outros, da *Corel Professional Photos* CD-ROM.
- Não usa realimentação de relevância, mas existe um projeto para incorporar esta técnica no sistema.
- A eficiência da recuperação foi medida em termos de precisão e chamada, aplicadas a alguns conjuntos de consultas que operaram com combinações das características usadas. Os resultados foram satisfatórios, mas o desempenho depende altamente da escolha das consultas usadas como exemplo.

6.12. SOM-AIR

- Os primeiros artigos a seu respeito foram publicados por volta de 1998.
- O sistema [21] suporta recuperação de imagens geográficas baseadas em textura. A abordagem incorpora várias técnicas de processamento de imagens, incluindo os filtros de Gabor, realce de imagens e compressão de imagens para a criação de uma biblioteca digital de imagens.
- O sistema aplica técnicas de análise de informação para classificar imagens

geográficas com base na característica textura extraída automaticamente delas.

- A partir dos testes concluiu-se que em algumas circunstâncias usar apenas a característica textura não é suficiente para representar uma imagem.
- A seleção de um algoritmo para representar imagens varia com o tipo da imagem. Por exemplo, um algoritmo pode servir para trabalhar com imagens médicas, mas não ser adequado para imagens geográficas.
- A representação é gerada pelos filtros de Gabor e resulta em um vetor de características. É feita uma classificação das imagens com base na característica textura e é criado um metadado para representá-la.
- O usuário pode navegar pelo banco de dados através de um *browsing* ou formular uma consulta usando uma imagem exemplo.
- Foi adotada uma técnica automática de classificação de imagens, chamada SOM (*Self-Organizing Map*). Esta técnica é baseada em redes neurais.
- O sistema foi avaliado com e sem o uso de realimentação de relevância por parte do usuário, para que o desempenho nos dois casos pudesse ser medido.
- A coleção de imagens usada para testes consiste de 800 fotos aéreas fornecidas pelo *Map and Imagery Laboratory of Davidson Library* da Universidade da Califórnia, em Santa Bárbara (UCSB – *University of California, Santa Barbara*). As áreas que aparecem nas fotos incluem ilhas como *Channel*, *Anacapa* e *Santa Barbara* e ainda áreas residenciais, fazendas e praias.
- A avaliação foi conduzida de tal maneira que o desempenho do sistema foi diretamente comparado com o do ser humano. A qualidade dos testes foram medidas em função de precisão e chamada. E a conclusão é que o desempenho do sistema foi semelhante ao do usuário humano.

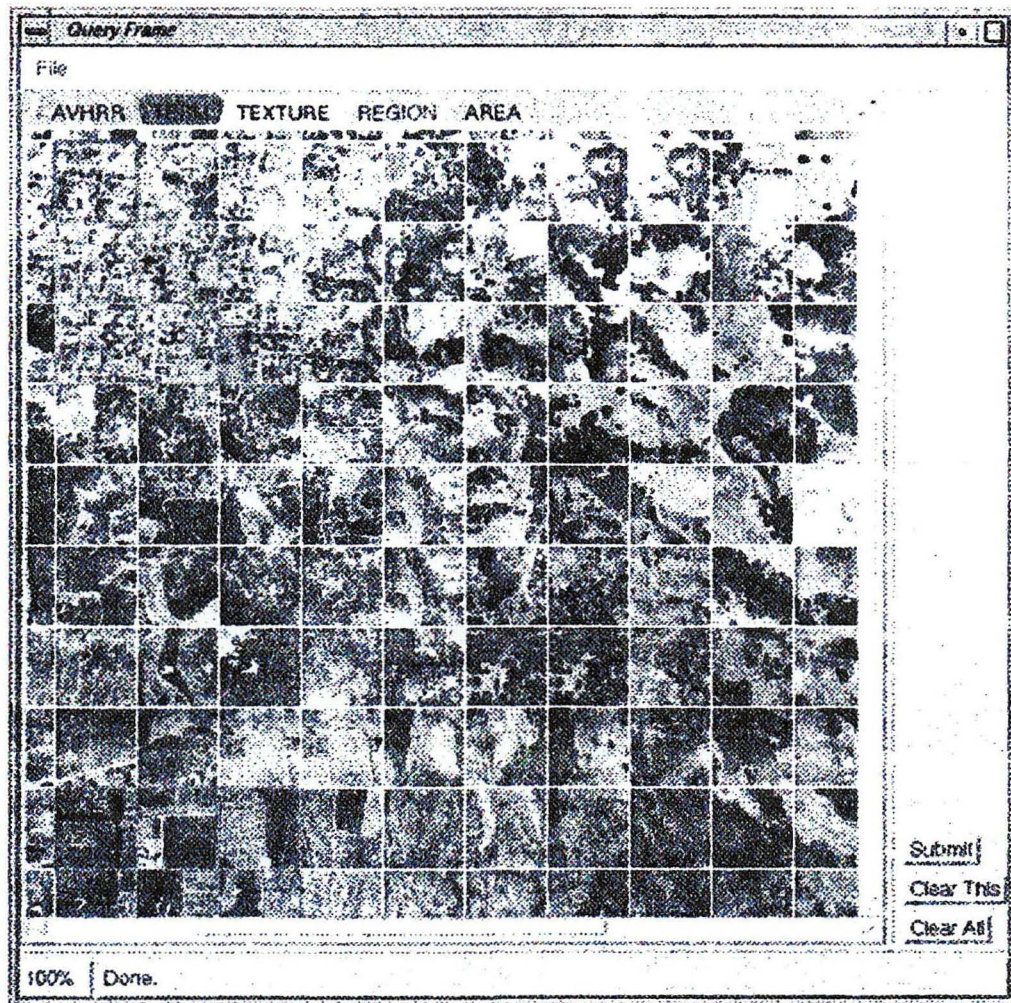


FIG 10.: EXEMPLO DE INTERFACE DO SISTEMA SOM-AIR. QUADRO DE CONSULTA.

6.13. MARCO

- Os primeiros artigos a seu respeito foram publicados por volta de 1994.
- É um sistema para aquisição, armazenamento, indexação e recuperação de imagens [22].
- O sistema MAGELLAN (*Map Acquisition of GEographic Labels by Legend Analysis*) é usado para processar as imagens e é capaz de extrair dados sobre o conteúdo dos mapas e seus objetos. Estes dados são usados para indexar as imagens.
- A imagem é segmentada e rotulada e para cada região é montado um vetor de características, como informações sobre o formato da região (circularidade,

retangularidade), informações que possam distinguir a região das demais, interseção entre objetos, centro de gravidade da região e as coordenadas da localização da região dentro da imagem.

- O usuário interage com o sistema MAGELLAN durante a extração dos dados, interpretando as legendas e os símbolos especiais dos mapas, que são importantes para a sua aplicação.
- Um índice é formado com características espaciais das imagens, ou seja, o relacionamento espacial entre os objetos encontrados na imagem.
- Outros dados, como a resolução do equipamento de aquisição e a escala do mapa, são armazenados e também podem ser usados para a criação de índices.
- O sistema MAGELLAN usa um vetor de características para representar cada símbolo do mapa apontado pelo usuário como importante para a sua aplicação.
- Os índices são construídos com base em dados textuais e espaciais das imagens.
- A estrutura de índice usada pelo MAGELLAN é a k-d tree.
- Os índices espaciais são implementados usando uma estrutura do tipo quadtree.
- As consultas podem ser feitas usando linguagem natural e são traduzidas para uma linguagem semelhante à linguagem SQL.
- Além disso, o usuário tem a sua disposição uma GUI (*Graphical User Interface*) que permite que o usuário escolha um tipo de consulta de seu interesse. A especificação da consulta é muito mais fácil, mas a variedade de consultas é limitada.
- Durante a consulta, o usuário escolhe a categoria de consultas através do botão apropriado e um ícone representando uma classe e o sistema retorna as imagens daquela classe.
- O sistema MAGELLAN permite o uso de realimentação de relevância feita pelo usuário.
- As imagens utilizadas são vários tipos de mapas geográficos escaneados. A coleção de teste é composta por mapas da Finlândia.

- A avaliação do sistema foi feita em termos de precisão e chamada e foi feita uma comparação empírica com base no tempo de execução de uma consulta, de acordo com a categoria escolhida. Os resultados demonstram a aplicabilidade do sistema.

6.14. PICHUNTER

- Os primeiros artigos a seu respeito foram publicados por volta de 1996.
- Uma contribuição do trabalho [6] [32] é a redução do problema da busca de imagens para três tarefas: projetar um espaço de ações do usuário; construir um modelo do usuário e selecionar uma estratégia para mostrar as imagens.
- As características usadas para representar uma imagem foram: o percentual de pixels com uma das onze cores (vermelho, verde, azul, preto, cinza, branco, laranja, amarelo, lilás, marrom, rosa); as dimensões da imagem, uma medida de contraste global; informação sobre a distribuição espacial de cores. Uma das versões do sistema ainda usou características textuais.
- Cada imagem tem uma entrada no vetor que representa a probabilidade de ser o alvo. A distribuição é alterada com base nas ações do usuário.
- Para representar cores, foi usado o histograma de cores.
- A principal abordagem para consulta é a busca por uma imagem alvo específica. Mas, as buscas ainda podem ser classificadas em três categorias: Busca pelo objetivo – o usuário deve encontrar uma imagem objetivo no banco de dados. (usada no sistema); Busca por categoria – o usuário procura por imagens que pertencem a determinada categoria; Navegação (Pesquisa aberta) – o usuário não tem um objetivo bem definido e navega pelo banco de dados.
- O PicHunter usa uma interface simples projetada para busca por imagens alvo com o mínimo de treinamento.
- O sistema seleciona algumas imagens na inicialização do mesmo. O usuário seleciona as imagens mais semelhantes à imagem alvo e o sistema busca novas imagens no banco de dados de acordo com a escolha do usuário. O processo é

repetido até que o usuário encontre a imagem procurada ou desista da busca.

- Um vetor de características é usado para representar o esquema de indexação do sistema.
- Usa realimentação de relevância com uma abordagem Bayesiana, representando sua incerteza sobre o objetivo do usuário através de uma distribuição de probabilidade sobre possíveis objetivos. O sistema foi avaliado com e sem o uso da realimentação. Os resultados indicaram que o desempenho aumenta muito depois que o usuário interage com o sistema uma vez, mas a mudança passa a ser menor depois da segunda interação.
- O banco de dados utilizado para os testes foi um conjunto de 4522 imagens da Corel, cujos temas variavam de cavalos, jardins, águias, monumentos egípcios e outros. A versão do sistema que suportava consultas baseadas em texto foi submetida a testes usando-se também anotações textuais.
- A medida de desempenho que foi usada durante todo o desenvolvimento do projeto é o número médio de imagens necessárias para convergir para o alvo específico. Esta medida foi útil para comparar as várias versões do sistema e, segundo [32] é um grande passo para se estabelecer um *benchmark* para comparações úteis entre sistemas de recuperação de imagens por conteúdo.
- A avaliação foi feita com base na satisfação do usuário. As várias versões do sistema foram submetidas a testes e os resultados, avaliados. Sem uma medida objetiva de avaliação, os resultados foram surpreendentes, dependendo do nível de conhecimento do usuário.
- Uma possível extensão seria modificar o sistema de forma que permita consultas por categoria.

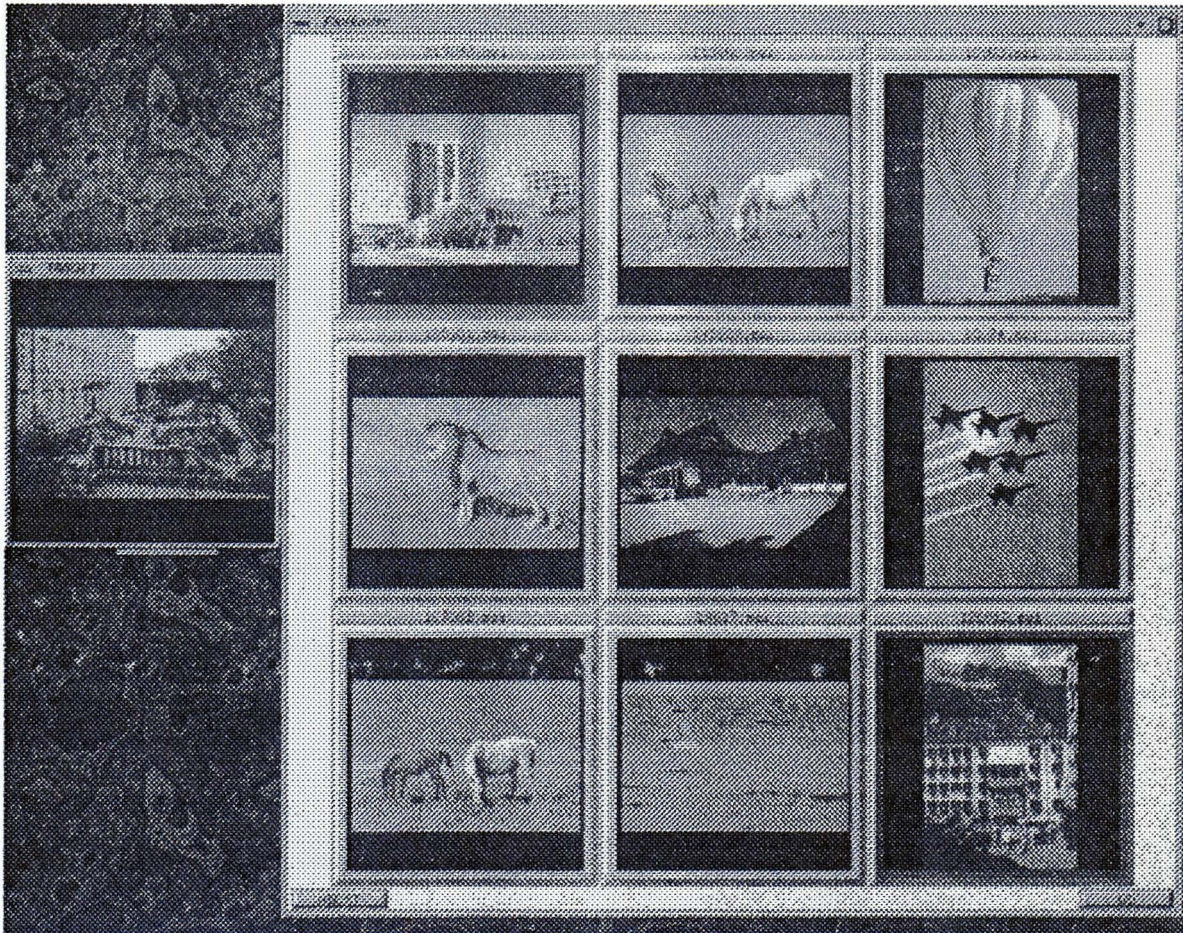


FIG 11.: EXEMPLO DE INTERFACE DO SISTEMA PICHUNTER.

6.15. FOCUS

- FOCUS (*Fast Object Color-based qUery System*) [35].
- Os primeiros artigos a seu respeito foram publicados por volta de 1996.
- Usa a característica cor dentro do espaço de cores HSV (*hue, saturation, value*) e um histograma de cores para representá-la. Para armazenar o histograma, não foram consideradas as cores com valor zero. Foi usado um método de detecção de picos do histograma para que fossem consideradas apenas as cores mais frequentes na imagem.
- A imagem é dividida em células e um histograma é construído para cada célula e os picos são armazenados.
- Não foi possível usar segmentação porque o plano de fundo de uma imagem pode

ser um objeto de destaque de outra imagem.

- Usa os relacionamentos espaciais entre as regiões de cores.
- A consulta é feita com o uso de uma imagem exemplo e o sistema procura por imagens semelhantes.
- O objeto procurado pode estar embutido em muitas imagens diferentes e pode ocupar uma parte bem pequena da imagem.
- A recuperação é feita em duas fases. Na primeira, foi dada mais ênfase à velocidade da busca. Na segunda, o objetivo é remover resultados falsos da lista produzida pela primeira fase. A segunda fase só é iniciada se o usuário optar por executá-la.
- Para a construção do índice foi usada uma estrutura de árvore B+ para armazenar os picos dos histogramas locais.
- A localização espacial das cores é armazenada usando-se um grafo, cujos nós representam os picos dos histogramas locais e cujas ligações são feitas com base nos picos de regiões adjacentes.
- Não usa realimentação de relevância.
- O banco de dados usado para testes consiste de 1.200 imagens coloridas diferentes. Existem 400 anúncios de revistas e 800 imagens coloridas de natureza, incluindo pássaros, peixes, flores, animais e vegetais.
- O desafio maior é encontrar um objeto em um dos anúncios das revistas, já que o objeto pode estar em qualquer posição dentro da imagem, em qualquer tamanho e orientação.
- O sistema foi avaliado usando precisão e chamada e os resultados foram satisfatórios, considerando-se a relação entre as duas medidas.

6.16. FOOTSCAN

- O sistema está sendo desenvolvido pela UFPR – Universidade Federal do Paraná, pelo departamento de Informática, em parceria com o departamento de Pediatria da

Unidade de Neonatologia do Hospital de Clínicas do Paraná.

- O principal objetivo do sistema é medir o tempo de gestação de um bebê recém-nascido prematuramente. Com esta informação, será possível administrar o tratamento correto para a sobrevivência da criança e o seu neurodesenvolvimento.
- O método FootScan pretende determinar a idade gestacional através da análise dos sulcos e pregas cutâneas da superfície da planta do pé do recém-nascido. O conjunto de imagens que será utilizado pelo sistema é composto, portanto, por imagens dos pés dos recém-nascidos obtidas entre 6 e 12 horas de vida, através de um sensor.

Considerando esta amostra de sistemas de recuperação de imagens por conteúdo, foi elaborada uma tabela que apresenta os principais itens de cada sistema, como as características armazenadas, os métodos de indexação e de consulta e o tipo de avaliação usada. A partir da tabela, foi estabelecida uma classificação dos sistemas estudados e algumas conclusões podem ser observadas para auxiliar a construção de novos sistemas deste tipo.

Outros sistemas foram pesquisados [47], [46], [45], [44], [43], [42], [18], mas não fazem parte da coleção porque os artigos publicados a seu respeito não apresentaram todas as informações necessárias ao estudo que se segue.

7. CLASSIFICAÇÃO DE SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO

A partir da coleção de sistemas de recuperação de imagens por conteúdo apresentada, foi elaborada uma tabela que apresenta os principais itens de cada sistema, como as características armazenadas, o tipo de representação para cada característica, os métodos de indexação e de consulta, a escolha pelo uso de realimentação de relevância ou não, e o tipo de avaliação usada. A partir da tabela, foi estabelecida uma classificação dos sistemas estudados e algumas conclusões podem ser observadas para auxiliar a construção de novos sistemas deste tipo.

Sistema	Características	Representação	Indexação	Método de Consulta	Realimentação de Relevância	Categoria de imagens	Avaliação
QBIC	Cor, textura e forma	Para cores, histograma de cores. Para textura, uma representação matemática de aspereza, contraste e direcionalidade. Para forma, características como área, circularidade, direção quanto ao eixo maior e um conjunto de ângulos tangentes ao perímetro do objeto.	Usa redução de dimensão, R*-tree	Baseadas em imagens exemplo, esboços ou desenhos feitos pelo usuário	Não	7.450 imagens da <i>Mediasource Series</i> de imagens e áudio da <i>Applied Optical Media Corp.</i> , 4.100 imagens do CD <i>PhotoDisc</i> , 950 imagens da coleção <i>Corel Professional Photo CD</i> , e 450 imagens da coleção IBM.	A avaliação do sistema foi feita de forma subjetiva, em função de testes com alguns exemplos de consultas.
Virage	Metadado e características, extração manual, cor e forma extraídas automaticamente, layout de cores, textura e estrutura.	Histogramas para cor. Para forma, lista com cor, área, razão entre perímetro e área e centralidade.	Estrutura de Quadtree.	Consulta por exemplo, com imagem ou valores de características	Sim	imagens aéreas, nas quais os objetos de interesse são edifícios, veículos terrestres, aeronaves, estradas.	a resposta do usuário, indicando se a recuperação foi satisfatória e a eficiência do sistema, em função do domínio da aplicação.
Photobook	Conteúdo da imagem e anotações textuais. Cor, forma e faces	FEM para forma. KLT para faces.	Árvores	O usuário escolhe uma imagem do banco de dados ou uma categoria e o sistema retorna imagens semelhantes.	Sim	Pessoas, quadros de vídeo, ferramentas manuais, peixes e imagens médicas 3-D	Precisão e rechamada
VisualSeek	Cor e textura	Conjuntos de cores para cor e Transformada de Wavelet para textura	Árvore binária para cor, R-tree para localização espacial	Suporta consultas baseadas em características visuais e seus relacionamentos espaciais.	Sim	Mais de um milhão de imagens da WWW.	Gráfico precisão versus rechamada
Netra	Cor, forma, textura e localização espacial	cor através do histograma de cores; forma, através de	Redes Neurais; para	Consulta por imagem exemplo ou por	Não	Foram usadas imagens aéreas e de satélites para testes. 2.500	A avaliação foi feita com

Sistema	Características	Representação	Indexação	Método de Consulta	Realimentação de Relevância	Categoria de imagens	Avaliação
		pontos da borda do objeto; textura, decomposição de Gabor	características s textura e forma, estrutura de dados SS-tree; para localização espacial, quadtree	características.		imagens coloridas da galeria de fotos da Corel	base em exemplos de consultas e nos resultados obtidos.
MARS	cor, textura, forma e lay-out de cor, e ainda uma anotação manual sobre o conteúdo das imagens	cor é representada por um histograma de cores; Wavelet, matriz de ocorrência e vetor de características, para textura; forma, descritor de Fourier	estrutura de dados do tipo árvore híbrida que incorpora as vantagens da k-d tree e da estrutura baseada nos limites das regiões.	O usuário escolhe algumas imagens da coleção e especifica as características de interesse.	Compara com e sem	uma coleção de imagens de artefatos antigos do <i>Fowler Museum of Cultural History</i> , que consiste de 286 imagens e a coleção Corel de imagens, com cerca de 70.000 imagens da natureza.	O sistema foi avaliado em termos de precisão e rechamada
Surfimage	Cor, forma e textura e ainda características de alto nível	A cor, através de um histograma; textura, através de <i>wavelet</i> ; forma, pela transformada de Fourier.	Em termos estatísticos, cada imagem tem uma distribuição paramétrica.	Por imagem exemplo, classes de imagens, consultas múltiplas e refinamento de consulta	Sim	Vistex, com 384 texturas homogêneas, Columbia, com 1440 imagens de 20 objetos 3D, o ORL, de faces, o MIT, com 7.652 imagens de faces, o BTphoto, de cidades e países, e outros	Precisão e rechamada. Usa o gráfico PR
SamMatch	Cor e forma e tamanho das regiões.	Histogramas de cores e uma tabela para a forma das regiões	através de um filtro	Por imagem exemplo ou esboço	Não	15.808 imagens de rosas, pássaros, balões, prédios, céus, estátuas, comidas...	média da rechamada e tempo e espaço
Epic	Nome do fotógrafo, data da foto, palavras-chave e características	Características espaciais por pares ordenados: um rótulo representando um objeto da imagem e as coordenadas do		duas versões: uma que permite apenas consultas visuais e outra que permite	Sim	800 fotografias tiradas da <i>National Trust for Scotland</i> . Envolvem exteriores de castelos, jardins, lagos,	centrada na satisfação do usuário, e rechamada.

Sistema	Características	Representação	Indexação	Método de Consulta	Realimentação de Relevância	Categoria de imagens	Avaliação
	espaciais (objeto e suas coordenadas).	objeto dentro da imagem.		apenas consultas textuais.		ornamentos, arranha-céus e paisagens.	
MOSAIC	Cor, tamanho do cluster e localização espacial	Clusters (retângulo de contorno), espaço de cor CIEL *u*v*	Estrutura multinível;; Para os níveis que usam cor e localização espacial foi usada uma estrutura R-tree, para o tamanho do <i>cluster</i> , um array.	Baseadas em texto ou visuais	Não	12.000 imagens de animais, arte, computadores, alimentos, natureza, viagens, símbolos, transporte.	precisão e rechamada
WaveGuide	Cor, textura e forma dos objetos	Wavelet para textura e forma, sistema YUV para cor, através do histograma de cores da imagem	uma estrutura de pacotes <i>wavelet</i> .	Consultas baseadas em exemplo	Não	2.127 imagens naturais, incluindo paisagens, animais, edifícios, pessoas, flores, plantas e outros, da <i>Corel Professional Photos CD-ROM</i> .	precisão e rechamada
SOM-AIR	Textura, extraída automaticamente das imagens.	gerada pelos filtros de Gabor e resulta em um vetor de características.	SOM, baseada em redes neurais.	<i>browsing</i> ou formular uma consulta usando uma imagem exemplo.	com e sem o uso de realimentação de relevância por parte do usuário	800 fotos aéreas, que incluem ilhas como <i>Channel</i> , <i>Anacapa</i> e <i>Santa Barbara</i> e ainda áreas residenciais, fazendas e praias.	precisão e rechamada
MARCO	Formato da região (circularidade, retangularidade), exceções, interseção entre objetos, centro de gravidade e as coordenadas da localização.	Vetor de características.	k-d tree para índices textuais e quadtree para espaciais.	O usuário escolhe a categoria de consultas através do botão apropriado e um ícone representando uma classe. Outra possibilidade é usar linguagem natural.	Apenas durante o processo de extração de características.	Mapas da Finlândia	Em termos de precisão e rechamada e tempo de execução das consultas.

Sistema	Características	Representação	Indexação	Método de Consulta	Realimentação de Relevância	Categoria de imagens	Avaliação
PicHunter	o percentual de pixels com uma das onze cores; as dimensões da imagem, uma medida de contraste global; a distribuição espacial de cores. E ainda características textuais.	cores, o histograma de cores.	vetor de características	busca por uma imagem alvo específica; Busca por categoria; Navegação	Usa realimentação de relevância com uma abordagem bayesiana, e foi avaliado sem o uso também.	4522 imagens da Corel, cujos temas variavam de cavalos, jardins, águias, monumentos egípcios e outros	número médio de imagens necessárias para convergir para o alvo específico; com base na satisfação do usuário
FOCUS	Cor e seus relacionamentos espaciais	Histograma de cores, dentro do espaço de cores HSV	árvore B+	uso de uma imagem exemplo	Não	400 anúncios de revistas e 800 imagens coloridas de natureza, incluindo pássaros, peixes, flores, animais e vegetais.	precisão e rechamada

TAB2.: COLEÇÃO DE SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO

Com o estudo destes sistemas e analisando a tabela, algumas considerações podem ser feitas:

1. A grande maioria dos sistemas usa a característica cor. Essa escolha se justifica porque essa é uma característica facilmente observada por seres humanos e porque o processamento desta característica é facilitado pela existência de vários sistemas de cores (como RGB, YUV e outros) e porque pode ser representada por campos numéricos, o que torna o processamento mais rápido. Outra característica importante é o fato de que esta característica é invariante quanto ao tamanho da imagem e à rotação da mesma.
2. A escolha da característica cor implica, na maioria das vezes, no uso de uma representação do tipo histograma de cores. A comparação entre dois histogramas para testar a similaridade entre duas imagens é um processo custoso, mas ainda é uma das melhores opções.
3. uso da característica textura é mais freqüente em sistemas que trabalham com imagens do tipo paisagem. Isso porque ela permite a diferenciação entre a imagem de um oceano e a de um carro azul, por exemplo, já que a textura reconhece outras particularidades que se repetem em uma determinada região da imagem, além da cor.
4. Para representar textura, na maioria das vezes é usada a abordagem *wavelet*. Um dos motivos desta escolha é o fato de que esta abordagem consiste de funções que decompõem os sinais em diferentes componentes de freqüência e analisa cada componente em diferentes resoluções. Além disso, a transformada de *wavelet* não possui apenas uma função base e sim uma classe delas. Dependendo da aplicação desejada e da natureza das imagens, a escolha da função base é um diferencial.
5. A característica forma foi usada principalmente quando existem pouco objetos nas imagens e não há muitos casos de sobreposição. Isso acontece com maior freqüência quando há muitos objetos semanticamente diferentes e que podem possuir a mesma cor. Quando há muitos casos de sobreposição, a identificação de

um objeto se torna muito difícil.

6. A representação da característica forma normalmente é feita através de um vetor de características envolvendo detalhes do objeto, como área, perímetro, centro, circularidade, direção com relação a um determinado eixo, posição relativa dentro da imagem.
7. uso de características semânticas do tipo texto é mais freqüente em sistemas cujos usuários são especialistas do contexto da área de aplicação. Nestes casos, há a possibilidade de o anotador e o usuário terem a mesma visão sobre a imagem e menor probabilidade de haver ambigüidade na representação. Esta característica normalmente é armazenada com o uso de uma string de caracteres.
8. Ainda podem ocorrer situações em que o usuário esteja interessado em dados como a data da aquisição da imagem ou o nome do fotógrafo. Nestes casos, é necessária uma anotação manual destas características e a representação deve ser feita usando-se campos específicos, como o tipo data para esta informação e texto para o nome do fotógrafo.
9. A escolha das características a serem usadas para representar uma imagem depende do contexto da aplicação. Sistemas que tenham interesse por exemplo em encontrar imagens de satélite que apresentam paisagens devem se preocupar com os padrões de textura da imagem. Sistemas que trabalham com apenas um tipo de imagem, de carros por exemplo, podem optar por trabalhar apenas com a característica cor, ou incluir uma informação textual que indique o modelo do carro ou o ano, já que é muito provável que o usuário tenha conhecimento sobre esta informação.
10. A escolha da representação mais adequada para cada característica deve ser feita com base nas abordagens disponíveis e no tipo de imagens que pertencem ao banco de dados. Uma representação que use campos numéricos tem a facilidade de aumentar a velocidade do processamento, mas tem a desvantagem de não oferecer informação semântica sobre a imagem. Além disso, a própria representação numérica pode provocar uma sobrecarga computacional.

11. A escolha pelo método de consulta visual, principalmente através de imagem exemplo, é praticamente unânime e pesquisas como [34] revelam a preferência do usuário com relação a esse método de trabalho. O usuário, em geral, tem mais facilidade em mostrar um exemplo do que está procurando do que explicitar sua expectativa através de palavras.
12. Outra escolha praticamente unânime é o uso de uma estrutura do tipo árvore como índice para auxiliar a busca. Além disso, a maioria dos sistemas se preocupa em usar uma abordagem para reduzir a dimensão do vetor de características da imagem.
13. Ainda o uso da abordagem de realimentação de relevância durante o processo de busca foi bastante aceita e testada por vários pesquisadores [4], [11], [31]. O uso desta abordagem pode ser aplicado a qualquer tipo de sistema. Uma preocupação deve ser a implementação de um critério de parada da consulta, para que o sistema não permita que o usuário não mude o objetivo muitas vezes durante a busca, o que poderia degradar em muito o resultado da busca.
14. Uma dificuldade apontada por vários pesquisadores foi encontrar um banco de dados para testar o sistema desenvolvido e comparar os resultados produzidos com os resultados de outros sistemas da mesma natureza. As categorias de imagens mais utilizadas pelos sistemas estudados foram as imagens coloridas da Corel, incluindo principalmente fotos de elementos da natureza, imagens de monumentos ou obras de arte ou imagens de satélite.
15. Quanto à avaliação dos sistemas, observou-se um consenso no que concerne à necessidade de métricas mais adequadas a sistemas desta natureza. Os sistemas estudados foram avaliados por seus desenvolvedores através de precisão e chamada (ou sua representação gráfica) ou através da resposta do usuário, indicando se o desempenho do sistema foi satisfatório às suas expectativas.
16. Considerando os sistemas estudados e as observações anotadas acima, é possível estabelecer uma classificação destes sistemas em função dos vários itens

analisados. Alguns sistemas podem pertencer a duas ou mais categorias.

17. Quanto à categoria de imagens utilizadas para testes:

17.1. Sistemas gerais que trabalham com imagens coloridas de vários temas – QBIC, Photobook (módulo de forma), VisualSEEk, Netra (quando trabalha com imagens da Corel), MARS (quando trabalha com imagens da Corel), Surfimage, SamMatch, MOSAIC, PicHunter, FOCUS e MARCO. Para estes sistemas, observa-se o interesse maior em determinar a similaridade em função das cores em destaque nas imagens. Houve uma preocupação em considerar uma característica textual além da cor, porque a cor não oferece nenhum poder semântico. Outra preocupação foi oferecer ao usuário um método de consulta que fosse simples e não exigisse nenhum conhecimento prévio em uma linguagem de consulta.

17.2. Sistemas de paisagens que trabalham com texturas para encontrar imagens de paisagens, museus, ou algo semelhante, quando a cor não é suficiente para distinguir as imagens – MARS, Epic e WaveGuide. As imagens de museus normalmente têm um número pequeno de objetos em destaque em um plano de fundo facilmente identificável [33].

17.3. Sistemas de imagens aéreas que trabalham com imagens geradas por equipamentos de satélites – Virage, Netra e SOM-AIR.

17.4. Sistemas de imagens médicas que trabalham com imagens geradas por equipamentos de radiologia, ressonância magnética e outros que apresentam imagens da anatomia do corpo humano – FootScan.

Para sintetizar o estudo realizado sobre a coleção de sistemas apresentada, foi elaborada uma taxonomia.

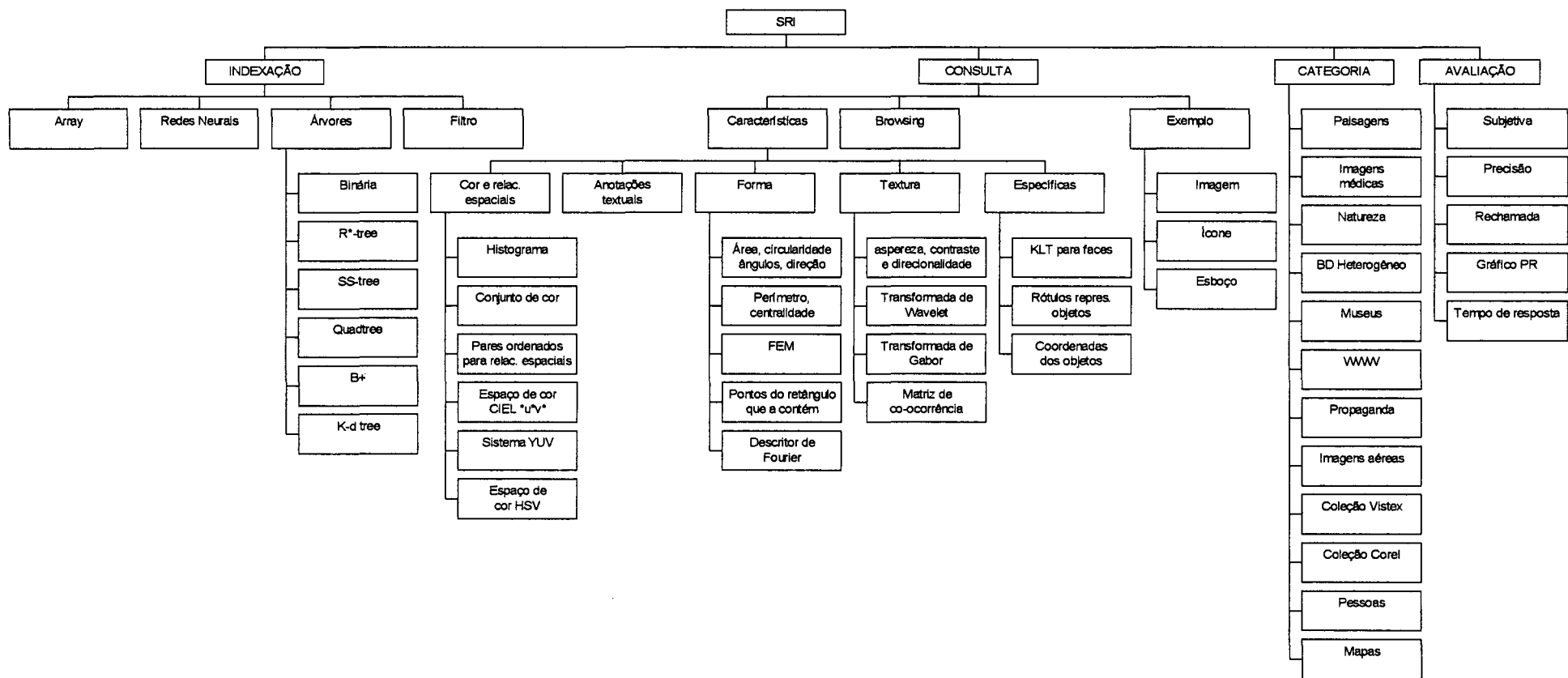


FIG.12: PRINCIPAIS PECULIARIDADES DOS SISTEMAS DE RECUPERAÇÃO DE IMAGENS POR CONTEÚDO

SISTEMA	CARACTERÍSTICAS	REPRESENTAÇÃO	INDEXAÇÃO	MÉTODO DE CONSUTA
QBIC	Cor	Histograma	R*-tree	Exemplos ou esboços
	Textura	Aspereza, contraste e direcionalidade		
	Forma	Área, circularidade, direção e ângulos		
Virage	Cor	Histograma	Quadtree	Exemplo ou características
	Textura			
	Forma	Cor, área, perímetro e centralidade		
Photobook	Texto		Árvore	Exemplo
	Cor			
	Forma	FEM		
	Face	KLT		
VisualSEEk	Cor	Conjunto de cores	R-tree	Características visuais e seus relacionamentos espaciais
	Textura	Wavelet		
Netra	Cor	Histograma	Redes Neurais, SS-tree e quadtree	Exemplo ou características
	Forma	Borda do objeto		
	Textura	Gabor		
	Localização espacial			
MARS	Cor	Histograma	Árvore híbrida	Exemplo e características
	Forma	Descritor de Fourier		
	Textura	Wavelet e matriz de co-ocorrência		
	Texto			
Surfimage	Cor	Histograma	Distribuição paramétrica	Exemplo e classes de imagens
	Forma	Descritor de Fourier		
	Textura	Wavelet		
	Texto			
SamMatch	Cor	Histograma	Filtro	Exemplo ou esboço
	Forma	Tabela		
	Tamanho das regiões			
Epic	Nome do fotógrafo			Visuais ou textuais
	Data da foto			
	Texto			
	Carac. espaciais	Pares ordenados: rótulo e localização		
MOSAIC	Cor	Espaço CIEL *u*v*	Estrutura multinível	Texto ou visuais
	Cluster	Retângulo de contorno		

WaveGuide	Cor	Sistema YUV	Estrutura de pacotes Wavelet	Exemplo
	Textura	Wavelet		
	Forma	Wavelet		
SOM-AIR	Textura	Transformada de Gabor	SOM	Exemplo ou navegação
MARCO	Características da região	Vetor de características	k-d tree e quadtree	Categoria de imagens ou linguagem natural
PicHunter	Cor	Histograma	Vetor de características	Exemplo, categoria ou navegação
	Texto			
FOCUS	Cor	Histograma, no espaço HSV	Árvore B+	Exemplo

TAB3: FORMA SINTÉTICA DA TABELA DE CARACTERÍSTICAS DA COLEÇÃO DE SISTEMAS

8. CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho apresenta uma extensa revisão bibliográfica sobre a base de sistemas de recuperação de imagens por conteúdo, dando destaque à extração de característica, indexação multidimensional e projeto dos métodos de consulta disponíveis no sistema.

Alguns sistemas de recuperação de imagens por conteúdo foram estudados e vários itens sobre o seu projeto foram destacados. Com base neste estudo, foi elaborada uma classificação para estes sistemas, em função do conjunto de imagens usado para testes, que indica o contexto da aplicação, e suas principais preocupações durante a fase de projeto.

Uma taxonomia foi elaborada para sintetizar o estudo sobre o projeto de sistemas de recuperação de imagens por conteúdo. O estudo abrange os vários itens analisados durante a fase de projeto e oferece uma visão geral sobre os sistemas estudados.

Também foi feito um estudo sobre as métricas de avaliação existentes para sistemas desta natureza. A conclusão é que ainda não existem métricas específicas para este contexto. As métricas precisão e chamada foram adaptadas da área de recuperação de informações e nem sempre são capazes de avaliar adequadamente os sistemas em estudo. Alguns autores [32] optaram por desenvolver um método de avaliar o sistema centrado na satisfação do usuário, apesar de ser um método subjetivo.

As escolhas por determinadas características, por uma estrutura de indexação e por um método de consulta adequado ainda estão dependentes do contexto da aplicação. O tipo de imagens que pertencem ao conjunto com o qual o sistema terá que lidar influencia de modo considerável no projeto do sistema.

A expansão da WWW (*World Wide Web*) é espantosa. A cada dia muitos documentos são adicionados à Web, sendo muitos deles imagens. Para melhor organizar e recuperar estas informações, são necessários sistemas de busca baseados

em Web. Já existem soluções para lidar com informações do tipo texto, como os sites AltaVista e Inforseek. Para imagens na Web, alguns trabalhos bons estão tomando lugar [31], mas ainda não se comparam aos similares para texto.

Outro problema é a falta de uma ligação entre os índices de características visuais de baixo nível e informações de nível semântico mais significativo [5]. Observa-se que navegadores e comparação baseada em texto ainda são mais populares que busca baseada em características. Os sistemas que usam características visuais resolvem problemas particulares, para usuários e conteúdos de imagens diferentes [5].

A maioria dos sistemas de busca de imagens para Web consegue lidar com um número pequeno de imagens se for considerada a enorme coleção existente [5]. E, para isso, uma busca por todas as imagens não diminui muito o desempenho do sistema. Mas, como as coleções estão aumentando cada vez mais, a velocidade da recuperação está se tornando um problema. Embora esta área tenha conseguido alguns progressos [23], ainda há a necessidade urgente por técnicas de indexação multidimensional realmente eficientes.

Em [23], os autores sugerem o uso de mineração de dados (*datamining*) como uma boa ferramenta para auxiliar o processo de indexação, já que os objetos podem agora ser vistos como pontos do espaço 2-D ou 3-D, revelando agrupamentos potenciais, correlações entre atributos e outras regularidades que *datamining* procura.

Durante a fase de projeto da interface, a maioria dos sistemas de recuperação de imagens por conteúdo limitam-se a tratar determinados métodos de consulta. O desenvolvimento de uma linguagem para consultas visuais é uma das tarefas mais difíceis para os pesquisadores [4]. Mas, o desenvolvimento de uma linguagem que consiga fazer com que o usuário ofereça informação suficiente para que o sistema determine a similaridade entre duas imagens e, ao mesmo tempo, não exija que o usuário tenha um conhecimento prévio sobre linguagens de consultas a bancos de dados convencionais, é de interesse para muitos pesquisadores.

Uma técnica avança em função do seu critério de avaliação de domínio [5].

Precisão e chamada são usadas em recuperação de informação baseada em texto. Uma boa métrica pode levar a técnica para a direção correta enquanto que uma métrica não muito eficiente pode enganar o esforço do pesquisador [5]. Correntemente, alguns sistemas de Recuperação de Imagens medem seu desempenho baseado na relação entre custo e tempo para encontrar as imagens corretas [5]. Outros avaliam desempenho usando precisão e chamada, termos emprestados da recuperação baseada em texto.

Estes critérios medem o desempenho dos sistemas atuais, mas estão longe de serem considerados satisfatórios. O maior motivo que causa esta dificuldade é a subjetividade do conteúdo da imagem [5]. Isto é, a subjetividade da percepção da imagem impede a definição de um critério de avaliação objetivo. Mas ainda é preciso encontrar um caminho de avaliar o desempenho do sistema para guiar os esforços das pesquisas na direção certa.

Para evitar que este estudo se limite a adaptações de técnicas existentes em recuperação de informação, é importante que seja feita uma revisão sistemática de métodos de avaliação usados em recuperação de informação. Além disso, é preciso verificar quais deles realmente são convenientes ao estudo da recuperação de imagens por conteúdo.

Uma tarefa igualmente importante é estabelecer um conjunto de teste bem balanceado, de forma que seja possível fazer uma avaliação confiável e, ao mesmo tempo, não prejudique a velocidade dos testes. Para compressão de vídeo, a comunidade MPEG desenvolveu seqüências de testes bem balanceadas. Para Recuperação de Informação baseada em texto, também já existe teste de base em alta escala. Para recuperação de imagens, a comunidade MPEG-7 recentemente começou a coletar dados para testes [5].

O usuário final de um sistema de Recuperação de Imagens é um ser humano. Portanto, o estudo da percepção humana de conteúdo de imagens a um nível psicológico é crucial. Este tópico está ganhando cada vez mais atenção [5], com o

objetivo de se explorar a percepção do ser humano quanto ao conteúdo de uma imagem. Além disso, é interessante que se possa integrar este “modelo humano” aos sistemas de Recuperação de Imagens. As equipes que estudam este tópico são as mesmas que começaram os estudos sobre realimentação de relevância [8] em Recuperação de Imagens. Isto porque é preciso usar a realimentação de relevância para “decodificar” a percepção humana depois de analisada [5].

Estudos mais recentes da percepção humana têm seu foco nos aspectos psicológicos da percepção humana. Papathomas et al. [6] conduziu experimentos estudando a importância de usar (a) informação sobre a semântica, (b) memória de entrada, e (c) julgamento relativo ou absoluto de similaridades das imagens. Os resultados mostraram que o melhor desempenho é obtido apenas usando dados semânticos, com memória e julgamento de similaridade relativo. A combinação de dados visuais e semânticos somente alcança o segundo melhor resultado. Uma das razões que dados visuais não ajudam neste caso pode ser o tamanho limitado do conjunto de teste – se o conjunto de dados não é grande o suficiente, o sistema pode não ser capaz de utilizar a informação adicional dos dados visuais.

O ideal é que se direcione as pesquisas no sentido de criar medidas de avaliação que sejam capazes de refletir com precisão a satisfação de um usuário humano, especialista no contexto da aplicação.

Muitos artigos das áreas de Bancos de Dados e Visão Computacional têm como título “Banco de Dados de Imagens” [5]. Mas, na realidade, a maioria dos sistemas destas áreas não é verdadeiramente de Bancos de Dados de Imagens. São palavras-chave baseadas em texto, no caso da área de Bancos de Dados, e apenas um arquivo contendo muitas imagens, no caso da Visão Computacional. Existem muitas pesquisas [37] para se construir um verdadeiro Banco de Dados de Imagens, mas os sistemas ainda não estão completos. Um trabalho foi desenvolvido recentemente para modelagem de Banco de Dados de Imagens Médicas [49].

Um sistema de Banco de Dados de Imagens de sucesso requer um esforço de

pesquisa interdisciplinar [5]. Além da integração de Gerenciamento de Bancos de Dados e Visão Computacional, a pesquisa da tradicional área de Recuperação de Informação é também uma parte indispensável. Apesar de o foco das pesquisas da área de Recuperação de Informação ser a recuperação de documentos baseada em texto, muitos modelos e técnicas de recuperação úteis podem ser adaptados à Recuperação de Imagens [5]. Alguns exemplos de sucesso incluem a adaptação do modelo de recuperação booleana em Recuperação de Imagens, e a utilização de realimentação de relevância em Recuperação de Imagens [8] [10].

Existem ainda muitas questões abertas relacionadas à pesquisa de recuperação de imagens por conteúdo a serem resolvidas. O sucesso destas pesquisas futuras fará com que os sistemas possam ser colocados em prática e considerados satisfatórios a vários usuários.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Vendrig, J., *Filter Image Browsing: a study to image retrieval in large pictorial databases.*, Department of Intelligent Sensoric Information Systems, Faculty WINS, Universiteit van Amsterdam, Msc Thesis, February, 1997.
- [2] Smith, J. R. and Chang, S., *Tools and Techniques for Color Image Retrieval.*, In Proc. of the SPIE Storage and Retrieval for Image and Video Database III, volume 2460, Msc Thesis, February, 1996.
- [3] Aksoy, S. and Haralick, R., *Content-based Image Database Retrieval Using Variances of Gray Level Spatial Dependencies.*, IAPR International Workshop on Multimedia Information Analysis & Retrieval (MINAR'98), August 1998, Hong Kong. Pages 3 – 19.
- [4] Gupta, A. and Jain, R., *Visual Information Retrieval.*, Communications of the ACM, may, 1997/Vol. 40. Pages 71 – 79.
- [5] Huang, T. S., Rui, Y., and Chang, S.-F., *Image Retrieval: Past, Present and Future.*, Journal of Visual Communication and Image Representation, Volume 10, 1999. Pages 1 – 23.
- [6] Papathomas, T. V., Conway T. E., Cox, I. J., Ghosn, J., Miller, M. L., Minka, T. P. and Yianilos, P. N., *Psychophysical studies of the performance of an image database retrieval system.*, IS&T / SPIE Conf on Human Vision and Electronic Imaging III, 1998. Pages 591 – 602.
- [7] Wu, D., Agrawal, D., Abbadi, A. E., Singh, A. and Smith, T. R., *Efficient Retrieval for Browsing Large Image Databases.*, Proceedings of the fifth International conference on Information and Knowledge management, 1996. Pages 11 – 18.
- [8] Rui, Y., Huang, T. S., Ortega, M. and Mehtotra, S., *Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval.*, IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Segmentation, Description and Retrieval of Video Content, vol. 8, no 5, September 1998. Pages 644 – 655.
- [9] Marsicoi, M. D., Cinque, L. and Levialdi, S., *Indexing pictorial documents by their content: a survey of current techniques.*, Image and Vision Computing, 15, February, 1997. Pages 119 – 141.
- [10] Rui, Y., Huang, T. S., and Mehtotra, S., *Content-Based Image Retrieval with Relevance Feedback in MARS.*, Proceedings of IEEE International conference on Image Processing, Santa Barbara, California, USA, October, 1997, ppII815 – 818.
- [11] Pentland, A., Picard, R. W. and Sclaroff, S., *Photobook: Content-Based Manipulation of Image Databases.*, International Journal of Computer Vision, 18 (3), 1996. Pages 233 – 254.

- [12] Nastar, C., Mitschke, M., Meilhac, C. and Boujemaa, N., *Surfimage: a Flexible Content-Based Image Retrieval System.*, ACM Multimedia'98, Bristol, UK, 1998. Pages 339 – 344.
- [13] Hua, K. A., Khanh, V. and Oh, J. H., *SamMatch: A Flexible and Efficient Sampling-Based Image Retrieval Technique for Large Image Databases.*, Proceedings of the seventh ACM International conference on Multimedia, 1999. Pages 225 – 234.
- [14] Rui, Y. and Huang, T. S., *A Novel Relevance Feedback Technique in Image Retrieval.*, ACM Multimedia, Orlando, FL, USA, October, 1999. Pages 67 – 70.
- [15] Jose, J. M., Furner, J. and Harper, D. J., *Spatial querying for image retrieval: a user-oriented evaluation.*, Proceedings of the 21st ACM SIGIR Conference on Research and development in information retrieval, Melbourne, Australia, 1998. Pages 232 – 240.
- [16] Goh, S. and Tan, K., *MOSAIC: A fast multi-feature image retrieval system.*, Data & Knowledge Engineering, 33 (3), Jun, 2000. Pages 219 – 239.
- [17] Liang, K. and Kuo, C., *WaveGuide: A Joint Wavelet-Based Image Representation and Description System.*, IEEE Transactions on Image Processing, Vol. 8 No. 11, November, 1999. Pages 1619 – 1629.
- [18] Kobayashi, H., Okouchi, Y. and Ota, S., *Image Retrieval System Using KANSEI Features.*, 5th Pacific Rim International Conference on Artificial Intelligence (PRICAI'98), Singapore, November, 1998. Pages 626 – 635.
- [19] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A. and Jain, R., *Content-Based Image Retrieval at the End of the Early Years.*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, December, 2000. Pages 1349 – 1380.
- [20] Müller, H., Müller, W., Squire, D. M., Marchand-Maillet, S. and Pun, T., *Performance evaluation in content-based image retrieval: overview and proposals.*, Pattern Recognition Letters, vol. 22, 2001. Pages 593 – 601.
- [21] Zhu, B., Ramsey, M. and Chen, H., *Creating a Large-Scale Content-Based Airphoto Image Digital Library.*, IEEE Transactions on Image Processing, vol. 9, No 1, January, 2000. Pages 163 – 167.
- [22] Samet, H. and Soffer, A., *MARCO: MAp Retrieval by COntent.*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, No 8, August, 1996. Pages 783 – 798.
- [23] Faloutsos, C. and Lin, K.-I., *FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets.*, ACM SIGMOD, San Jose, CA, USA, 1995. Pages 163 – 174.
- [24] Orphanoudakis, S. C., Chronaki, C. and Kostomanolakis, S., *I²C: A System for the Indexing, Storage, and Retrieval of Medical Images by Content.*, Medical Informatics, vol. 19(2), 1994. Pages 109 – 122.

- [25] Chang, S.-K., Shi, Q.-Y. and Yan, C.-W., *Iconic Indexing by 2D Strings.*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 9, 1987. Pages 413 – 428.
- [26] Araújo, A. A., and Guimarães, S. J. F., *Recuperação de informação visual com base no conteúdo em imagens e vídeos digitais.*, RITA – Revista de Informática Teórica e Aplicada, vol. 7, no 2, Edição Especial em Computação Gráfica e Processamento de Imagens, Instituto de Informática, UFRGS, Porto Alegre – RS – Brasil, December, 2000. Pages 43 – 71.
- [27] Belongie, A. P., and Shapiro, L. G., *Color and texture-based image segmentation using EM and its application to content-based image retrieval.*, In Proceedings of the Sixth International Conference on Computer Vision (ICCV'98), Bombay, India, January, 1998.
- [28] Dy, J. G., Brodley, C. E., Kak, A., Shyu, C.-R. and Broderick, L. S., *The customized-queries approach to CBIR using EM.*, In Proceedings 1999 IEEE Conf. Computer Vision and Pattern Recognition (CVPR'99), Fort Collins, CO, USA, 1999. Pages 400 – 406.
- [29] Vendrig, J., Worring, M., and Smeulders, A. W. M., *Filter image browsing: Exploiting interaction in image retrieval.*, In Third International Conference On Visual Information Systems (VISUAL'99), no 1614 in Lecture Notes in Computer Science. Springer-Verlag, Amsterdam, The Netherlands, 1999. Pages 147 – 154.
- [30] Berman, A. P., and Shapiro, L. G., *Efficient content-based retrieval: Experimental results.*, In IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'99), Fort Collins, CO, USA, June, 1999. Pages 55 – 61.
- [31] Smith, J. R., and Chang, S.-F., *VisualSEEK: a fully automated content-based image query system.*, ACM Multimedia, Boston, MA, USA, 1996. Pages 87 – 98.
- [32] Cox, I. J., Miller, M. L., Minka, T. P., Papathomas, T. V. and Yanilos, P. N., *The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments.*, IEEE Transactions on Image Processing, vol. 9, no 1, January, 2000. Pages 20 – 37.
- [33] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D. and Yanker, P., *Query by Image and Video Content: The QBIC System.*, IEEE Computer, Vol. 28 No. 9, September, 1995. Pages 23 – 32.
- [34] Ma, W.-Y., and Manjunath, B. S., *NeTra: A toolbox for navigating large image databases.*, Multimedia Systems, Vol. 7 No. 3, May, 1999. Pages 184 – 198.
- [35] Das, M., Riseman, E. M. and Draper, B. A., *FOCUS: Searching for Multi-colored Objects in a Diverse Image Database.*, In Proc. of IEEE CVPR'97, June, 1997. Pages 756 – 761.
- [36] Assfalg, J. and Pala, P., *Querying by Photographs: A VR Metaphor for Image Retrieval.*, IEEE Multimedia Computing and Systems, January – March, 2000. Pages 52 – 59.

- [37] Ortega, M., Rui, Y., Chakrabarti, K., Porkaew, K., Mehrotra, S. and Huang, T. S., *Supporting Ranked Boolean Similarity Queries in MARS.*, IEEE Transactions on Knowledge and Data Engineering, Vol. 10, No 6, November – December, 1998. Pages 905 – 924.
- [38] Chang, T., and Kuo, C.-C., *Texture analysis and classification with tree-structured wavelet transform.*, IEEE Transactions on Image Processing, Vol. 3, No 4, October, 1993. Pages 429 – 441.
- [39] Laine, A. F., and Fan, J., *Texture classification by wavelet packet signatures.*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No 11, November, 1993. Pages 1186 – 1191.
- [40] Direne, A. I., *Designing Intelligent Systems for Teaching Visual Concepts.*, International Journal of Artificial Intelligence in Education, Vol. 8, 1997. Pages 44 – 70.
- [41] Gudivada, V. N. and Raghavan, V. V., *Design and Evaluation of Algorithms for Image Retrieval by Spatial Similarity.*, ACM Transactions on Information Systems, vol. 13, No 2, April, 1995. Pages 115 – 144.
- [42] Berman, A. P. and Shapiro, L. G., *A Flexible Image Database System for Content-Based Retrieval.*, Computer Vision and Image Understanding, vol. 75, Nos 1/2, July/August, 1999. Pages 175 – 195.
- [43] Chang, S. K., Yan, C. W., Dimitroff, D. C. and Arndt, T., *An Intelligent Image Database System.*, IEEE Transaction on Software Engineering, vol. 14, No 5, May, 1988. Pages 681 – 688.
- [44] Nabil, M., Ngu, A. and Sheperd, J., *Picture Similarity Retrieval Using the 2D Projection Interval Representation.*, IEEE Transaction on Knowledge and Data Engineering, vol. 8, No 4, August, 1996. Pages 533 – 539.
- [45] Huang, C.-L. and Huang, D.-H., *A content-based image retrieval system.*, Image and Vision Computing, vol. 16, December, 1998. Pages 149 – 163.
- [46] Nappi, N., Polese, G. and Tortora, G., *FIRST: Fractal Indexing and Retrieval SysTem for Image Databases.*, Image and Vision Computing, vol. 16, December, 1998. Pages 1019 – 1031.
- [47] Ratha, N. K., Karu, K., Chen, S. and Jain, A. K., *A Real-Time Matching System for Large Fingerprint Databases.*, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 18, No 8, August, 1996. Pages 799 – 813.
- [48] Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J., *Active Shape Models – Their Training and Application.*, Computer Vision and Image Understanding, vol. 61, No 1, January, 1995. Pages 38 – 59.

[49] Marchaukoski, J. N., *Princípios para a Construção de Banco de Dados de Imagens Médicas.*, Departamento de Informática, Universidade Federal do Paraná (UFPR), Msc Dissertação, Novembro, 2001.