

Vocabulários Visuais Aplicados à Detecção de Edifícios em Fotografias Históricas

Natália C. Batista¹, Ana Paula B. Lopes^{1,2}, e Arnaldo de A. Araújo^{1*}

¹ Departamento de Ciência da Computação
Universidade Federal de Minas Gerais – UFMG
31270-901, Belo Horizonte, MG, Brasil

² Departamento de Ciências Exatas e Tecnológicas
Universidade Estadual de Santa Cruz – UESC
45662-000, Ilhéus, BA, Brasil
{natalia,paula,arnaldo}@dcc.ufmg.br
<http://www.npdi.dcc.ufmg.br>

Abstract. The digitization of historical documents is an effective strategy of make them publicly available while preventing degradation of the original sources. The public Archives of the Brazilian Minas Gerais state (APM), has a collection of digitized historical photographs. The availability of digital copies makes it possible to apply Content-Based Image Retrieval techniques to alleviate the huge manual effort that is put into their description and indexing nowadays. In this work, it is proposed a technique based on bag-of-keypoints descriptors to identify images containing buildings in the APM collection. Bag-of keypoints is an efficient image representation technique, which has been proved robust to occlusion and several kinds of variations. The results of experiments show that, despite of the poor quality of the images, this technique is able to provide recognition rates comparable to those reported by similar papers dealing with modern high quality images.

Key-words: Digital Image Processing, Computer Vision, Image classification, Bag-of-keypoints, Visual vocabulary, Historical photographs

1 Introdução

A preservação de acervos de documentos históricos inclui, muitas vezes, sua digitalização, que apresenta-se como forma eficaz de viabilizar o acesso público a grandes acervos e equalizar o severo compromisso entre conservação e acesso. Frequentemente, documentos históricos apresentam alto grau de degradação, causada pela ação do tempo ou danos sofridos (escurecimento do papel, ressecamento da tinta, umidade, exposição à luz, armazenamento e manipulação inadequados, interferência frente-verso), resultando em imagens digitais com baixo nível de legibilidade e em alguns casos extremos, totalmente ilegíveis [1].

* Os autores agradecem o apoio financeiro das agências nacionais CNPq, CAPES e FAPEMIG e ao Dr. Eduardo A. do Valle Jr., pelas valiosas discussões e sugestões.

A digitalização contribui para evitar o manuseio desses documentos e consequentemente a aceleração do processo de degradação, produzindo centenas de gigabytes de informações. Devido à expansão de projetos de digitalização de acervos históricos e preservação digital da memória, um vasto número de imagens de documentos históricos têm sido gerados por diversos projetos espalhados pelo mundo. Essa enorme quantidade de dados do patrimônio cultural de uma cidade ou estado, por exemplo, necessita de meios para catalogá-los e organizá-los, provendo aos seus usuários - tais como historiadores, estudantes e pesquisadores - ferramentas para consulta e recuperação dos documentos desejados, de forma fácil e rápida.

O Arquivo Público Mineiro (APM) ³ é um exemplo de instituição que executa projetos de digitalização em seu acervo, que contém documentos produzidos e acumulados por órgãos da Administração Pública de Minas Gerais e diversos arquivos privados. O APM recolhe e conserva importante patrimônio histórico e cultural, abrangendo os séculos XVIII, XIX e parte do século XX. Além de documentos manuscritos e impressos, reúne mapas, fotografias, filmes, livros, dentre outros [2]. O acervo de fotografias é composto por aproximadamente 75.000 documentos, das quais 14.000 fotografias já estão organizadas para a digitalização e em torno de 6.000 fotografias já foram digitalizadas, indexadas e disponibilizadas em um sistema de informação. A Figura 1 mostra imagens do acervo do APM de fotografias, nas quais pode-se observar alguns exemplos do conteúdo e também da degradação visual.



Figura 1. Exemplos de documentos digitalizados da coleção de fotografias do APM.

Atualmente, no APM, a recuperação das imagens é realizada com base na informação textual, obtida pelo processo manual de indexação, que consiste em atribuir termos de pesquisa para as fotografias. A tarefa de descrição e indexação possui relação direta com as dimensões do acervo e com o número de conceitos anotados, podendo durar vários anos, pois em geral é uma tarefa realizada de forma manual, descrevendo-se um documento por vez [3]. Nesse contexto, ferramentas capazes de auxiliar no processo de anotação podem tornar-se valiosas.

³ <http://www.siaapm.cultura.mg.gov.br>, acessado em 16 de maio de 2009.

Em [2] é realizada uma análise da base de imagens do APM e um levantamento baseado em entrevistas com usuários do sistema de informação existente, resultando em uma listagem abrangente de diferentes necessidades de processamento do acervo fotográfico. Tais necessidades incluem: separar imagens que possuem apenas bustos ou rostos e grupos de pessoas; separar imagens de ambientes internos (interior) e de ambientes externos (paisagens naturais), bem como encontrar fotografias de áreas rurais/montanhas/urbanas; e identificar objetos nas imagens (pontes, igrejas, construções, vegetação, fachadas, placas, linhas de transmissão, linhas férreas e etc).

Várias dessas tarefas requerem técnicas de Recuperação de Imagens com Base no Conteúdo (RIBC). Nessa área de pesquisa, um dos maiores desafios consiste em encontrar descritores capazes de expressar, a partir de características de baixo nível – como cor, textura e forma – o conteúdo semântico da imagem. Neste trabalho, propõe-se uma metodologia para auxiliar na identificação automática de imagens que contêm fachadas de edifícios (incluindo casas, igrejas e outras construções semelhantes) no acervo digitalizado do APM, colaborando com a indexação e recuperação dessas imagens, com base no conteúdo visual.

A próxima seção descreve os trabalhos relacionados mais relevantes encontrados na literatura. A Seção 3 detalha a abordagem escolhida, explicando as principais decisões de implementação. A metodologia utilizada para a classificação das imagens é abordada na Seção 4 e os resultados experimentais e as conclusões são expostos nas Seções 5 e 6.

2 Trabalhos relacionados

De um modo geral, pode-se classificar objetos de uma cena em dois tipos: objetos feitos pelo homem e objetos naturais. Objetos feitos pelo homem são caracterizados principalmente pela regularidade aparente e o relacionamento estrutural e espacial entre as características de seus componentes. Objetos naturais, como árvores, rios e nuvens, podem coexistir com objetos feitos pelo homem, o que torna não-trivial a tarefa de detectar sua presença. Edifícios são objetos feitos pelo homem, e sua presença na imagem gera um grande número de bordas, junções, linhas e grupos paralelos em comparação com uma imagem que não contém edifícios de forma predominante. Essas estruturas podem ser generalizadas pela presença de quinas, janelas, portas, limites do edifício, etc.

A maior parte das abordagens encontradas na literatura para recuperar imagens de edifícios utiliza características de linhas e bordas. Em [4], regras do agrupamento perceptivo são aplicadas na recuperação por classificação de cenas que contém objetos feitos pelo homem e objetos naturais por meio da extração de linhas, coterminações, junções, grupos paralelos e polígonos. Nesse trabalho, o espaço de características é particionado em três classes: estruturada (imagens contendo estrutura significativa exibida por objetos feitos pelo homem), não-estruturada (imagens contendo nenhum objeto feito pelo homem) e intermediária (imagens contendo uma mistura das duas classes). Outros trabalhos que realizam o reconhecimento e localização de objetos feitos pelo homem são [5], que utiliza

agrupamento de linha coerente que explora cor, orientação e características espaciais de segmentos de linha e [6], que extrai características de blocos da imagem e classifica cada bloco individualmente, capturando dependências espaciais nos dados usando um modelo baseado em um campo causal multiescalar aleatório.

Em recuperação de imagens de edifícios especificamente, tem-se os trabalhos [7] e [8]. Em [7] a similaridade entre duas imagens é avaliada pela distância de centróides de picos das bandas, sendo cada imagem dividida no espaço da transformação de Hough. Em [8], o reconhecimento é executado em dois estágios hierárquicos. O primeiro estágio utiliza um método de indexação baseado em histogramas locais de cor computados em estruturas com orientação dominante na imagem e no segundo estágio, seleciona-se os melhores candidatos da etapa anterior por meio do casamento de pontos de interesse invariantes à escala.

Uma das vantagens das abordagens que utilizam características de linhas e bordas é que a segmentação e a representação detalhada do objeto não são necessárias, ou seja, a decisão relativa à presença de objetos feitos pelo homem pode ser feita sem a necessidade de localizar e reconhecer um objeto específico. Entretanto, isso requer maior conhecimento sobre as propriedades dos objetos, limitando a abordagem devido à pouca capacidade de generalização.

Até o limite de nosso conhecimento, todos os trabalhos encontrados na literatura utilizam bases de imagens de fotografias recentes, contento em sua maioria edifícios modernos. Deve-se notar que essas bases possuem características distintas da base de fotografias do APM, em que o desgaste, esmaecimento, estado de conservação, artefatos da digitalização, dentre outros, tornam a detecção de edifícios mais desafiadora nessas imagens.

3 Categorização de imagens

A recuperação de imagens pode ser realizada por meio do reconhecimento de categorias de objetos, no qual um detector especializado em uma classe de interesse é aplicado a uma dada imagem, determinando se um objeto dessa classe está presente ou não. O paradigma atual consiste em coletar manualmente um grande conjunto de bons exemplares da categoria do objeto desejado (conjunto de treino), treinar um classificador nesse conjunto e então avaliar o modelo obtido em novas imagens (conjunto de teste) [9].

Dentro desse paradigma, o trabalho [10] apresenta uma abordagem computacionalmente eficiente que tem sido bem sucedida para categorização de objetos e cenas. O método é denominado *bag-of-keypoints*, traduzido neste trabalho por “histograma de palavras visuais”. O *bag-of-keypoints*⁴ é uma analogia às representações *bag-of-words* usadas em recuperação de informação em textos, na qual cada documento é representado por um vetor das palavras que ocorrem nele [11].

A idéia básica desse método é descrever uma imagem como uma coleção não-ordenada de características locais em torno de regiões de pontos de interesse. Para uma representação compacta, um vocabulário visual é geralmente

⁴ Alguns trabalhos utilizam a denominação *bag-of-features*, *bag-of-visual-words* ou *bag-of-words* referindo-se ao mesmo método.

construído por meio do agrupamento (*clustering*) dessas características. Cada grupo (*cluster*) de pontos de interesse é tratado como uma palavra visual no vocabulário. Por meio do mapeamento dos pontos de interesse em uma imagem ao vocabulário visual, pode-se descrever a imagem como um vetor de características de acordo com a presença ou contagem de cada palavra visual [12]. Esse vetor de características é o *bag-of-keypoints* de uma imagem e um classificador pode ser aplicado aos vetores de características de determinado conjunto de imagens para determinar à qual categoria cada imagem pertence.

O método do histograma de palavras visuais e suas extensões (por exemplo, [13] e [14]) têm se destacado recentemente, apresentando robustez a diversos tipos de variação eficiência. Além disso, é capaz de lidar com grande variedade de objetos e cenas utilizando o mesmo algoritmo, já que não faz suposições explícitas a respeito dos objetos e cenas sendo procurados. Devido a essas características, neste trabalho optou-se por uma abordagem baseada em histograma de palavras visuais para realizar a classificação de um subconjunto das imagens do APM.

As próximas subseções descrevem em mais detalhes as opções escolhidas para cada passo da abordagem baseada em histograma de palavras visuais.

3.1 Detecção e descrição dos pontos de interesse

Pontos de interesse são pontos da imagem que possuem grande quantidade de informação em termos de mudanças locais no sinal [15]. Entre os vários métodos existentes para amostrar os pontos de interesse de uma imagem (densamente, aleatoriamente, utilizando um detector de pontos de interesse, etc) e para descrever localmente esses pontos, a Transformação de características invariantes à escala (SIFT, do inglês *Scale-Invariant Feature Transform*) [16] é um dos mais utilizados e que apresenta os melhores resultados em diversas tarefas [11][12].

O SIFT transforma as informações de uma imagem em um conjunto de coordenadas relativas às características locais, razoavelmente invariantes à escala, mudanças de iluminação, rotação, perspectiva e ruído. Cada ponto de interesse é associado a uma localização na imagem, uma escala e uma orientação. O descritor do ponto de interesse é criado computando-se a magnitude do gradiente e a orientação em volta da localização desse ponto em uma escala escolhida. A Figura 2 mostra os pontos de interesse detectados pelo SIFT em algumas imagens do acervo do APM.

Como o descritor SIFT possui 128 dimensões, para tornar a próxima etapa de processamento factível, o método PCA (*Principal Components Analysis*) [17] é utilizado para reduzir o vetor de descritores para 30 dimensões. PCA é uma técnica estatística capaz de identificar padrões em dados e expressá-los de tal forma que suas diferenças e similaridades sejam realçadas, além de possibilitar a compressão dos dados pela redução do número de dimensões, sem grande perda de informação.



Figura 2. Imagens do APM marcadas com pontos SIFT detectados (cruzes vermelhas).

3.2 Construção do vocabulário visual

O vocabulário dos pontos de interesse representativos é formado pelos centróides dos grupos resultantes do agrupamento realizado com o algoritmo *k-means*, que é um método simples e com complexidade computacional reduzida [18]. O *k-means* particiona um conjunto de pontos entre k subconjuntos disjuntos de forma a minimizar a distância intra-classe e maximizar a distância inter-classe. Neste trabalho utiliza-se a distância Euclidiana.

Diferentemente do vocabulário textual de recuperação de informação em textos, o tamanho do vocabulário é determinado pelo número de grupos de pontos de interesse. Tamanhos típicos de vocabulários visuais encontrados na literatura variam de 100 a 1000 palavras. Vários tamanhos de vocabulário foram testados gerando resultados similares, optou-se então por usar 100 palavras visuais. A Figura 3 mostra um exemplo com as palavras visuais do vocabulário formado a partir da detecção de pontos de interesse de um conjunto de imagens do APM. Nota-se que as palavras visuais desse exemplo estão relacionadas a padrões de janelas de edifícios.



Figura 3. Parte de um determinado grupo do vocabulário visual das imagens do APM contendo 128 regiões de 13×13 pixels. O centro de cada região é um ponto de interesse pertencente ao grupo. O número médio de pontos de interesse utilizados para construir o vocabulário é 37210 (que corresponde a 40% do total, por razões de eficiência).

3.3 Construção do histograma de palavras visuais

Nessa abordagem de histograma de palavras visuais, uma imagem é caracterizada por um histograma que conta o número de pontos de interesse dessa imagem associados à cada palavra visual extraída do vocabulário.

As distâncias de cada descritor da imagem a todos os centróides determinados pelo *k-means* foram calculadas de acordo com a distância Euclidiana, no espaço de características. O centróide que possui a menor distância do descritor é determinado, e adiciona-se uma unidade ao *bin* correspondente do histograma. Cada ponto de interesse influencia igualmente a representação e o histograma é normalizado de modo que cada *bin* represente a frequência relativa de uma palavra na imagem considerada. O histograma de palavras visuais ou *bag-of-keypoints* resultante da imagem compõe as características que serão usadas para classificar a imagem.

3.4 Classificação

Após a construção dos vetores de características, ocorre a categorização visual do conjunto de imagens. Cada categoria é mapeada em uma classe, que está associada a determinado rótulo. Pelo método de aprendizado supervisionado, o classificador realiza duas fases: treino e teste. Na fase de treino são utilizadas imagens cujo rótulo é conhecido para gerar um modelo, que utiliza um procedimento de decisão estatístico para distinguir as classes. O modelo é então utilizado na fase de teste para prever a classe das imagens que não possuem um rótulo.

O classificador utilizado para esse fim é o *Support Vector Machines* (SVM) [19], muito utilizado em abordagens de histograma de palavras visuais devido ao seu desempenho [12] [10]. SVM é um método baseado na construção de hiperplanos que separam os exemplos positivos e negativos de uma classe com a maior margem possível. No SVM, o espaço de entrada pode ser mapeado para um espaço de características de maior dimensionalidade por meio do *kernel* utilizado. Idealmente, um *kernel* deve enfatizar somente regiões contendo o conceito alvo e ao mesmo tempo tolerar as variações do fundo da imagem sem amplificar seu efeito. Embora existam vários *kernels* de propósito geral, não é claro qual é o mais eficaz para abordagens de histogramas de palavras visuais no contexto da classificação. Entretanto, na maioria dos trabalhos existentes o *kernel* escolhido é o linear ou o Gaussiano [12]. Neste trabalho é utilizado o *kernel* Gaussiano.

4 Metodologia

Um conjunto de 300 imagens foram selecionadas aleatoriamente do acervo de fotografias digitalizadas do APM. Essas fotografias encontram-se originalmente na resolução de 600 dpi (*dots per inch*), no formato TIFF (*Tagged Image File Format*), digitalizadas utilizando-se o sistema de cores RGB (*Red Green Blue*) e foram reduzidas a 72 dpi no formato PGM (*Portable Gray Map*), em tons de cinza, para redução da complexidade computacional.

Para a construção do *ground-truth*, foram convidados usuários voluntários para realizar a anotação manual, sendo que cada usuário avaliou 100 imagens, escolhendo, para cada uma delas, uma das classes: edifício, não-edifício e intermediária. Essa partição do espaço de características é semelhante à utilizada por [4]. A classe edifício deveria ser escolhida se o usuário julgasse aparecer alguma

fachada de prédio, casa, ou outra construção civil na maior parte da fotografia. A classe não-edifício deveria ser escolhida se o usuário julgasse não aparecer nenhum desses elementos e a classe intermediária deveria ser escolhida se o usuário julgasse que a imagem não está em nenhuma das duas classes anteriores. Dessa forma, cada imagem foi anotada por 3 usuários distintos e a classe a qual a imagem pertence foi definida como aquela escolhida pelo maior número de usuários, semelhante a uma votação. Esse processo gerou 68 imagens da classe edifício, 45 da classe intermediária e 187 da classe não-edifício.

O conjunto de imagens foi dividido em três subconjuntos por sorteio. Os experimentos foram executados 3 vezes utilizando-se dois subconjuntos para treino e o subconjunto restante para teste (*cross-validation*). A cada execução, o vocabulário foi reconstruído usando 40% dos descritores das imagens de treino.

5 Resultados experimentais

Os experimentos foram realizados em dois grupos. O primeiro grupo, com a configuração descrita na Seção 3, resultou em uma taxa média de acertos para a classe edifício de 42,2% e para a classe não-edifício de 90,9%.

No segundo grupo de experimentos, as imagens pertencentes à classe intermediária foram retiradas do conjunto de treino, que foi balanceado para obter um número semelhante de imagens da classe edifício e não-edifício, resultando em uma taxa média de acertos de 76,9% para a classe edifício e 72,0% para a classe não-edifício. A avaliação foi realizada pela porcentagem de imagens classificadas corretamente das classes edifício e não-edifício individualmente, dentro de um intervalo com confiança de 90%. Os resultados são resumidos na Tabela 1 e indicam que a classe intermediária confunde o classificador, pois a taxa de acerto da classe edifícios é aumentada significativamente quando as imagens da classe intermediária são retiradas do conjunto de treino, ainda que as imagens possam conter objetos de ambas classes, embora em menor proporção. As imagens dessa classe apresentam dificuldades de serem classificadas tanto pelos usuários quanto pelo classificador, pois não pertencem claramente a nenhuma das duas classes (edifícios e não-edifícios).

Grupo de experimentos	Classe edifício	Classe não-edifício
1	37,5% – 47,9%	87,2% – 94,6%
2	72,1% – 81,8%	67,6% – 76,5%

Tabela 1. Resultados da classificação. São mostrados os intervalos com confiança de 90%. A taxa média de acertos da classe intermediária aproximadamente 2% no primeiro grupo de experimentos.

Apesar de não terem sido encontradas bases de imagens semelhantes à utilizada neste trabalho, a comparação com os trabalhos [20] e [4], que utilizam bases de imagens recentes, pode ilustrar o potencial da abordagem de histograma

de pontos de interesse, mesmo sob ruído elevado e baixa qualidade das imagens. Os trabalhos citados apresentam para a classe edifício taxas de acerto de 92,0% e 77,6%, respectivamente.

6 Conclusões

Este trabalho apresentou a aplicação da abordagem de histograma de pontos de interesse a imagens de fotografias históricas com o objetivo de classificá-las como contendo ou não edifícios. Os experimentos executados em um conjunto de imagens do acervo do APM mostraram que a utilização de algoritmos em uma base de imagens históricas não é uma tarefa trivial, pois apresentam variados níveis de degradação. Os resultados experimentais indicam o potencial do método proposto para a tarefa de detecção de edifícios, mesmo com a baixa qualidade visual das imagens utilizadas.

O desenvolvimento de técnicas que melhorem a indexação e o acesso ao material histórico do APM são importantes para valorização do acervo digital da instituição. Alguns exemplos de trabalhos já realizados nesse acervo são a indexação de pessoas baseada na detecção e identificação facial [21], a identificação automática dos tipos de impressões fotográficas [22] e a recuperação por imagem-exemplo robusta a distorções [23]. Nesse contexto, este trabalho irá cooperar com o desenvolvimento da recuperação de imagens que contém edifícios, informação que também pode ser útil para separar fotografias de áreas rurais ou urbanas.

Como trabalhos futuros, pretende-se amostrar mais pontos de interesse nas imagens (conforme [12], um maior número de pontos de interesse detectados pode significar mais informação discriminativa para a classificação) e estender a abordagem para detecção de outros objetos, tais como placas e linhas férreas.

Referências

1. Bertholdo, F.A.R.: Técnicas de limiarização para melhorar a qualidade visual de documentos históricos. Dissertação de mestrado, Universidade Federal de Minas Gerais (2007)
2. Oliveira, C.J.S., de A. Araújo, A.: Levantamento dos projetos que envolvem bases de imagens digitais do Arquivo Público Mineiro. Relatório técnico RT.DCC.001/2004 (2004)
3. Kennedy, L.S., Chang, S.F., Kozintsev, I.V.: To search or to label?: predicting the performance of search-based automatic image classifiers. In: Proceedings of the ACM International Workshop on Multimedia Information Retrieval, New York, NY, USA, ACM (2006) 249–258
4. Iqbal, Q., Aggarwal, J.: Retrieval by classification of images containing large man-made objects using perceptual grouping. *Pattern Recognition Journal* **35**(7) (2002)
5. Li, Y., Shapiro, L.G.: Consistent line clusters for building recognition in cbir. In: Proceedings of the International Conference on Pattern Recognition. Volume 3., Washington, DC, USA, IEEE Computer Society (2002) 30952
6. Kumar, S., Hebert, M.: Man-made structure detection in natural images using a causal multiscale random field. In: Proceedings of the IEEE Computer Society

- Conference on Computer Vision and Pattern Recognition. Volume 1. (2003) 119–126
7. Yuan, X., Li, C.T.: Cbir approach to building image retrieval based on linear edge distribution. In: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, Los Alamitos, CA, USA, IEEE Computer Society (2006) 95
 8. Zhang, W., Košecká, J.: Hierarchical building recognition. *Image Vision Computing* **25**(5) (2007) 704–716
 9. Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: Learning object categories from google’s image search. (2005)
 10. Csurka, G., Dance, C., Willamowski, J., Fan, L., Bray, C.: Visual categorization with bags of keypoints. In: Proceedings of the Workshop on Statistical Learning in Computer Vision. (2004) 1–22
 11. Nowak, E., Jurie, F., Triggs, B.: Sampling strategies for bag-of-features image classification. In: Proceedings of the European Conference on Computer Vision, Springer (2006) 490–503
 12. Jiang, Y.G., Ngo, C.W., Yang, J.: Towards optimal bag-of-features for object categorization and semantic video retrieval. In: Proceedings of the International Conference on Image and Video Retrieval, New York, NY, USA, ACM (2007) 494–501
 13. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, IEEE Computer Society (2006) 2169–2178
 14. Perronnin, F.: Universal and adapted vocabularies for generic visual categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(7) (2008) 1243–1256
 15. Agarwal, S., Awan, A., Roth, D.: Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2004)
 16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60**(2) (2004) 91–110
 17. Forsyth, D.A., Ponce, J.: *Computer Vision: A Modern Approach*. Prentice Hall of India (2006)
 18. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Computing Surveys* **31**(3) (1999) 264–323
 19. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* **2**(2) (1998) 121–167
 20. Jurie, F., Triggs, B.: Creating efficient codebooks for visual recognition. In: Proceedings of the IEEE International Conference on Computer Vision. (2005)
 21. Lopes, A., Oliveira, C., de A. Araujo, A.: Face recognition aiding historical photographs indexing using a two-stage training scheme and an enhanced distance measure. In: Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing. (2008)
 22. Oliveira, C.J.S., Batista, N.C., de A. Araújo, A., de S. Pio, J.L.: Photographic prints identification. In: Proceedings of the International Conference on Systems, Signals and Image Processings and Semantic Multimodal Analysis of Digital Media, Budapest, Hungary (2006) 231–234
 23. Valle, E., Cord, M., Philipp-Foliguet, S.: Content-based retrieval of images for cultural institutions using local descriptors. In: *Geometric Modeling and Imaging—New Trends*. (2006)