

Uma nova abordagem CBIR baseada em realimentação de relevância e classificação por OPF

André Tavares da Silva , Léo Pini Magalhães (Orientador) , Alexandre Xavier Falcão (Co-Orientador)

Departamento de Engenharia de Computação e Automação Industrial (DCA)
Faculdade de Engenharia Elétrica e de Computação (FEEC)
Universidade Estadual de Campinas (Unicamp)
Caixa Postal 6101, 13083-970 – Campinas, SP, Brasil

{atavares,leopini}@dca.fee.unicamp.br, afalcao@ic.unicamp.br

Abstract – More recently, some CBIR approaches have shown the use of relevance feedback to train a pattern classifier which selects relevant images for retrieval. This paper revisits this strategy by using an optimum-path forest (OPF) classifier. During relevance feedback iterations, the proposed method uses the OPF classifier to decide which database images are relevant or not. Images just classified as relevant are sorted and presented to the user for a new iteration. Such images are ordered according to the normalized distance using relevant and irrelevant prototypes, computed previously by the OPF classifier. Our experiments show that the proposed approach requires few iterations, being faster and more effective than methods based on support vector machines.

Keywords – CBIR, relevance feedback, image processing.

1. Introdução

Com o crescimento da internet e a popularização dos dispositivos para captura de imagens como câmeras digitais e *scanners*, a disponibilidade de coleções de imagens tem crescido rapidamente nos últimos anos [3]. Por isso, os usuários necessitam cada vez mais de ferramentas eficientes para pesquisar, navegar e recuperar essas informações em diferentes domínios, como sensoriamento remoto, moda, prevenção de crime, publicidade, medicina, arquitetura, entre outros. Para este propósito, têm sido desenvolvidos muitos sistemas de recuperação de imagens.

Existem duas linhas principais: baseados em texto e em conteúdo. Na abordagem baseada em texto, o processo de recuperação consiste em comparar os termos de uma consulta textual, definida por um usuário, com as anotações associadas às imagens e, a partir dessa comparação, retornar um conjunto de imagens. Existem duas principais desvantagens nesta abordagem: a necessidade de um trabalho humano considerável para realizar as anotações e a imprecisão das anotações devido à subjetividade da percepção humana [1], já que diferentes pessoas podem associar diferentes anotações para uma mesma imagem.

Para superar essas desvantagens em sistemas de recuperação de imagens, foram introduzidos os sistemas de recuperação de imagens baseados em conteúdo (CBIR - *content-based image retrieval*). Nos sistemas CBIR, as imagens são indexadas pelo seu conteúdo visual, tais como cor, textura e forma.

Nesses sistemas, a anotação manual não é necessária. O processo de busca consiste basicamente em, dado um padrão de consulta (por exemplo uma imagem), calcular a sua similaridade em relação às imagens armazenadas na base, exibindo as mais similares.

Um grande desafio na recuperação de imagens é saber interpretar o desejo do usuário [5]. E esse desejo varia conforme a realidade cultural de cada pessoa, ou seja, imagens têm significado diferente para cada indivíduo e depende do conhecimento e vivência que cada um tem sobre as imagens.

A técnica de realimentação de relevância tem sido bastante utilizada para diminuir a lacuna semântica existente entre os sistemas computacionais e a subjetividade das pessoas. Essa técnica possibilita ao usuário expressar sua necessidade na especificação de uma consulta sem precisar recorrer a propriedades de mais baixo nível para representação da imagem. O usuário informa quais as imagens ele considera relevantes em um conjunto de imagens retornado pelo sistema. O algoritmo de realimentação de relevância aprende a vontade do usuário durante um determinado número de iterações. Dessa forma, o sistema retorna imagens cada vez mais similares à vontade do usuário, aprendendo o conceito estabelecido por ele.

Neste trabalho, propomos uma nova abordagem para CBIR com realimentação de relevância utilizando informações de imagens relevantes e ir-

relevantes selecionada pelo usuário. Para um determinado conjunto de imagens relevantes e irrelevantes, o método calcula uma OPF (Optimum-Path Forest) [4]. Apenas as imagens classificadas como relevantes são ordenadas pela distância e apresentadas ao usuário na próxima iteração. Mostramos que essa estratégia é realmente muito eficaz reduzindo consideravelmente o número de iterações necessárias.

Este artigo está organizado da seguinte forma: na seção 2 são expostos alguns conceitos básicos sobre realimentação de relevância, na seção 3 é apresentada a proposta da tese e por fim, são apresentadas as contribuições e perspectivas sobre o presente trabalho.

2. Conceitos Básicos

Esta seção apresenta alguns conceitos básicos sobre recuperação de imagens por conteúdo adotados neste trabalho.

2.1. Descritores

A extração de características (descritores) é a base da recuperação de informação visual [6]. A percepção visual dos objetos é subjetiva e por isso não existe uma única representação e nem mesmo uma melhor representação para uma dada característica. As principais características extraídas de imagens são cor, textura e forma.

A cor é provavelmente a característica mais utilizada para recuperação visual. Ela é relativamente robusta por apresentar independência do tamanho da imagem e da orientação da mesma. Uma vantagem do uso desse tipo de classificador é que as cores podem ser facilmente associadas a descrições textuais (nome da cor), facilitando a utilização em muitos sistemas CBIR. O histograma de cores é a característica mais utilizada para representar imagens.

Textura é uma propriedade presente em praticamente todas as estruturas, como nuvens, vegetação, paredes, cabelo e outros. Ela contém informação importante sobre o arranjo estrutural da superfície e sua relação com o ambiente.

Os primeiros descritores para descrever a forma de objetos em uma imagem se resumiam em definir informações simples como comprimento, perímetro, área e algumas relações entre essas me-

das (regularidade e compacidade). Mais tarde, outros descritores mais complexos foram definidos, como os pontos de saliência, saliências de segmentos (*Segment Saliences*), Fourier, invariantes de momento (*moment invariants*), entre outros.

2.2. Realimentação de Relevância

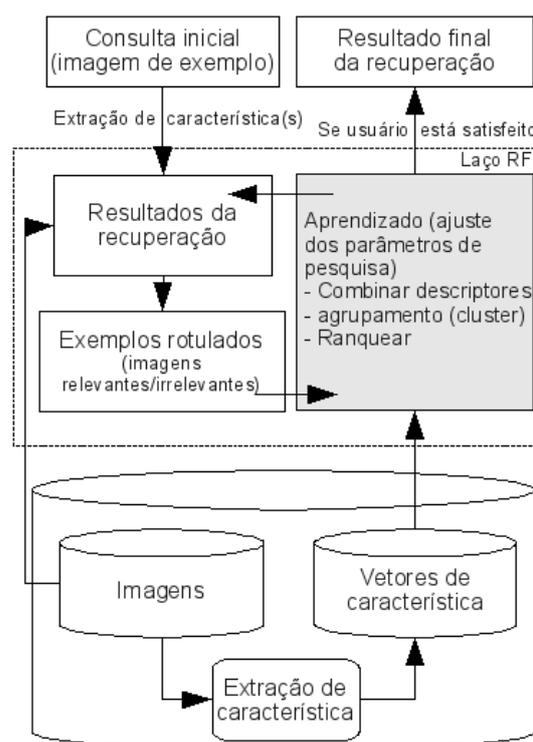


Figura 1. Arquitetura de um sistema de recuperação de imagens por conteúdo com realimentação de relevância.

Esse mecanismo (figura 1) tem por objetivo possibilitar que o usuário expresse a sua necessidade na especificação de uma consulta, sem recorrer a ajustes de 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 de realimentação de relevância “aprende” quais propriedades visuais melhor definem as imagens relevantes a partir das informações fornecidas pelo usuário, ou seja, das imagens por ele indicadas. Assim, após um determinado número de iterações, o sistema retorna as imagens mais similares à imagem de consulta.

Realimentação de relevância endereça duas questões referentes ao processo de recuperação de

imagens por conteúdo. A primeira delas reside na lacuna semântica (*semantic gap*) entre as propriedades visuais de alto nível, através dos quais o usuário tem a percepção da informação visual, e a descrição de baixo nível utilizada para a representação das imagens [3]. 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 realimentação de relevância essas duas questões são contornadas de forma transparente para o usuário.

O algoritmo de aprendizado é um ponto crucial para a definição de um mecanismo de realimentação de relevância. Em alguns trabalhos, o aprendizado consiste em estimar o vetor de característica que melhor representa o padrão de consulta. Em outros, atribuem-se pesos para cada posição do vetor de características e para cada descritor utilizado. Assim, o aprendizado consiste em estimar esses pesos, de forma a melhor representar a percepção visual do usuário. Existem diversos métodos para combinação de descritores, como Movimento de Ponto de Consulta, Aprendizado Probabilístico e Máquinas de Vetores de Suporte.

3. Proposta

OPF é um método de classificação, que representa cada classe de objetos por uma ou mais árvores de caminhos ótimos cujas raízes são amostras chamadas de protótipos. As amostras de treinamento são os nós de um grafo completo, cujos arcos são ponderados pela distância entre os vetores de características de seus nós. Na realimentação de relevância, temos duas classes: imagens relevantes escolhidas pelo usuário e as irrelevantes. Os protótipos escolhidos pelo classificador OPF, são então utilizados para classificar as imagens de acordo com a seleção do usuário.

Seja \mathcal{Z} um de banco de dados de imagem. Para cada imagem $t \in \mathcal{Z}$, temos um vetor de características $\vec{v}(t) \in \mathbb{R}^n$. Ou seja, cada imagem pode ser considerada um ponto no espaço \mathbb{R}^n . A distância $d(s, t)$ entre duas imagens s e t é a distância entre seus vetores de características. Para um ponto inicial de pesquisa s , o método proposto retorna as N imagens $t \in \mathcal{Z}$ mais próximas a s (pesquisa por similaridade). Devido à lacuna semântica, as imagens mais próximas a s podem não ser as mais relevantes

para um determinado usuário. Marcando as imagens que um usuário considera relevante ou não, são criados dois conjuntos: uma lista $\mathcal{I} \subset \mathcal{Z}$ de imagens irrelevantes e uma lista $\mathcal{R} \subset \mathcal{Z}$ de imagens relevantes. O método usa então os conjuntos \mathcal{R} e \mathcal{I} para treinamento da OPF. Apenas N imagens classificadas como relevantes mais próximas serão retornadas ao usuário na próxima interação.

Para realizar essa ordenação, são usados os protótipos relevantes (\mathcal{A}) e irrelevantes (\mathcal{B}) escolhidos na fase de treinamento. O método calcula a distância média $\bar{d}_{\mathcal{A}}(t, \mathcal{A})$ entre cada imagem $t \in \mathcal{Z}$ do banco de dados e imagens do conjunto de protótipos relevantes \mathcal{A} . Também é calculada a distância média $\bar{d}_{\mathcal{B}}(t, \mathcal{B})$ entre imagens do conjunto de protótipos irrelevantes \mathcal{B} . Finalmente, uma a distância média normalizada $\bar{d}(t, \mathcal{A}, \mathcal{B})$ é calculada entre protótipos relevantes e irrelevantes: $\frac{d_{\mathcal{A}}(t, \mathcal{A})}{d_{\mathcal{A}}(t, \mathcal{A}) + d_{\mathcal{B}}(t, \mathcal{B})}$.

Após classificar cada imagem do banco de dados, o método retorna ao usuário um novo conjunto de N imagens relevantes, que contém os menores valores de $\bar{d}(t, \mathcal{A}, \mathcal{B})$. Esse processo é repetido durante algumas interações T e, finalmente, o sistema retorna todas as imagens relevantes obtidas nesse processo.

Usamos o descritor BIC com a distância dLog [7] para avaliar nosso método e comparamos sua eficiência usando a curva precisão-revocação contra dois outros métodos: um baseado em SVM proposto por Tong et al. [8] e outro usando *multi-point query* apenas com imagens relevantes. O primeiro, denominado SAL (SVM Active Learning), é também chamado por *SVM_{ACTIVE}* ou *SVM_{AL}* na literatura. Foi escolhido por ser baseado em uma técnica que é considerada o estado da arte em classificação de imagens. O segundo, chamado de QPM (Query Point Movement) [8], foi selecionado para mostrar a importância das imagens selecionadas como irrelevantes. Nossa abordagem é chamada aqui como OPF, já que é baseada em um classificador OPF.

As imagens a seguir, mostram a curva média de precisão-revocação para duas bases de dados (Corel e PASCAL) após três iterações comparando os métodos QPM, SAL e OPF. Quanto mais alta a curva, melhor o resultado. Podemos ver que nosso método supera a performance dos outros dois. Resultados obtidos por este trabalho podem ser vistos

em [2].

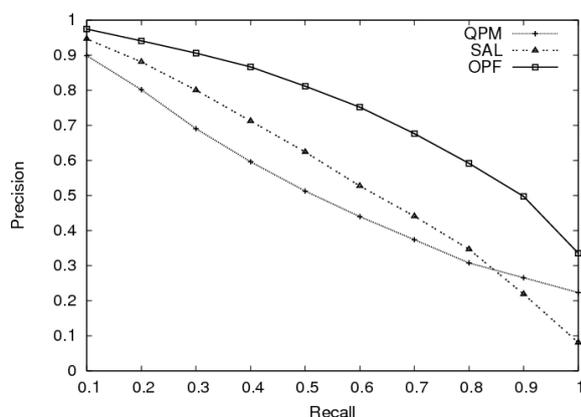


Figura 2. Curva média de precisão-revocação para a base Corel após três iterações.

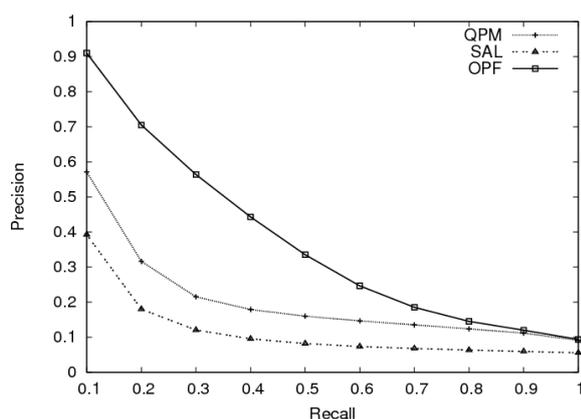


Figura 3. Curva média de precisão-revocação para a base PASCAL após três iterações.

4. Considerações Finais

Pretende-se neste trabalho desenvolver um algoritmo de aprendizado que consiga aprender o mais rápido possível a vontade do usuário.

Existe atualmente uma grande quantidade de imagens digitais disponível, principalmente na Web. Essas imagens vêm sendo geradas, manipuladas e armazenadas em diferentes locais. Para recuperar imagens, é necessário um método eficiente e eficaz para isso.

Os resultados obtidos mostram que nosso método, chamado OPF, necessita de poucas iterações e supera a performance dos outros métodos testados. O número de imagens perdidas pela classificação é insignificante (entre 0,2 e 3%). Além disso,

uma vantagem do nosso método é que ele é aproximadamente vinte vezes mais rápido que o SAL.

Como trabalho futuro, pretendemos usar múltiplos descritores e técnicas para combiná-los. Também pretendemos comparar nosso método com outros mais recentes.

Referências

- [1] P. Alshuth, T. Hermes, J. Kreyb, and M. Roper. *Intelligent Retrieval for Images and Videos*, volume 8. World Scientific, 1997.
- [2] A. T. da Silva, A. X. Falcão, and L. P. Magalhães. A new cbir approach based on relevance feedback and optimum-path forest classification. In *Proc. WSCG 2010*, Plzen, Czech Republic, 2010.
- [3] Ying Liua, Dengsheng Zhanga, Guojun Lua, and Wei-Ying Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, (40):262–282, 2007.
- [4] J. P. Papa, A. X. Falcão, and C. T. N. Suzuki. Supervised pattern classification based on optimum-path forest. *International Journal of Imaging Systems and Technology*, 19(2):120–131, 2009.
- [5] Yong Rui, Thomas S. Huang, and Shih-Fu Chang. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, (10):39–62, 1999.
- [6] Arnold W.M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, Dezembro 2000.
- [7] R. O. Stehling, M. A. Nascimento, and A. X. Falcão. A compact and efficient image retrieval approach based on border/interior pixel classification. In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, pages 102–109, New York, NY, USA, 2002. ACM.
- [8] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia*, pages 107–118, New York, NY, USA, 2001. ACM.