

Sistema para pesquisa de imagens com Retroacção de Relevância*

Rui M. Jesus¹, Arnaldo J. Abrantes¹, Jorge S. Marques²

¹M2A - Instituto Superior de Engenharia de Lisboa

Rua Conselheiro Emídio Navarro, nº 1, 1940-014 Lisboa, Portugal

Telefone: 218317227, rjesus@deetc.isel.ipl.pt, aja@deetc.isel.ipl.pt

² ISR/IST, Av. Rovisco Pais 1049-001 Lisboa, Portugal

Recentemente, a retroacção de relevância tem sido utilizada para melhorar o desempenho dos sistemas de pesquisa em base de dados de imagens. Este artigo apresenta um método de Retroacção de Relevância baseado no classificador de Mínimos Quadrados Regularizado e numa técnica de selecção de imagens que permite aumentar a capacidade de aprendizagem do método. São apresentados resultados de testes experimentais.

1. INTRODUÇÃO

Os sistemas de pesquisa de imagens na web ou em base de dados procuram encontrar conjuntos de imagens relevantes para o utilizador. Esta tarefa é difícil porque não é possível anotar manualmente grandes bases de dados, utilizando palavras-chave que descrevam a informação contida na imagem. Por isso, os sistemas automáticos utilizam características de baixo nível da imagem (e.g. cor, textura, forma). No entanto, estes sistemas apresentam um desempenho fraco porque têm dificuldade em capturar conceitos semânticos, e.g., flores ou pessoas. Este problema é conhecido como *falha semântica* [1]. Para ultrapassar esta dificuldade, o utilizador deve poder interagir com o sistema providenciando informação adicional durante o processo de busca. Esta operação é conhecida como *retroacção de relevância* [2].

A primeira técnica de *retroacção de relevância* apresentada para sistemas de pesquisa baseados em conteúdo (CBIR) foi inspirada no trabalho de Rocchio [3] proposto no contexto da procura de documentos. Esta técnica movimenta a *query* (vector de características) na direcção das características das imagens relevantes no espaço de características. Rui Yong *et al.* [4] melhorou esta técnica atribuindo diferentes pesos a cada componente do vector de características. Estas duas aproximações exibem um mau desempenho quando a distribuição das imagens relevantes no espaço características é multimodal. Estas dificuldades são ultrapassadas utilizando *queries* multi-ponto [5] constituídas por vários vectores de características de várias imagens. Inicialmente a *query* é caracterizada pelo vector de características de uma imagem. Depois de cada iteração, um subconjunto de imagens relevantes é escolhido e as suas características são adicionadas ao modelo da *query*. As imagens mais próximas da *query* multi-ponto são retornadas pelo sistema ao utilizador. Mais recentemente, foram propostas várias técnicas de retroacção de relevância baseadas em métodos de classificação [6-8], que classificam as imagens da base de dados como relevantes ou não relevantes. Nesta operação têm sido extensivamente utilizadas as Máquinas de Vectores de Suporte (SVM) [7,9].

Este artigo apresenta um sistema de pesquisa de imagens com retroacção de relevância. Esta informação é utilizada para construir interactivamente um conjunto de imagens relevantes e não relevantes (conjunto de treino) e para recursivamente treinar um classificador binário. Esta tarefa é realizada pelo classificador de

* Este trabalho foi parcialmente publicado em [11].

Mínimos Quadrados Regularizado (RLSC), recentemente proposto em [10]. Este trabalho também propõe uma técnica para escolher as imagens a serem avaliadas pelo utilizador durante o processo de pesquisa. A selecção de imagens a visualizar faz um compromisso entre dois critérios conflituosos: a escolha entre as imagens mais relevantes e as mais informativas [6] de acordo com o classificador.

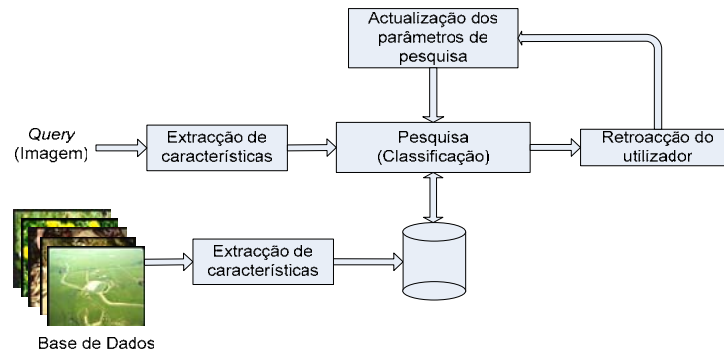


Figura 1. Sistema de pesquisa de imagens com retroacção de relevância.

2. SISTEMA DE PESQUISA DE IMAGENS

O objectivo do sistema proposto neste trabalho é procurar imagens semelhantes à imagem da *query* numa base de dados, utilizando o classificador RLS. Este é um problema especial de classificação, porque a maior parte dos pontos (vectores de características das imagens) não estão classificadas e apenas o vector de características da *query* é classificado como relevante. Assim, o conjunto de treino é constituído apenas por um único vector. Para melhorar o desempenho pede-se ao utilizador para visualizar algumas imagens adicionais e classifica-las quanto à sua relevância (ver figura 1). Em cada iteração o sistema mostra ao utilizador um conjunto de novas imagens e pede-lhe que as classifique como positivas ou negativas. O classificador é actualizado com base no novo conjunto de treino e o processo repete-se.

O classificador escolhido para esta tarefa deve ter uma boa capacidade de generalização porque o número de pontos classificados é sempre uma pequena fracção de toda a base de dados. Para resolver este problema é necessário solucionar duas questões principais:

- Como projectar um classificador com um conjunto de treino pequeno?
- Como escolher as melhores imagens para serem classificadas pelo utilizador?

Estes dois pontos são abordados nas secções 3 e 4.

3. CLASSIFICAÇÃO USANDO O RLSC

Esta secção descreve o classificador de Mínimos Quadrados Regularizado [10] utilizado no algoritmo de retroacção de relevância proposto neste artigo. Dado um conjunto de treino $S_m = \{(x_i, y_i)_{i=1}^m\}$, onde x_i designa o i -ésimo padrão e $y_i \in \{-1, 1\}$ a sua classe, pretende-se estimar a fronteira de decisão entre as duas classes (relevante/ não relevante). A função discriminante utilizada é,

$$f(x) = \left(\sum_{i=1}^m c_i K(x_i, x) \right), \quad (1)$$

onde $K(x, x')$ representa o kernel gaussiano $K(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$, m é o número de pontos de treino e $c = [c_1, \dots, c_m]^T$ é um vector de coeficientes que é obtido minimizando o funcional regularizado [10],

$$\frac{1}{m} \sum_{i=1}^m [y_i - f(x_i)]^2 + \gamma \|f\|_k^2, \quad (2)$$

onde γ é um número real positivo, e $\|f\|_k^2$ é a norma do espaço H_K associado ao kernel K [10]. A minimização da equação (2) conduz a,

$$(m\mathcal{I} + K)c = y, \quad (3)$$

onde \mathcal{I} é a matriz identidade, K é uma matriz definida positiva com os elementos $K_{i,j} = K(x_i, x_j)$ e y é um vector com coordenadas y_i . O treino do classificador reduz-se à solução de um sistema linear de m equações.

Os pontos $\{x_i\}$ com $f(x_i) \leq 0$, são classificados na classe 2 ($y_i = -1$), e os pontos com $f(x_i) > 0$ são classificados na classe 1 ($y_i = 1$).

4. SELECÇÃO DOS MELHORES PONTOS

Como a *query* inicial é constituída por uma única imagem o classificador inicial tem um desempenho muito pobre. Procura-se que o utilizador forneça informação adicional o que é feito pedindo ao utilizador que visualize algumas imagens adicionais seleccionadas pelo sistema e as classifique em relevantes/não relevantes. Esta selecção tem um papel importante no processo de aprendizagem porque guia a construção do conjunto de treino. Neste trabalho é utilizada uma aproximação híbrida para seleccionar N imagens em cada iteração. As imagens são seleccionadas de acordo com dois critérios. Primeiro, é seleccionado um conjunto de imagens mais informativas i.e., imagens que permitem que o sistema aprenda a distribuição dos dados e melhore o classificador. Também é mostrado o conjunto das imagens mais relevantes seleccionadas pelo classificador. O conjunto de pontos mais informativos é obtido da seguinte forma. Primeiro, todos os pontos em que,

$$|f(x_i)| < T, \quad (4)$$

são considerados como pontos ambíguos, sendo T um limiar definido pelo utilizador. A seguir, todos os pontos são ordenados de acordo com o número de vizinhos dentro de uma circunferência de raio R ,

$$N_i = \left| \left\{ x_j : \|x_i - x_j\| < R \right\} \right| \quad i \neq j \quad (5)$$

O ponto com maior N_i é escolhido por representar mais pontos ambíguos e os seus vizinhos dentro da circunferência são removidos. O processo é repetido $N/2$ vezes para seleccionar $N/2$ pontos. Esta estratégia permite obter informação em regiões do espaço de características que ainda não foram exploradas, e permite afinar a fronteira de decisão.

5. RESULTADOS EXPERIMENTAIS

O método de retroacção de relevância proposto foi testado com dados sintéticos e dados reais.

5.1 Dados Sintéticos – Misturas de Gaussianas

O algoritmo de retroacção de relevância foi testado utilizando um conjunto de vectores de características gerados aleatoriamente por uma mistura de 7 Gaussianas com uma forte sobreposição entre os modos

Gaussianos (see fig. 2a). Os pontos gerados por duas destas Gaussianas (25%) foram classificados na classe 1 (relevante) e os restantes pontos na classe 2 (não relevantes). Foram escolhidos aleatoriamente dois pontos de cada classe para treinar o classificador RLS.

As figuras 2b-d, mostram os resultados obtidos pelo algoritmo em três iterações, onde se pode visualizar os pontos classificados como relevantes (amarelos), os não relevantes (verdes), os pontos ambíguos (vermelhos) e os pontos ambíguos seleccionados (cruzes azuis). O número de pontos ambíguos diminui com o número de iterações. Neste exemplo foram seleccionados os pontos ambíguos e 10 pontos relevantes em cada iteração.

A figura 3, mostra a evolução da probabilidade de erro do classificador RLS que diminui durante a experiência. A figura 3a, mostra a evolução da probabilidade de erro. Cada curva é obtida utilizando um conjunto de teste de comprimento diferente. O desempenho do algoritmo proposto não é alterado com o aumento do número de pontos não classificados no conjunto de teste. A figura 3b, compara duas técnicas para a selecção dos pontos a visualizar pelo utilizador: a estratégia da secção 4 e a selecção aleatória. O método proposto apresenta melhores resultados.

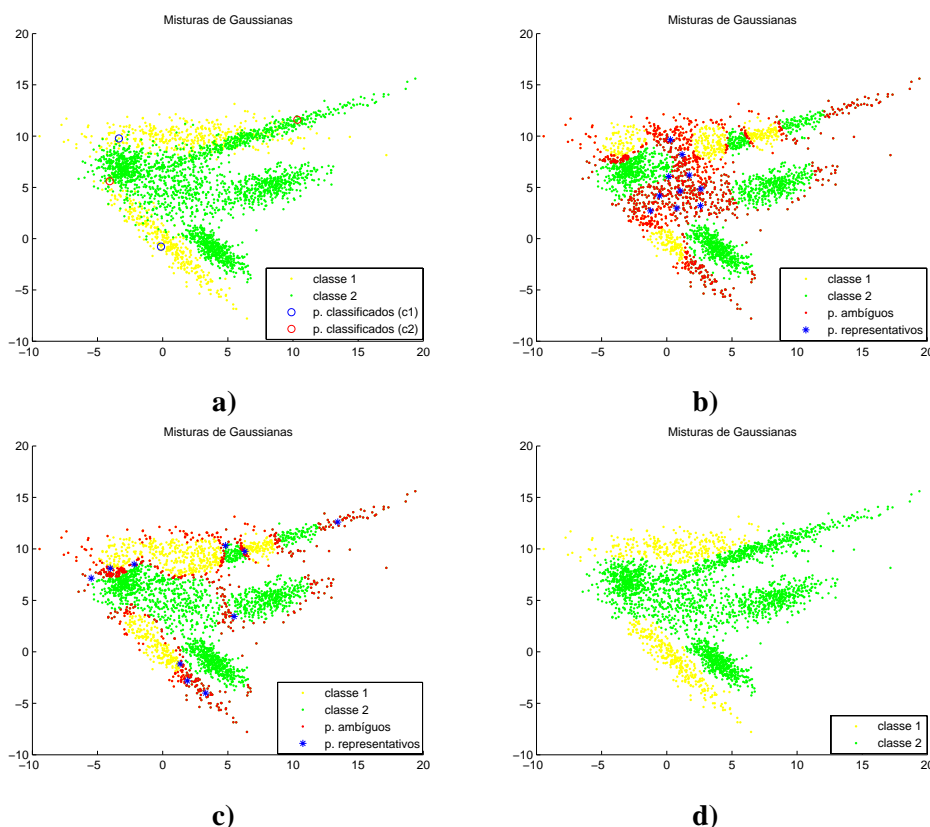


Figura 2. Desempenho do sistema com dados sintéticos: a) ground truth; b)c) resultados intermédios em 2 iterações; d) resultados finais.

5.2 Dados Reais – Imagens

O sistema proposto foi testado com uma base de dados com 630 imagens, 530 seleccionadas do conjunto da UC Berkeley e 100 imagens do CD people II da Corel stock photo library. Estas imagens foram divididas em 23 classes pertencentes a 5 grupos: cenários, animais, automóveis, flores e pessoas.

De forma a avaliar a técnica de retroacção de relevância proposta, todas as imagens da base de dados foram testadas como *query* e para cada uma delas foram calculadas as curvas do precisão e do *recall*. Os valores médios destas curvas são apresentados na figura 4. Cada figura apresenta uma curva com os resultados obtidos sem retroacção de relevância e 2 curvas obtidas depois de 3 iterações (14 imagens por iteração). Estas duas curvas foram obtidas utilizando duas estratégias para a selecção das imagens seleccionadas:

selecção aleatória e método da secção 4. Ambas as estratégias exibem melhor desempenho que o sistema sem retroacção de relevância. No entanto, o método descrito na secção 4 apresenta o melhor desempenho.

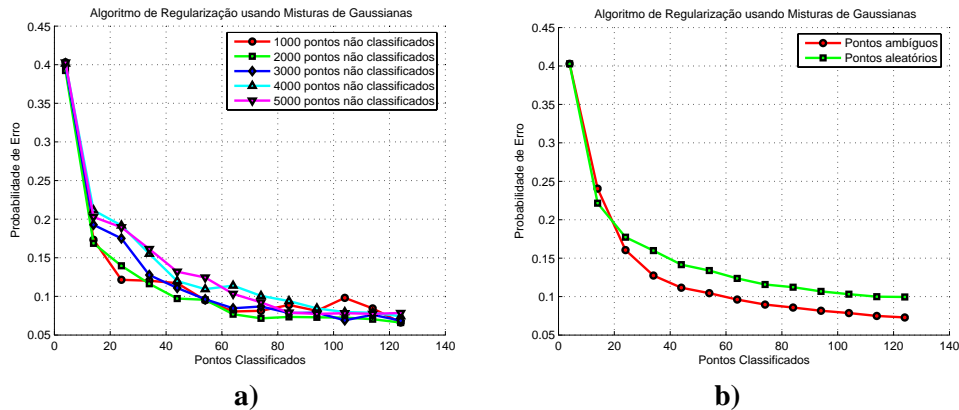


Figura 3. Evolução da probabilidade de erro: a) utilizando vários comprimentos para o conjunto de teste ; b) selecção dos melhores pontos versus selecção aleatória.

Estes testes foram obtidos mostrando 14 imagens por iteração ao longo de 3 iterações. Para comparar com estes resultados foram realizados testes semelhantes utilizando 10 imagens em 4 iterações e 20 imagens em 2 iterações. A figura 5 e a tabela 1 mostram os resultados obtidos com estes três testes. Os três casos apresentam resultados semelhantes. O segundo caso é mais adequado quando o utilizador pretende resultados com poucas imagens para classificar em cada iteração. No entanto, se o requisito for poucas iterações, o terceiro caso é o mais apropriado. O primeiro caso representa um bom compromisso entre o número de imagens para retroacção e o número de iterações.

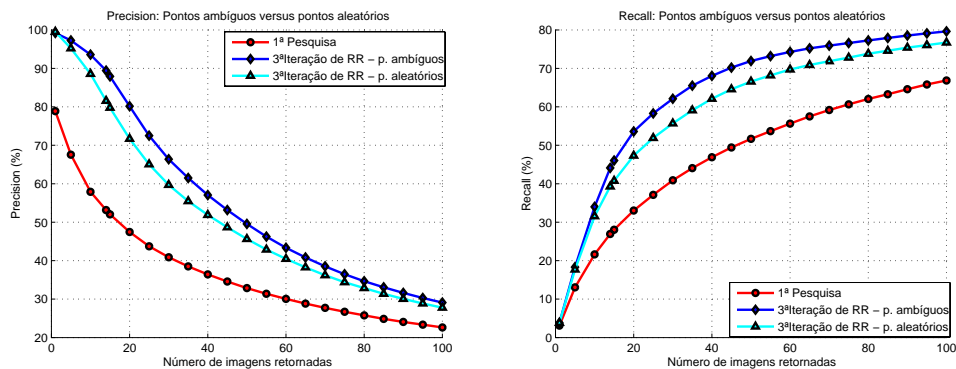


Figura 4. Média da precisão e do recall obtida pelo algoritmo de retroacção de relevância utilizando a técnica de selecção dos melhores pontos versus a selecção aleatória.

6. CONCLUSÕES E TRABALHO FUTURO

Este artigo apresenta um sistema de pesquisa de imagens numa base de dados com retroacção de relevância. O método de retroacção de relevância proposto é baseado no classificador de Mínimos Quadrados Regularizado e numa técnica para escolher conjuntos de imagens durante o processo de aprendizagem. O algoritmo de retroacção de relevância foi testado com uma base de dados com 630 imagens sendo obtido um bom desempenho com uma precisão média de 80% e um recall médio de 54% (14 imagens retornadas) ao fim de três iterações. No futuro será necessário testar o método de retroacção de relevância com bases de dados maiores. Em relação ao classificador, é necessário acrescentar técnicas de aprendizagem considerando a informação obtida em outras sessões e por outros utilizadores.

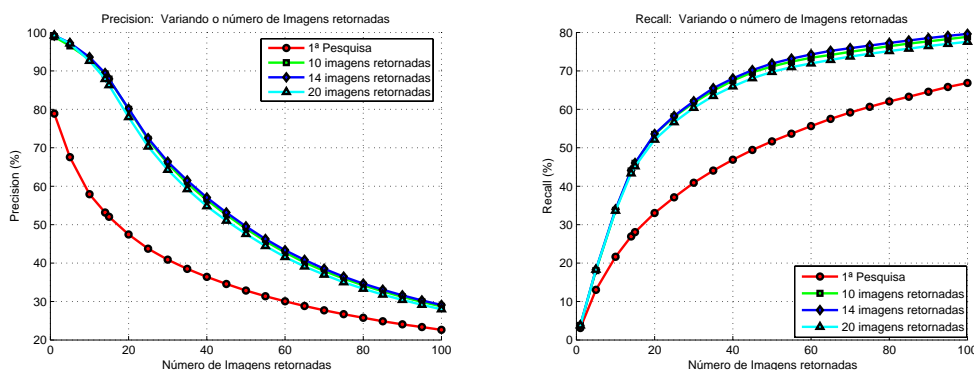


Figura 5. Media da precisão e do recall obtidos com o algoritmo de retroação de relevância utilizando 10, 14, 20 imagens para retroação em cada iteração.

Iteração	10 imagens para retroação	14 imagens para retroação	20 imagens para retroação
1º	0,43	0,37	0,31
2º	0,12	0,10	0,07
3º	0,07	0,06	-
4º	0,05	-	-

Tabela 1. Evolução (em média) da probabilidade de erro obtida pelo classificador RLS.

Referências

- [1] Arnold W. M. Smeulders and Marcel Worring and Simone Santini and Amarnath Gupta and Ramesh Jain, "Content-Based Image Retrieval at the End of the Early Years", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 22, No. 12, 2000, p. 1349-1380.
- [2] Alberto Del Bimbo, *Visual information retrieval*, Morgan Kaufmann Publishers Inc., 1999.
- [3] J. J. Rocchio, Jr., "Relevance Feedback in Information Retrieval," in Gerard Salton, editor, *The SMART Retrieval System: Experiments in Automatic Document Processing*, Prentice-Hall, Inc., 1971, p. 313-323.
- [4] Yong Rui, Thomas S. Huang, and Sharad Mehrotra., "Relevance feedback techniques in interactive content-based image retrieval", in *Storage and Retrieval for Image and Video Databases (SPIE)*, 1998, p. 25-36,.
- [5] Kriengkrai Porkaew and Kaushik Chakrabarti, "Query refinement for multimedia similarity retrieval in MARS", in *ACM Multimedia (1)*, 1999, p. 235-238.
- [6] I. Cox, M. Miller, T. Minka, T. Papatthornas, and P. Yianilos, "The bayesian image retrieval system, pichunter: Theory, implementation, and psychophysical experiments", 2000.
- [7] Simon Tong and Edward Chang, "Support vector machine active learning for image retrieval", in *MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia*, ACM Press, 2001, p. 107-118.
- [8] M. E. J. Wood, N. W. Campbell, and B. T. Thomas, "Iterative refinement by relevance feedback in content-based digital image retrieval", in *ACM Multimedia 98*, September, 1998, p. 13-20.
- [9] Marin Ferecatu, Michel Crucianu, and Nozha Boujemaa, "Sample selection strategies for relevance feedback in region-based image retrieval", in *PCM (2)*, 2004, p. 497-504.
- [10] Tomaso Poggio and Steve Smale, "The mathematics of learning: Dealing with data", *Notice of American Mathematical Society*, Vol. 50, No. 5, 2003, p. 537-544.
- [11] Rui M. Jesus, Arnaldo J. Abrantes, and Jorge S. Marques. "Relevance feedback in CBIR using the RLS classifier". *5th EURASIP Conference focused on Speech and Image Processing, Multimedia communications and Services, Bratislava*, Junho, 2005.