Recuperação de Imagens: Desafios e Novos Rumos

Ricardo da S. Torres[1], **Javier A. M. Zegarra**[1], **Jefersson A. dos Santos**[1], **Cristiano D. Ferreira**[1], **Otávio A. B. Penatti**[1], **Fernanda Andaló**[1], **Jurandy Almeida**[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP) – Campinas, SP

{rtorres,feandalo,jurandy.almeida}@ic.unicamp.br

{jmontoyaz,jeferssonalex,crferreira,otaviopenatti}@gmail.com

Abstract. *Huge image collections have been created, managed and stored into image databases. Given the large size of these collections it is essential to provide efficient and effective mechanisms to retrieve images. This is the objective of the so-called* content-based image retrieval *– CBIR – systems. Traditionally, these systems are based on objective criteria to represent and compare images. However, users of CBIR systems tend to use subjective elements to compare images. The use of these elements have improved the effectiveness of content-based image retrieval systems. This paper discusses approaches that incorporate semantic information into content-based image retrieval process, highlighting some new challenges on this area.*

Resumo. *Atualmente, um grande conjunto de imagens digitais vem sendo gerado, manipulado e armazenado em bancos de imagens. Dado o tamanho desses acervos, prover meios de recuperar imagens de forma eficiente e eficaz é essencial. Esse é o objetivo dos Sistemas de* Recuperação de Imagens por Conteúdo. *Tradicionalmente, tais sistemas são baseados em critérios objetivos de baixo nível para representar e comparar imagens. Porém, usuários destes sistemas tendem a utilizar elementos subjetivos para comparar imagens. Considerar esses elementos tem proporcionado uma melhora, em termos de eficácia, no processo de recuperação de imagens. Este artigo discute propostas relacionadas à incorporação de informação semântica à recuperação de imagens por conteúdo, destacando novos desafios desse campo.*

1. Introdução

Este artigo relaciona-se com temas associados ao desafio *"Gestão da Informação em grandes volumes de dados multimídia distribuídos"* [Carvalho et al. 2006].

A criação de novos dispositivos para obtenção e armazenamento de imagens tem possibilitado a disseminação de grandes coleções de imagens nas mais diversas aplicações: medicina, sensoriamento remoto, sistemas de informação de biodiversidade, biometria, dentre outras. As soluções existentes para o gerenciamento integrado de imagens e dados específicos de cada aplicação baseiam-se na criação de infra-estruturas de armazenamento e no desenvolvimento de ferramentas para realização de consultas.

Por um lado, sistemas e ferramentas específicas têm sido propostos para a busca de imagens levando-se em conta suas descrições textuais associadas. Entretanto, o processo de anotação de imagens (atribuição de descritores textuais) costuma ser ineficiente,

pois é comum que os usuários não façam anotações de forma sistemática, por exemplo, preocupando-se em utilizar palavras semelhantes para uma mesma característica da imagem. Além disso, esse processo normalmente é pouco eficaz, uma vez que a interpretação do conteúdo visual de uma imagem varia de acordo com o conhecimento, o objetivo, a experiência e a percepção de cada usuário.

Por outro lado, várias pesquisas vêm sendo realizadas visando a especificação de algoritmos que manipulam o conteúdo das imagens (objetos e suas propriedades como forma, cor e textura). Estes algoritmos visam descrever o conteúdo das imagens de maneira que a descrição possa ser utilizada para indexá-las e manipulá-las em um banco de imagens.

Nos sistemas de *recuperação de imagens por conteúdo* (Content-based Image Retrieval – CBIR), o cálculo da similaridade entre duas imagens é realizado por meio de descritores. Um descritor pode ser caracterizado por: (i) um *algoritmo de extração de características*, baseado em técnicas de processamento de imagens, que codifica as propriedades da imagem em um *vetor de características*; e (ii) uma *medida de similaridade* (função de distância) que computa a similaridade entre duas imagens como uma função de distância entre seus vetores de características correspondentes. No domínio de recuperação de imagens por conteúdo, um descritor é considerado melhor que outro se sua utilização resulta em um número maior de imagens relevantes retornadas para uma dada consulta.

A criação de sistemas de recuperação de imagens combina pesquisa em bancos de dados e processamento de imagens, apresentando diversos desafios tanto do ponto de vista teórico quanto de implementação. Os problemas enfrentados vão desde questões de armazenamento até interfaces amigáveis. O elemento complicador reside no fato de que os objetos envolvidos (imagens) têm gerenciamento muito mais complexo do que os objetos textuais. Do ponto de vista de armazenamento, imagens ocupam muito espaço e a sua recuperação costuma ser dependente de aplicação e contexto. Além disso, o processo de recuperação de imagens deixa de ser uma questão de processamento de *strings* e passa a depender de outras características, inclusive de diferentes aspectos cognitivos relativos à interpretação visual. De fato, a escolha de imagens de consulta que melhor representam o que o usuário deseja buscar em um sistema de recuperação de imagens por conteúdo é uma tarefa difícil, uma vez que o usuário precisa traduzir, por meio de descritores, conceitos (idéias) em características de baixo nível (propriedades de *pixels*).

Este artigo discute os principais desafios de pesquisa relacionados às soluções de alguns desses problemas. Trata-se de desafios multi e interdisciplinares envolvendo pesquisa em áreas como Processamento de Imagens, Bancos de Dados, Interface Humano-Computador, Estatística, Psicologia, etc. Em especial, serão apresentados questões relacionadas à incorporação de aspectos semânticos no processo de recuperação de imagens.

2. Visão Geral de Sistemas de Recuperação de Imagens por Conteúdo

Um sistema de recuperação de imagens por conteúdo é centrado na noção de similaridade de imagens — dado um banco com um grande número de imagens, o usuário deseja recuperar as imagens mais similares a um padrão de consulta (normalmente uma imagem definida como exemplo). O processo de recuperação é baseado na comparação das imagens por meio de descritores [Torres and Falcão 2006].

A Figura 1 mostra a arquitetura típica de um sistema de recuperação de imagens por conteúdo [Torres and Falcão 2006]. Essa arquitetura possui duas funcionalidades principais: a inserção de dados e o processamento de consultas.

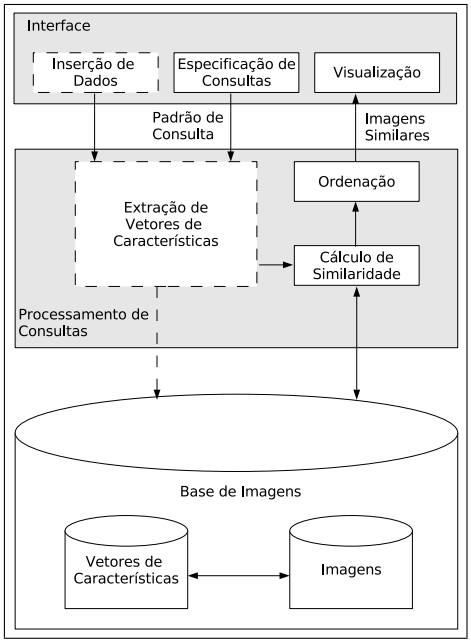

Figura 1. Arquitetura típica de um sistema de recuperação de imagens por conteúdo [Torres and Falcão 2006].

O subsistema de inserção de dados, representado por módulos e setas tracejados, é responsável por extrair os vetores de características das imagens e armazená-los na base de imagens. Geralmente, esse processo é realizado uma única vez para cada imagem e para cada descritor, sendo utilizado de maneira *offline*. Os vetores armazenados são usados posteriormente no processamento de consultas.

O processamento de consultas é organizado da seguinte forma: a interface permite ao usuário especificar uma consulta por meio de um padrão de consulta (por exemplo, uma imagem – *query by visual example*) e visualizar as imagens recuperadas. O módulo de processamento de consultas extrai o vetor de características do padrão de consulta e aplica uma métrica de distância, como a distância Euclidiana, para avaliar a similaridade entre a imagem de consulta e as imagens da base. Em seguida, esse módulo ordena as imagens da base de acordo com a similaridade e retorna as mais similares para o módulo de interface. Esse processo pode ser otimizado pela utilização de estruturas de indexação, como a *M-Tree* [Ciaccia et al. 1997].

3. Representação de Imagens a partir de Características de Baixo Nível

Para que sistemas de recuperação de imagens por conteúdo sejam viáveis, é desejável que as imagens possam ser descritas pelas suas propriedades intrínsecas, tais como forma, textura e cor, normalmente representadas em vetores de características. Neste sentido, os descritores de imagens são utilizados para extrair e comparar esses vetores, viabilizando a indexação e busca de imagens.

Formalmente, um vetor de características $\overrightarrow{v}_{\hat{I}}$ de uma imagem $\hat{I}$ pode ser considerado como um ponto no espaço $\mathbb{R}^n$: $\overrightarrow{v}_{\hat{I}} = (v_1, v_2, \ldots, v_n)$, onde n é a dimensão do vetor [Torres and Falcão 2006].

Por outro lado, um descritor de imagem D é definido como uma tupla (ϵ_D, δ_D) [Torres and Falcão 2006], onde $\epsilon_D : \{\hat{I}\} \rightarrow \mathbb{R}^n$ é uma função que extrai o vetor de características $\overrightarrow{v}_{\hat{I}}$ da imagem $\hat{I}$; e $\delta_D : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função que calcula a similaridade entre duas imagens, inversamente proporcional à distância entre seus vetores de características.

Note que esta definição de descritor considera que o desempenho e a eficácia de um descritor depende não somente do algoritmo de extração de vetor de característica empregado, mas também da função de similaridade usada. Trabalhos como [Sebastian et al. 2002, Bugatti et al. 2008] avaliam o uso de diferentes funções de distância.

Os descritores de imagens devem possuir propriedades que garantam sua eficácia quando usados na indexação e recuperação de imagens. A mais importante é a caracterização única de um determinado objeto a partir do vetor de características, facilitando a distinção entre imagens visualmente diferentes, de acordo com alguma métrica. A segunda propriedade mais importante é a invariância do descritor a algumas classes de transformações, como a rotação e a translação. Outras propriedades desejáveis são: insensibilidade a ruído, geração de vetores de características compactos (que requerem pouco espaço de armazenamento) e função de extração computacionalmente eficiente.

Nesta seção serão apresentadas as quatro características primárias mais importantes presentes nas imagens. Para cada uma destas serão apresentados os principais desafios a serem tratados.

3.1. Descritores de Cor

A cor é uma propriedade visual das imagens que geralmente simplifica a identificação de objetos e a sua posterior extração das cenas. Isso se deve ao fato de que seres humanos podem discernir milhares de tons e intensidades de cores, enquanto apenas duas dúzias de tons de cinza [Gonzalez and Woods 2001].

A Figura 2 mostra a representação esquemática de uma imagem sendo armazenada em um sistema que faz uso da informação de cor para descrever, representar, comparar e recuperar imagens. Após uma imagem ser fornecida como entrada, o seu conteúdo visual é analisado e resumido em um espaço de cores pré-estabelecido. Em seguida, são extraídas características visuais a partir das informações de cor. Por fim, representações compactas são escolhidas para as informações analisadas durante a etapa anterior. Essas representações determinam vetores de características que são armazenados e indexados em um banco de imagens.

De acordo com o esquema descrito acima, pode-se considerar a existência de três tópicos chave que precisam ser explorados para que se realize um processo automático de recuperação de imagens por cor: (1) qual espaço de cor deve ser utilizado para descrever, analisar e comparar imagens; (2) como descrever imagens por meio da sua distribuição de cores; e (3) como representar o conteúdo de uma imagem (características visuais) em um banco de imagens. A seguir cada um destes três desafios serão brevemente discutidos.

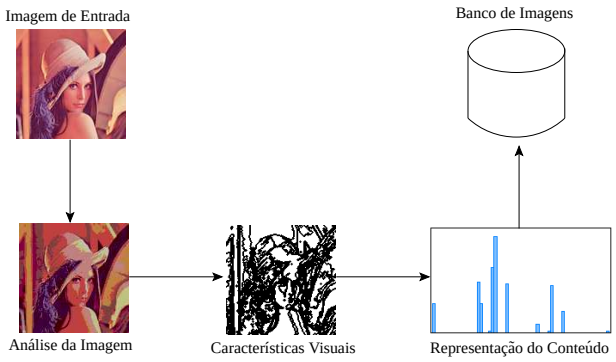

Figura 2. Representação esquemática de uma imagem sendo armazenada em um banco de dados de imagem.

A cor de um *pixel* é, em geral, representada por três valores, um para cada canal de um determinado espaço de cor. Um espaço de cor é uma especificação de um sistema de coordenadas 3D e um subespaço dentro desse sistema, no qual cada cor é representada por um único ponto [Gonzalez and Woods 2001]. A escolha de um espaço de cor no qual as imagens serão representadas, analisadas e comparadas é o primeiro passo em qualquer sistema de recuperação de imagens por cor. Outros desafios em aberto estão relacionados à complexidade dos algoritmos de análise das imagens, à utilização de espaço em disco para representar suas características visuais, à complexidade da função de distância utilizada para comparar as características visuais extraídas e, finalmente, à eficácia do processo de recuperação das imagens.

3.2. Descritores de Textura

Uma importante propriedade visual presente nas imagens e utilizada tanto na percepção humana, bem como no reconhecimento de padrões, é a textura. Embora a percepção de textura seja intuitiva (pois é possível reconhecê-la ao enxergá-la), ainda não existe uma definição precisa sobre a mesma. Contudo, os descritores de textura podem medir algumas das suas propriedades como regularidade, suavidade, granularidade, entre outras [Tamura et al. 1978].

A análise de textura é uma área que vem despertando grande interesse há alguns anos por diversas empresas e grupos de pesquisa espalhados pelo mundo. Como resultado desses estudos, diversas aplicações foram propostas, incluindo desde tarefas de recuperação de imagens por conteúdo [Liu et al. 2007] até tarefas de classificação, segmentação, síntese e obtenção de forma a partir de textura.

Estas aplicações caracterizam-se por um elemento em comum. Todas procuram responder a pergunta de como caracterizar/representar uma textura de forma que as suas propriedades mais importantes sejam extraídas. Apesar dos avanços apresentados na literatura, os métodos propostos geralmente foram avaliados em ambientes controlados que muitas vezes não representam situações reais onde o usuário interage com o sistema. Neste sentido, os sistemas de recuperação de imagens por conteúdo baseados em textura apresentam o desafio de obter representações compactas em ambientes *não-controlados* geralmente caracterizados por imagens de texturas com variações inter-classes e distorções geométricas. As variações inter-classes estão relacionadas com o fato de que as texturas de diferentes classes podem ser bem semelhantes. Este aspecto é

ilustrado na Figura 3.

Figura 3. Exemplos de inter-classes entre imagens de textura. Apesar da sua similaridade visual, cada uma das imagens apresentadas correspondem a classes diferentes.

Por outro lado, as distorções presentes nas imagens de textura referem-se tanto a orientações bem como escalas, podendo gerar novas caracterizações de texturas diferentes das imagens originais. A Figura 4 apresenta alguns exemplos destes tipos de distorções.

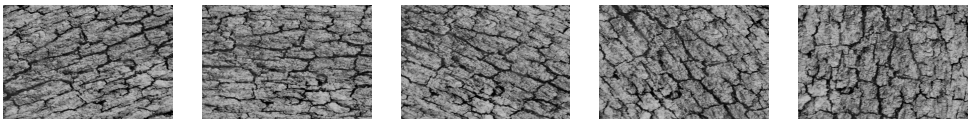

Figura 4. Exemplos de imagens de textura nas quais existem rotações.

3.3. Descritores de Forma

A forma é uma propriedade relevante, pois possibilita capturar as estruturas topológicas dos objetos presentes nas imagens. As características da forma são mais complexas de serem representadas do que a cor e a textura. Apesar disso, a forma é considerada como o arquétipo dos objetos que pertencem a uma mesma classe e, sendo assim, é muito importante na recuperação de imagens por conteúdo. Neste contexto, espera-se que as formas dos objetos sejam invariantes à escala, à rotação e à translação.

Para a aplicação de descritores de forma, muitas vezes as imagens devem passar previamente por processos de segmentação e representação. Entende-se por segmentação o processo de extração, ou distinção, de objetos relevantes em uma imagem. Como diferentes objetos requerem diferentes algoritmos de segmentação, torna-se complexa a tarefa de descrição de imagens por meio da forma. Por este motivo, geralmente as propostas de descritores não contemplam esta fase de segmentação, já supondo que as imagens foram segmentadas em um processo anterior. Por isso, a tarefa de segmentação de objetos relevantes é um primeiro desafio que os sistemas de recuperação por conteúdo baseados na forma precisam lidar.

Embora algumas vezes os dados obtidos na segmentação sejam utilizados diretamente para geração dos vetores de características, uma prática comum é a utilização de esquemas que compactam os dados em representações que são consideradas mais úteis na computação dos vetores [Gonzalez and Woods 2001].

Após o processo de segmentação e uma possível representação, as características devem ser descritas e comparadas. Mas não existe descrição matemática capaz de captar totalmente os aspectos das formas como são visualmente captados e entendidos. Além disso, a comparação destas características representa mais um importante desafio.

Os sistemas CBIR atuais que se baseiam na forma geralmente exploram informações de contorno dos objetos e/ou regiões do interior destes objetos [Zhang and Lu 2004]. Mas as características da forma disponíveis para cada tipo

de aplicação podem diferir bastante entre si. Um outro desafio é, então, encontrar uma descrição ideal para cada tipo de aplicação, considerando informações de contorno, interior ou uma combinação dos métodos, para obter uma recuperação de imagens mais eficiente.

3.4. Descritores de Relacionamento Espacial

Os descritores de relacionamento espacial entre objetos representam informações relativas ao posicionamento dos objetos na imagem e ao posicionamento dos objetos entre si. Eles podem caracterizar relacionamentos direcionais e topológicos. Relacionamentos direcionais são do tipo: "à direita", "acima", "abaixo". Relacionamentos topológicos podem ser: "dentro de", "adjacente a", "sobreposto a". Por meio desse tipo de descritor pode-se realizar uma busca do tipo: "encontrar imagens que contenham um homem embaixo de uma árvore". Assim seria levado em conta o posicionamento relativo dos objetos na imagem. Diversas aplicações podem se beneficiar das informações de relacionamento espacial entre os objetos, como aplicações geográficas e aplicações médicas.

Descritores deste tipo podem analisar uma imagem de diferentes maneiras. Alguns deles fazem representações simbólicas dos objetos da imagem e analisam seus relacionamentos espaciais por meio de *strings* [Chang et al. 1987]. Outros analisam cada ponto do contorno dos objetos [Penatti and Torres 2007] ou criam elementos envolventes para os objetos [Chang and Kim 2001] antes de extrair as relações espaciais.

O grande desafio para este tipo de descritor está envolvido com a segmentação dos objetos da imagem. A grande maioria deles necessita que os objetos já estejam segmentados na imagem. A segmentação de objetos é uma tarefa que depende do ajuste de parâmetros e, por isso, varia muito de aplicação para aplicação. Dessa maneira, o uso desses descritores fica limitado a aplicações específicas. Em aplicações de propósito geral, como a Web, por exemplo, o uso deles acaba sendo bastante restrito. Outros desafios para esta classe de descritores seriam: a capacidade de lidar com diferentes escalas dos objetos na imagem; a capacidade de lidar com oclusão parcial dos objetos; e a capacidade de identificar relacionamentos espaciais quando há um número grande de objetos na imagem.

4. Incorporação de Aspectos Semânticos no processo de Recuperação de Imagens

O processo de recuperação de imagens depende também de diferentes aspectos cognitivos relativos à interpretação visual. É difícil para um usuário escolher imagens de entrada que sejam representativas do que deseja buscar, uma vez que esta escolha se baseia na tradução de idéias em características de baixo nível (propriedades de *pixels*). Este problema é conhecido como *semantic gap* – descontinuidade semântica.

Dentre os principais desafios de pesquisa relacionados às soluções ao problema do *semantic gap*, destacam-se: a busca de técnicas para combinação de descritores (Seção 4.1 e Seção 4.2) e mecanismos para inclusão da percepção visual de cada usuário no processo de recuperação de imagens a partir de técnicas de realimentação de relevância –*relevance feedback* (Seção 4.3).

4.1. Fusão de Descritores

Em muitas aplicações, deseja-se recuperar uma imagem em função de múltiplas propriedades e/ou de uma mesma propriedade caracterizada por diferentes descritores. Neste cenário, um desafio de pesquisa relevante consiste em descobrir mecanismos para explorar as "vantagens" combinadas de diferentes descritores em uma aplicação específica.

Vários descritores vêm sendo propostos na literatura. Por um lado, diferentes descritores podem ser usados na caracterização de diferentes propriedades de uma imagem, apresentando, em geral, um bom desempenho em aplicações específicas. Por exemplo, descritores de forma e de cor podem ser utilizados em sistemas de recuperação que auxiliam o processo de identificação de espécies em um Sistema de Informação de Biodiversidade. Por outro lado, diferentes descritores podem ser utilizados para caracterizar uma mesma propriedade da imagem. O histograma de cor [Swain and Ballard 1991] e o CSIR [Almeida et al. 2008] são exemplos de descritores que costumam ser usados para caracterizar uma mesma propriedade. No entanto, não existem descritores que apresentam alto desempenho (eficácia e eficiência) em todas as aplicações.

4.2. Fusão de Evidências Textuais e Visuais

Uma outra variante para a recuperação de imagens é baseada em anotações textuais. Nesse tipo de abordagem, palavras-chaves são manualmente associadas às imagens da base. As consultas são realizadas de forma textual, por meio da comparação dos termos da consulta com as palavras-chaves atribuídas às imagens. Essa vertente possui dois problemas principais: o alto tempo necessário para a anotação manual das imagens e as possíveis inconsistências nas anotações. Porém, essa abordagem introduz a informação semântica ao processo de recuperação de imagens, sendo assim, complementar à recuperação de imagens por conteúdo. Dessa forma, uma nova vertente para recuperação de imagens consiste no desenvolvimento de um modelo que funda anotações textuais e o conteúdo das imagens, aproveitando assim, as qualidades complementares de ambas as abordagens.

4.3. Realimentação de Relevância

Uma das formas mais recorrentes de se realizar uma consulta em um sistema de recuperação de imagens por conteúdo é através da definição de um padrão de consulta fornecido pelo usuário. Nessa técnica, o sistema deve retornar as imagens mais similares a esse padrão de consulta, segundo os descritores utilizados. Porém, o conceito de similaridade é subjetivo. Por isso, é necessário prover um meio para que o usuário possa exprimir qual a sua necessidade em uma dada consulta. Uma forma de realizar essa tarefa é a partir do ajuste de determinados parâmetros do sistema de busca, como, por exemplo, a atribuição de pesos para cada descritor. Entretanto, para a realização dessa tarefa, seria necessário que o usuário conhecesse detalhes do processo de descrição de imagens, por exemplo o valor semântico relacionado aos elementos de vetores de características.

Uma das técnicas mais utilizadas para solucionar esse problema é chamada de *relevance feedback* – realimentação de relevância. *Relevance feedback* é uma técnica inicialmente utilizada na recuperação de informações por texto [Xu et al. 2003], mas que atualmente é alvo de pesquisa na área de recuperação de imagem por conteúdo em todo mundo [Kim et al. 2005, Rui et al. 1998, Zhou and Huang 2003]. Esse mecanismo tem por objetivo possibilitar que o usuário expresse a sua necessidade na especificação de

uma consulta, sem recorrer a propriedades de baixo nível utilizadas na representação de imagens. Para isso, o usuário apenas precisa indicar as imagens relevantes, e em certos casos também as irrelevantes, dentre um conjunto retornado pelo sistema. A cada iteração, o algoritmo busca "aprender" quais propriedades visuais melhor definem as imagens relevantes, a partir das informações fornecidas pelo usuário, ou seja, as imagens por ele indicadas. Dessa forma, a consulta é reformulada automaticamente e realizada novamente. E assim, após um determinado número de iterações, o sistema retorna as imagens mais similares à imagem de consulta.

Com isso, *relevance feedback* endereça duas questões referentes ao processo de recuperação de imagens por conteúdo. A primeira delas reside na diferença entre as propriedades visuais de alto nível, através das quais o usuário tem a percepção semântica da informação visual, e a descrição de baixo nível utilizada para a representação das imagens. A outra diz respeito ao caráter subjetivo da percepção da imagem pelo usuário. Diferentes pessoas, ou a mesma em diferentes circunstâncias, podem ter percepções visuais distintas de uma mesma imagem. Com *relevance feedback* essas duas questões são contornadas de forma transparente para o usuário.

Uma das etapas importantes do processo de *relevance feedback* em sistemas de recuperação de imagens por conteúdo consiste na (re-)combinação automática de descritores, considerando a indicação de imagens relevantes feita pelo usuário. Novamente, um grande desafio de pesquisa consiste na definição de mecanismos "inteligentes" para combinação de descritores levando-se em conta as propriedades visuais que cada um descreve, bem como a percepção visual de cada usuário do sistema.

5. Aplicações

Várias aplicações podem se beneficiar da incorporação de aspectos semânticos no processo de recuperação de imagens. A seguir, algumas destas aplicações são descritas.

5.1. Recuperação de Imagens na Web

Um aumento na quantidade de imagens vem ocorrendo ultimamente e pode ser observado principalmente na Web. Segundo [Lin et al. 2003] o Google indexava, em 2003, 425 milhões de imagens e recebia 150 milhões de consultas por dia em seu buscador de imagens. O número expressivo de imagens na Web aumenta a medida que a Internet e os dispositivos de captura de imagens se popularizam. Além disso, trata-se de coleção bastante heterogênea, compreendendo imagens de diferentes tipos, formatos e tamanhos. Essas peculiariedades tornam o desenvolvimento de um sistema de recuperação visual na Web um desafio para a computação.

Atualmente, os sistemas de busca de imagens mais comuns baseiam-se em descrições textuais, ou seja, as imagens são analisadas de acordo com anotações textuais realizadas pelas pessoas que as disponibilizaram. Um sistema deste tipo recebe como consulta um conjunto de palavras-chaves e, a partir delas, busca por imagens que tenham palavras-chaves relacionadas. Apesar dessa classe de sistemas mostrar-se eficaz em muitos casos, a subjetividade da anotação textual de imagens tende a fazer com que muita informação relevante não seja encontrada. Usuários diferentes tendem a ter diferentes interpretações de uma mesma imagem. Devido a isso, estes sistemas apresentam deficiências.

Uma outra abordagem mais recente, consiste em considerar as características visuais da imagem como elementos de análise. Os principais processos envolvidos numa busca seriam a extração de características das imagens e a comparação das características extraídas. A Web é uma aplicação que exige tempo de resposta pequeno. No entanto, tanto a extração de características e quanto a comparação delas pode ser lenta.

As duas abordagens citadas possuem suas vantagens e desvantagens. A abordagem textual tem a vantagem de analisar a descrição textual da imagem, que, na maioria das vezes, representa a semântica da imagem. Devido a isso, uma informação de mais alto nível é utilizada, sendo mais próxima da linguagem do usuário. Já a abordagem por conteúdo, embora permita um esquema de recuperação automático, usa características de baixo nível, que, de certa maneira, são mais distantes da semântica da imagem.

Apesar dos prós e contras já citados, ambas as abordagens sofrem de um problema em comum: compreender os desejos do usuário. No caso da descrição textual, a interpretação de uma imagem é muito subjetiva o que produz descrições textuais diferentes para possíveis imagens semelhantes no conceito de um determinado usuário. Isso faz com que os resultados de uma busca sejam muitas vezes diferentes do que o usuário procura. No caso da busca por conteúdo, não há associação de características visuais à semântica da imagem, portanto, nem sempre encontrar um padrão visual de acordo com o desejo do usuário equivale a encontrar o tipo de imagem que ele está procurando.

Diversas técnicas para ajudar na compreensão da intenção do usuário já foram propostas. Dentre elas, cabe citar o uso de técnicas de Relevance Feedback (ver Seção 4.3).

Outra possível solução para estes problemas seria o uso da busca textual em conjunto com a busca por conteúdo. Desta forma, componentes semânticos de alto nível seriam agregados a componentes visuais das imagens. O usuário poderia, por exemplo, iniciar uma consulta de maneira textual, e, assim que encontrasse uma imagem de resultado bastante relevante, poderia solicitar uma busca por conteúdo baseada naquela imagem.

Além das dificuldades apresentadas, sistemas de busca de imagens devem se preocupar também com elementos internos à sua arquitetura [Kherfi et al. 2004]. Os principais módulos de um sistema de recuperação de imagens na Web seriam a coleta de dados, a extração de informações das imagens, a geração de índices, a especificação das consultas e a recuperação e refinamento dos resultados (ver Figura 5).

5.2. Aplicações Agrícolas: Recuperação de Imagens de Sensoriamento Remoto

Uma imagem de sensoriamento remoto (ISR) é uma coleção de dados detectados por um sensor através das propriedades espectrais dos alvos que estão sendo imageados sem que haja contato com eles por meio da detecção do fluxo de energia radiante[Câmara et al. 1997]. Após o processo de detecção, os dados adquiridos pelos sensores são convertidos em formato digital. A especialização desses dados é constituída por *pixels*, que distribuídos em linhas e colunas formam uma imagem

Um exemplo de ISR é a Figura 6 onde foi detectado o fluxo de energia radiante de uma região do município de Monte Santo (sul do Estado de Minas Gerais), região tradicional no cultivo de café.

O uso de ISRs para mapeamento e extração de dados do mundo real é bas-

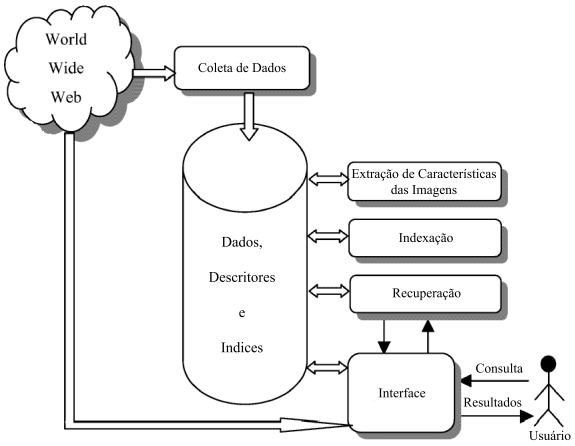

Figura 5. Estrutura geral de um sistema de busca de imagens na Web e seus principais módulos (adaptado de [Kherfi et al. 2004]).

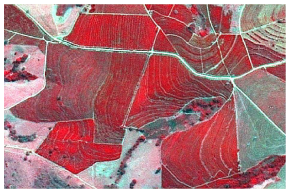

Figura 6. ISR extraída de uma região de café no município de Monte Santo (MG).

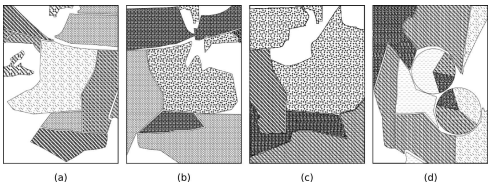

Figura 7. Exemplo de banco de dados contendo ISRs.

tante comum em Sistemas de Informação Geográfica (SIGs). Nessa classe de imagens, o problema da recuperação por conteúdo tem sido uma área de muito interesse em diferentes domínios, como na agricultura e no planejamento urbano. Del Val Cura [del Val Cura 2002] propôs um modelo para recuperação por conteúdo de Imagens de Sensoriamento Remoto utilizando três recursos principais: padrões de textura e cor como elemento básico de consulta, uso de múltiplos modelos matemáticos de representação e caracterização do conteúdo e um mecanismo de *relevance feedback* para o processo de consulta.

Além da recuperação de ISRs, outro problema importante diz respeito à vetorização de regiões de interesse. Em algumas aplicações de SIGs, frequentemente é necessário representar o mundo por meio de objetos identificáveis com geometria própria, ou seja, uma representação vetorial. Desse modo, para que sejam utilizadas por tais aplicações é preciso converter as ISRs do seu formato matricial para formato vetorial.

Uma atividade comum em aplicações que usam ISRs é identificar e criar um mapa diferenciando regiões de acordo com o tipo de vegetação cultivada em cada uma delas. A identificação e vetorização manual de regiões (por meio de uma ferramenta de edição gráfica) em ISRs é um processo eficaz, porém demanda muito tempo para ser executado, além da subjetividade imposta por cada usuário. Métodos automáticos existem, mas normalmente utilizando apenas características de baixo nível (cor e textura geralmente) não apresentam resultados satisfatórios. Viabilizar um meio de interação com o usuário pode

permitir o aprimoramento do reconhecimento de regiões. Desse modo, as técnicas de *relevance feedback* podem ser úteis para considerar a informação semântica fornecida pelo usuário.

Uma proposta que está sendo desenvolvida é a criação de um sistema que realize o reconhecimento semi-automático e vetorização das imagens. A seguir é apresentado um exemplo de uso e funcionamento do sistema. A Figura 7 ilustra um banco de dados com imagens de sensoriamento remoto de culturas agrícolas variadas.

Suponha que o usuário do sistema realize uma consulta "por regiões de imagem onde esteja sendo cultivado café". A Figura 8 mostra as imagens retornadas com destaque para as regiões reconhecidas como plantações de café utilizando características referentes à textura das regiões. O usuário seleciona então, as regiões reconhecidas que realmente correspondem às plantações de café e as que não foram reconhecidas, mas que entretanto, são áreas de cultivo de café. Na Figura, as regiões reconhecidas corretamente são marcadas com "V", enquanto as incorretas com "X".

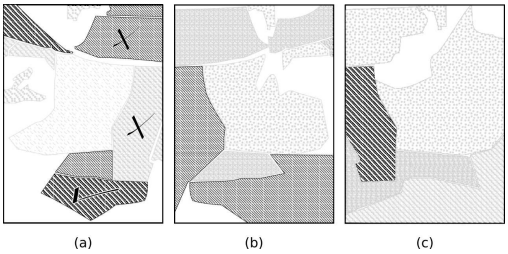

Figura 8. Exemplo de consulta por regiões com plantação de café.

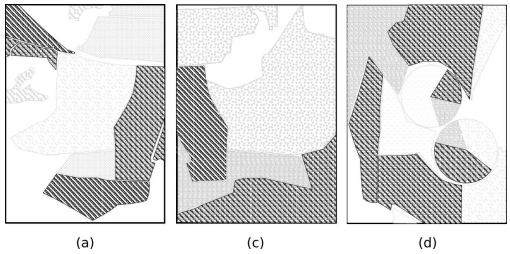

Figura 9. Consulta realizada após o *relevance feedback*.

Levando-se em conta as indicações do usuário, o sistema refaz a mesma consulta e retorna outro resultado, como mostra a Figura 9. Repare que a imagem (b), que possuía regiões reconhecidas como plantações de café na primeira consulta não aparece nos novos resultados. Em contrapartida a imagem (d), agora é retornada pois possui regiões com textura similares às que foram indicadas pelo usuário.

Se o resultado satisfizer o usuário, as regiões reconhecidas são vetorizadas e armazenadas no repositório. As imagens vetoriais podem ser utilizadas agora para diversas aplicações, como na confecção de mapas de plantações de café, por exemplo.

Em ambas aplicações citadas, considerar a opinão do usuário é bastante importante, visto que ISRs possuem propriedades bastante distintas. A distribuição de cores, texturas, forma e geometria dos objetos estão dentre as mais importantes características das ISRs utilizadas para o seu processamento. No caso das cores, o fato de que cada especialista possa definir sua própria composição de cores, e inclusive mudar essa composição dinamicamente, torna difícil armazenar descritores de cores baseados em sistemas que associam pesos diferentes às diferentes bandas do espectro visual (HSI, L*a*b, YIQ, etc). Outro problema é com relação aos modelos de descrição. A diversidade de padrões encontrados nas ISRs faz com que seja difícil encontrar um modelo universal para caracterização e modelagem de similaridade da imagem [del Val Cura 2002].

As aplicações descritas acima são exemplos de como as propriedades das ISRs afetam a sua utilização ao mesmo tempo em que ilustram como características de baixo

nível e propriedades semânticas podem ser combinadas para oferecer o resultado esperado pelo usuário.

6. Oportunidades de Pesquisas Multidisciplinares

A implementação de aplicações como as mencionadas na seção anterior requerem a realização de pesquisas multidisciplinares envolvendo várias áreas da computação e até mesmo de outras áreas de conhecimento: Processamento de Imagem, Bancos de Dados, Interface Humano-Computador, Inteligência Artificial, Estatística, Redes Sociais, etc. A seguir algumas oportunidades de pesquisas multidisciplinares são enumeradas:

1. Não existem muitas técnicas disponíveis para lidar com o distanciamento semântico (*semantic gap*) entre imagens e suas anotações. Novas ferramentas para marcar/anotar imagens (e suas regiões) e para extração automática de propriedades semânticas precisam ser desenvolvidas. Técnicas de Aprendizado de Máquina, Bancos de Dados e Processamento de Imagem precisam ser combinadas.
2. O grande volume de dados em aplicações que manipulam imagens exige pesquisa por métodos eficientes para sua organização, armazenamento e indexação nos mais variados tipos de redes. Este desafio implica, por exemplo, pesquisa em Bancos de Dados e Sistemas Distribuídos.
3. Descoberta de novas relações e mineração de padrões. Técnicas de mineração de texto poderão ser combinadas com descrições baseadas em aspectos visuais de imagens. Técnicas de Estatística e Inteligência Artificial precisam ser combinadas com técnicas de Processamento de Imagem.
4. Novas interfaces e técnicas de visualização de informação para anotação, navegação e busca dos diferentes tipos de dados (texto, imagem, vídeo, etc) precisam ser investigadas. Pesquisas nessa área exigirão a realização de estudos com usuários-alvo nos respectivos domínios de aplicação. Resultados de pesquisas em áreas como Psicologia e Interface Humano-Computador podem ser utilizados.
5. Aspectos de validação das mais diferentes técnicas desenvolvidas para CBIR ainda não despertaram muita atenção de pesquisadores da área. Verifica-se que há poucas coleções de referência (*benchmarks*) assim como medidas/critérios de comparação e avaliação comumente aceitos pela comunidade de CBIR. O uso de testes estatísticos pode ser combinado com o uso de técnicas de avaliação comumente usadas na área de Recuperação de Informação.
6. O processo de recuperação de imagens na Web pode tirar vantagem de relações sociais existentes entre usuários. Possivelmente, consultas similares feitas por usuários que compartilham os mesmos interesses poderiam retornar os mesmos resultados. Pesquisa nesta área envolve combinar técnicas extração de conhecimento de redes sociais, Bancos de Dados e Recuperação de Informação.

Referências importantes que apresentam uma visão geral da área de CBIR e discutem desafios de pesquisa incluem [Yan and Hauptmann 2007, Smeulders et al. 2000, Liu et al. 2007, Torres and Falcão 2006, Datta et al. 2008, Antani et al. 2002].

7. Conclusões

Este artigo discutiu alguns desafios de pesquisa relacionados ao gerenciamento e recuperação de imagens em grandes repositórios. Em especial, foram descritos conceitos

relacionados à busca de imagens a partir de características visuais, enfatizando limitações de soluções existentes na área. Além disso, foram apresentados novos desafios de pesquisa relacionados à incorporação de aspectos semânticos no processo de recuperação de imagens. Trata-se de processo multi e interdisciplinar que envolve pesquisa em áreas como Bancos de Dados, Processamento de Imagem, Inteligência Artificial, Estatística, dentre outras.

8. Agradecimentos

Este trabalho contou com apoio financeiro da FAPESP, CNPq, CAPES e Microsoft Research.

Referências

Almeida, J., Rocha, A., Torres, R., and Goldestein, S. (2008). Making Colors Worth more than a Thousand Words. In *The 23th Annual ACM Symposium on Applied Computing*, pages 1184–1190, Fortaleza.

Antani, A., Kasturi, R., and Jain, R. (2002). A Survey on the Use of Pattern Recognition Methods for Abstraction, Indexing and Retrieval of Images and Video. *Pattern Recognition*, 35(4):945–965.

Bugatti, P. H., Traina, A. J. M., and C. Traina, J. (2008). Assessing the best integration between distance-function and image-feature to answer similarity queries. In *Proceedings of the ACM symposium on Applied computing*, pages 1225–1230.

Carvalho, A. C. P. L. F., Brayner, A., Loureiro, A., Furtado, A. L., v. Staa, A., Lucena, C. J. P., S., C. S., Medeiros, C. M. B., Lucchesi, C. L., Silva, E. S., Wagner, F. R., Simon, I., Wainer, J., Maldonado, J. C., Oliveira, J. P. M., Ribeiro, L., Velho, L., calves, M. A. G., Baranauskas, M. C. C., Mattoso, M., Ziviani, N., Navaux, P. O. A., da S. Torres, R., Almeida, V. A. F., Jr., W. M., and Kohayakawa, Y. (2006). Grandes Desafios da Pesquisa em Computação no Brasil – 2006 - 2016. In *Seminário Grandes Desafios da Sociedade Brasileira de Computação*, São Paulo, SP, Brasil.

Chang, J.-W. and Kim, Y.-J. (2001). Spatial-match iconic image retrieval with ranking in multimedia databases. In *WAIM '01: Proceedings of the Second International Conference on Advances in Web-Age Information Management*, pages 3–13, London, UK.

Chang, S. K., Shi, Q. Y., and Yan, C. W. (1987). Iconic indexing by 2-d strings. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(3):413–428.

Ciaccia, P., Patella, M., and Zezula, P. (1997). M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. In *Proceedings of 23rd International Conference on Very Large Data Bases*, pages 426–435, Athens, Greece.

Câmara, G., Casanova, M., Hemerly, A., Magalhães, G., and Medeiros, C. (1997). *Anatomia de Sistemas de Informação Geográfica*. Sagres Editora, Curitiba-PR, Brasil.

Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2).

del Val Cura, L. M. (2002). *Um modelo para recuperação por conteudo de imagens de sensoriamento remoto*. PhD thesis, Universidade Estadual de Campinas. Instituto de Computação, Campinas, SP.

Gonzalez, R. C. and Woods, R. E. (2001). *Digital image processing*. Electrical and Computer Engineering Series. Addison-Wesley Longman Publishing Co. Inc., 2nd edition.

Kherfi, M. L., Ziou, D., and Bernardi, A. (2004). Image retrieval from the world wide web: Issues, techniques, and systems. *ACM Comput. Surv.*, 36(1):35–67.

Kim, D.-H., Chung, C.-W., and Barnard, K. (2005). Relevance feedback using adaptive clustering for image similarity retrieval. *Journal of Systems and Software*, 78(1):9–23.

Lin, W.-H., Jin, R., and Hauptmann, A. (2003). Web image retrieval re-ranking with relevance model. In *International Conference on Web Intelligence*, page 242.

Liu, Y., Zhang, D., Lu, G., and Ma, W.-Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262–282.

Penatti, O. B. and Torres, R. (2007). Descritor de Relacionamento Espacial baseado em Partições. In *XXVI Concurso de Trabalhos de Iniciação Científica, XXVII Congresso da Sociedade Brasileira de Computação*, Rio de Janeiro, Brazil.

Rui, Y., Huang, T. S., Ortega, M., and Mehrotra, S. (1998). Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655.

Sebastian, T. B., Klein, P. N., and Kimia, B. B. (2002). Shock-based indexing into large shape databases. In *ECCV*, pages 731–746.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380.

Swain, M. and Ballard, D. (1991). Color Indexing. *International Journal of Computer Vision*, 7(1):11–32.

Tamura, H., Mori, S., and Yamawaki, T. (1978). Texture features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8(6):460–473.

Torres, R. and Falcão, A. X. (2006). Content-Based Image Retrieval: Theory and Applications. *Revista de Informática Teórica e Aplicada*, 13(2):161–185.

Xu, Z., Xu, X., Yu, K., and Tresp, V. (2003). A Hybrid Relevance-Feedback Approach to Text Retrieval. *Proceedings of the 25th European Conference on Information Retrieval Research, Lecture Notes in Computer Science*, 2633:81–293.

Yan, R. and Hauptmann, A. G. (2007). A review of text and image retrieval approaches for broadcast news video. *Information Retrieval*, 10(4-5):445–484.

Zhang, D. and Lu, G. (2004). Review of Shape Representation and Description. *Pattern Recognition*, 37(1):1–19.

Zhou, X. S. and Huang, T. S. (2003). Relevance feedback in image retrieval: A comprehensive review. *Multimedia Systems*, 8(6):536–544.