

Uma abordagem CBIR baseada em realimentação de relevância e classificação por floresta de caminhos ótimos

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 approaches require 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 [4]. 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.

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. Para um determinado conjunto de imagens relevantes e irrelevantes, o método calcula uma OPF (Optimum-Path Forest) [5]. 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. Utilizamos este método geral aplicado a diferentes paradigmas.

Este artigo está organizado da seguinte forma: na seção 2 são expostos alguns conceitos básicos sobre realimentação de relevância, na seção 3 são apresentados os métodos desenvolvidos na tese e por fim, são apresentadas as contribuições do presente trabalho.

2. Realimentação de Relevância

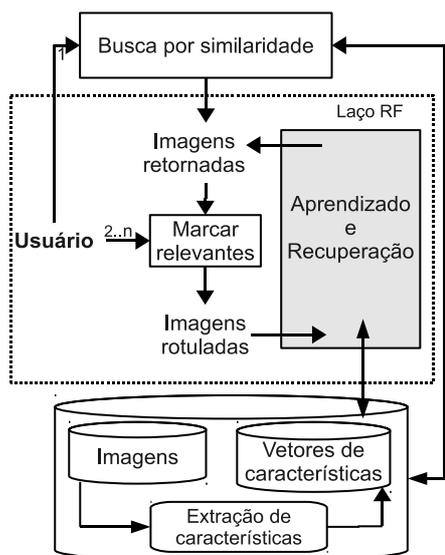


Figura 1. Arquitetura de um sistema de recuperação de imagens por conteúdo com realimentação de relevância.

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 [4]. 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, algoritmos genéticos e programação genética. Técnicas probabilísticas como Bayes e de aprendizado de máquina como redes neurais artificiais e máquinas de vetor de suporte (SVM) são largamente utilizadas em métodos de realimentação de relevância.

3. Métodos de CBIR baseados em OPF

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, é criado um conjunto \mathcal{T} de imagens rotuladas como relevantes ou irrelevantes. A partir desse conjunto rotulado, é criada uma árvore de expansão mínima (MST – Minimum Spanning Tree) de um grafo completo formado pelos elementos em \mathcal{T} . Os elementos adjacentes na MST com diferentes rótulos em \mathcal{T} são definidos como *protótipos*, isto é, elementos mais próximos entre relevantes e irrelevantes no caso desta tese. Removendo-se os arcos entre as diferentes classes, tais amostras adjacentes tornam-se protótipos e são armazenados nos conjuntos \mathcal{S}_R e \mathcal{S}_I de protótipos relevantes e irrelevantes respectivamente. O método usa então os conjuntos \mathcal{S}_R e \mathcal{S}_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{S}_R) e irrelevantes (\mathcal{S}_I) escolhidos na fase de treinamento. O método calcula a distância média normalizada $\bar{d}(t, \mathcal{S}_R, \mathcal{S}_I)$ entre cada imagem $t \in \mathcal{Z}$ e os dois conjuntos de protótipos.

$$\bar{d}(t, \mathcal{S}_R, \mathcal{S}_I) = \frac{\bar{d}(t, \mathcal{S}_R)}{\bar{d}(t, \mathcal{S}_R) + \bar{d}(t, \mathcal{S}_I)}, \quad (1)$$

$$\bar{d}(t, \mathcal{S}_R) = \frac{1}{|\mathcal{S}_R|} \sum_{\forall s \in \mathcal{S}_R} d(s, t), \quad (2)$$

$$\bar{d}(t, \mathcal{S}_I) = \frac{1}{|\mathcal{S}_I|} \sum_{\forall s \in \mathcal{S}_I} d(s, t). \quad (3)$$

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 $\bar{d}(t, \mathcal{S}_R, \mathcal{S}_I)$. Esse processo é repetido durante algumas interações T e, finalmente, o sistema retorna todas as imagens relevantes obtidas nesse processo.

Existem dois paradigmas de aprendizagem em realimentação de relevância em relação às imagens retornadas. No primeiro, a cada interação tenta-se retornar sempre as imagens que o usuário considera mais relevantes, sendo o paradigma mais utilizado em CBIR. Nesta tese este paradigma é denominado como *guloso*. Em outros casos, o usuário estabelece quantas iterações o sistema deverá aprender antes de retornar as imagens ordenadas por relevância. Nas iterações anteriores, o sistema apresenta as imagens mais informativas para auxiliar a aprendizagem do sistema. Este último paradigma nesta tese é chamado de *planejado*.

Usamos o descritor BIC com a distância d_{Log} [9] para avaliar nosso método e comparamos sua eficiência usando a curva precisão-revoação contra dois outros métodos: um planejado baseado em SVM proposto por Tong et al. [10] (denominado SV_{MAL}) e um guloso chamado QEX (*query expansion method*) [6] usando apenas imagens relevantes. Nossa abordagem gulosa é chamada aqui como $GOPF_{RF}$ e a abordagem planejada é denominada $POPF_{RF}$. A Figura 2 mostra a curva média de precisão vs. revocação para a base de dados Corel [11] após três iterações comparando os métodos QEX,

SAL, $GOPF_{RF}$ e OPF. Quanto mais alta a curva, melhor o resultado. Podemos ver que nosso método é mais eficiente do que os demais.

Foram desenvolvidos dois métodos para utilizar combinação de descritores à técnica de realimentação de relevância usando OPF. OPF_{MSPS} utiliza uma abordagem denominada MSPS (*Multi-Scale Parameter Search*) [7] para determinar os parâmetros mais adequados de uma função de combinação, enquanto OPF_{GP} utiliza programação genética para geração de funções de combinação. Esses dois métodos são comparados a um método baseado em programação genética considerado como estado-da-arte para combinação de descritores chamado GP^+ [3] e com o resultado de $GOPF_{RF}$ usando o melhor descritor. A comparação com GP^+ avalia a eficiência dos métodos na combinação de descritores. Além disso, a combinação de descritores é justificada apenas se sua eficácia for superior à utilizar apenas um descritor. Esta é a razão de apresentar a comparação dos resultados com os obtidos usando um único descritor. A Figura 3 mostra a curva média de precisão vs. revocação para a base de dados MPEG7 [8] após três iterações comparando os métodos OPF_{MSPS} , OPF_{GP} , GP^+ e $GOPF_{RF}$ usando o descritor TSDIZ [2].

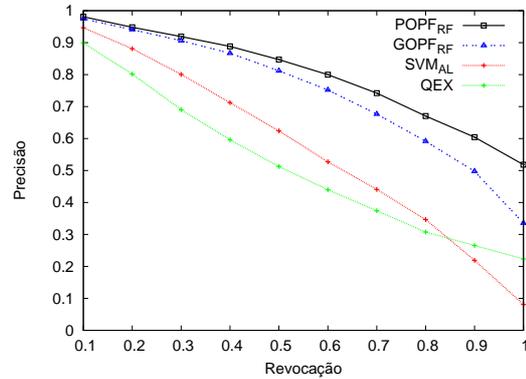


Figura 2. Curva média de precisão-revoação para a base Corel após três iterações.

Também foi desenvolvido nesta tese um método que utiliza a classificação tanto a nível de pixel quanto a nível de descritor como utilizado até então. Neste contexto, o usuário marca quais objetos ele considera relevante em uma imagem e o sistema classifica em todas as outras o que é objeto e o que não interessa (fundo). O usuário pode durante a realimentação de relevância não apenas definir quais as imagens são relevantes ou não, como

também selecionar o que foi erroneamente classificado como objeto ou fundo.

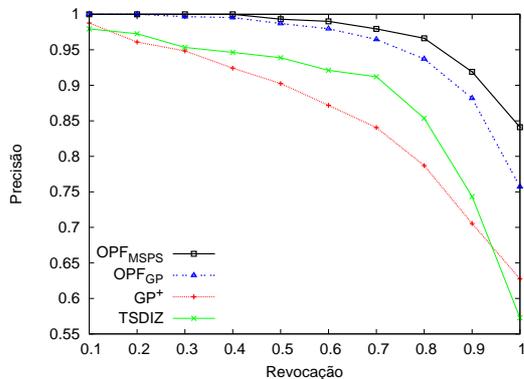


Figura 3. Curva média de precisão-revocação para a base PASCAL após três iterações.

4. Considerações Finais

Nesta tese foi criado um novo método de realimentação de relevância para recuperação de imagens por conteúdo. Esta é a primeira vez que o classificador baseado em floresta de caminhos ótimos é usado e testado para conjuntos de treinamentos pequenos, conforme exigido pela técnica de realimentação de relevância. Também é desenvolvida uma nova métrica para buscar as imagens ordenadas pela maior possibilidade de serem relevantes, baseada na distância média entre os protótipos relevantes e irrelevantes definidos pelo classificador OPF.

Os resultados obtidos mostram que nossos métodos necessitam de poucas iterações e supera a performance dos outros métodos testados. Além disso, uma vantagem do nosso método é que ele é em média 52 vezes mais rápido que os métodos baseados em SVM.

É demonstrado que este novo método de realimentação de relevância pode ser aplicado utilizando-se tanto o paradigma guloso quanto o paradigma planejado. Além disso, a utilização de múltiplos descritores pode ser aplicada diretamente, como mostrado usando tanto MSPS quanto programação genética. O trabalho também propõe um método que atua nos níveis de pixel e de descritor.

Referências

[1] P. Alshuth, T. Hermes, J. Kreyb, and M. Roper. *Intelligent Retrieval for Images and Videos*, volume 8. World Scientific, 1997.

[2] F. A. Andaló, P. A. V. Miranda, R. S. Torres, and A. X. Falcão. Shape feature extraction and description based on tensor scale. *Pattern Recognition*, 43(1):26–36, 2010.

[3] C.D. Ferreira, J.A. Santos, R. da S. Torres, M.A. Gonçalves, R.C. Rezende, and Weiguo Fan. Relevance feedback based on genetic programming for image retrieval. *Pattern Recognition Letters*, 32(1):27–37, 2011.

[4] 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.

[5] 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.

[6] Kriengkrai Porkaew, Kaushik Chakrabarti, and Sharad Mehrotra. Query refinement for multimedia similarity retrieval in mars. In *Proceedings of ACM Multimedia*, pages 235–238, 1999.

[7] G. C. S. Ruppert, F. O. Favretto, A. X. Falcão, C. L. Yassuda, and F. P. G. Bergamo. Fast and accurate image registration using the multiscale parametric space and grayscale watershed transform. In *Systems, Signals and Image Processing*, pages 17–19, Rio de Janeiro, June 2010. IEEE Computer Society.

[8] T. Sikora. The mpeg-7 visual standard for content description—an overview. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):696–702, 2001.

[9] R. O. Stehling, M. A. Nascimento, and A. X. Falcão. A compact and efficient image retrieval approach based on border/interior pixel classification. In *Information and knowledge management*, pages 102–109, New York, NY, USA, 2002. ACM.

[10] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *ACM international conference on Multimedia*, pages 107–118, New York, NY, USA, 2001. ACM.

[11] J. Z. Wang, J. Li, and G. Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:947–963, 2001.