Semantic Discriminative Projections for Image Retrieval

He-Ping Song, Qun-Sheng Yang, and Yin-Wei Zhan

Faculty of Computer, GuangDong University of Technology
510006 Guangzhou, P.R. China
hepingsong@gmail.com, jsjqsy@sina.com, ywzhan@gdut.edu.cn

Abstract. Subspace learning has attracted much attention in image retrieval. In this paper, we present a subspace learning approach called Semantic Discriminative Projections (SDP), which learns the semantic subspace through integrating the descriptive information and discriminative information. We first construct one graph to characterize the similarity of contented-based features, another to describe the semantic dissimilarity. Then we formulate constrained optimization problem with a penalized difference form. Therefore, we can avoid the singularity problem and get the optimal dimensionality while learning a semantic subspace. Furthermore, SDP may be conducted in the original space or in the reproducing kernel Hilbert space into which images are mapped. This gives rise to kernel SDP. We investigate extensive experiments to verify the effectiveness of our approach. Experimental results show that our approach achieves better retrieval performance than state-of-art methods.

1 Introduction

With the development of digital imaging technology and the popularity of World Wide Web, Gigabytes of images are generated every day. It is a challenge that manage effectively images visual content. Content Based Image Retrieval (CBIR) is receiving research interest for this purpose [1,2,3,4]. However, there are still many open issues to be solved. Firstly, the visual content such as color, shape, texture, is extracted from an image as feature vectors. The dimensionality of feature space is usually very high. It ranges from tens to hundreds of thousands in most cases. Traditional machine learning approaches fail to learn in such a high-dimensional feature space. This is the well-known curse of dimensionality. Secondly, the low-level image features used in CBIR are often visual characterized, but it doesn't exist the directly connection with high-level semantic concepts, i.e. so-called semantic gap.

To alleviate the open issues, more and more attention has been drawn on the dimensionality reduction techniques. ISOMAP [5], Locally Linear Embedding (LLE) [6] and Laplacian eigenmaps [7] usher in manifold learning, these algorithms discover the intrinsic structure and preserve the local or global property of training data. Tenenbaum et al. [5] uses geodesic distance instead of Euclidean

H. Li et al. (Eds.): AIRS 2008, LNCS 4993, pp. 34–43, 2008.

distance to estimate distance between points. Multidimensional Scaling [8] is applied to discover the embedding space. Saul et al. [6] assumes there are smooth local patches that could be approximately linear, meanwhile a point in the training space could be approximated by linear combination of its neighbors. The projected space minimizes the reconstruction error using neighborhood correlation. Laplacian eigenmaps preserves locality information and makes neighbors close to each other in the projected space. These algorithms are unsupervised and limited to a nonlinear map. It is hard to map entire data space to low-dimensional space.

Locality Preserving Projections (LPP) [9], and Local Discriminant Embedding (LDE) [10] are proposed to extend the nonlinear learning approaches. These algorithms are all motivated by Laplacian eigenmaps. He et al. [9] uses a neighborhood graph to characterize locality preserving property that nearest neighbors in the original space should be nearest neighbors in the projected space. LDE constructs two neighborhood graphs, one prevents neighbors from different category and another preserves the locality through the affinity matrix using neighborhood information. LPP and LDE are effectively used to map data in entire image space. But only one neighborhood graph is used to discover the intrinsic structure, and LLE doesn't utilize the label information. LDE only keeps the neighborhood images from different classes away. LPP and LDE need to compute the inverse matrix, suffering from the singularity problem.

Bridge low-level visual feature to the high-level semantic is a great challenge in CBIR. We use Laplacian to learn the images semantic subspace in order to achieve more discriminative image representation for CBIR. In our work, both visual similarity and semantic dissimilarity are applied to construct neighborhood graph since they not only contain the descriptive information of the unlabeled images but also the discriminative information of the labeled images utilized in learning. We introduce a penalty γ to formulate a constrained optimization problem in the difference form, so that the optimal projection can be found by eigenvalue decomposition. Information of conjunctive graphs is represented by a affinity matrix, and it is much more computationally efficient in time and storage than LPP and LDE. On the other hand,the learnt subspace can preserve both local geometry and relevance information. Previous works often neglect the singularity problem and the optimal dimensionality, but we will determine the optimal dimensionality and avoid the singularity problem simultaneously.

This paper is organized as follows. In section 2, we introduce the SDP approach, kernel trick is used to the nonlinear learning approach in section 3, followed by the experiment results are discussed in section 4, and lastly we conclude our paper in section 5.

2 Laplacian Based Subspace Learning

In this section, we introduce our learning approach for image retrieval which respects the local descriptive and discriminative information of the original image space.

Suppose n training images, $\{\mathrm{x}_i\}_{i=1}^{n} \in \mathbb{R}^D$. we can construct two graph. G^S denotes the semantic similarity via semantic label and G^V denotes the visual resemblance by exploring the neighborhood of each image in the geometric space. W^S and W^V denotes the affinity matrix of G^S and G^V respectively. W^S and W^V are computed as follows:

$$W_{ij}^S = \begin{cases} 1 & \mathrm{x}_i \ , \mathrm{x}_j \text{ share the same class label,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

$$W_{ij}^V = \begin{cases} 1 & \mathrm{x}_i \in k\text{-NN of } \mathrm{x}_j \text{ or } \mathrm{x}_j \in k\text{-NN of } \mathrm{x}_i, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Where $k-\mathrm{NN}(\cdot)$ denotes the k-nearest-neighbors.
We integrate two kinds of information:

$$\begin{aligned} W^- &= \overline{W^S} \\ W^+ &= W^S \wedge W^V \end{aligned} \tag{3}$$

Where "$\wedge$" denotes the Meet of two zero-one matrices.
We utilize the penalized difference form to formulate following constrained optimization problem.

$$P = \arg \max_{P^T P = I} \sum_{i,j} \|P^T \mathrm{x}_i - P^T \mathrm{x}_j\|^2 (W_{ij}^- - \gamma W_{ij}^+) \tag{4}$$

where γ is a penalized coefficient, the constraint $P^T P = I$ avoids trivial solution, and I is the $d \times d$ identity matrix, d is the reduced dimensionality.

The above formulation exhibits the implication that local neighbors with semantic dissimilarity should separate each other and different semantic classes are far away from each other; the images with similar semantic and visual content will be clustered together, preserving the intrinsic structure.
We rewrite (4) in the form of trace, and get the following formulation:

$$\begin{aligned} \mathcal{J} &= \sum_{i,j} \|P^T \mathrm{x}_i - P^T \mathrm{x}_j\|^2 (W_{ij}^- - \gamma W_{ij}^+) \\ &= \sum_{i,j} tr\{(P^T \mathrm{x}_i - P^T \mathrm{x}_j)(P^T \mathrm{x}_i - P^T \mathrm{x}_j)^T\}(W_{ij}^- - \gamma W_{ij}^+) \\ &= \sum_{i,j} tr\{P^T (\mathrm{x}_i - \mathrm{x}_j)(\mathrm{x}_i - \mathrm{x}_j)^T P\}(W_{ij}^- - \gamma W_{ij}^+) \\ &= tr\{P^T \sum_{i,j} (\mathrm{x}_i - \mathrm{x}_j)(W_{ij}^- - \gamma W_{ij}^+)(\mathrm{x}_i - \mathrm{x}_j)^T P\} \\ &= 2tr\{P^T[(XD^- X^T - XW^- X^T) - \gamma(XD^+ X^T - XW^+ X^T)]P\} \\ &= 2tr\{P^T X(L^- - \gamma L^+)X^T P\} \end{aligned} \tag{5}$$

Where $L^- = D^- - W^-$, and D^- is a diagonal matrix with $D_{ii}^- = \sum_j W_{ij}^-$. Analogously, $L^+ = D^+ - W^+$ with $D_{ii}^+ = \sum_j W_{ij}^+$. Thus the optimization problem can be formulated as:

$$\begin{aligned} P = & \quad \arg \max_{P^T P=I} tr\{P^T X(L^- - \gamma L^+)X^T P\} \\ \Rightarrow P = & \quad \mathrm{eig}(X(L^- - \gamma L^+)X^T) \end{aligned} \tag{6}$$

We take no dimensionality reduction as the baseline, therefore (5) could be zero. We could get a positive scalar γ when the dimensionality is reduced [11]. Then we have:

$$\mathcal{J} = tr(X(L^- - \gamma L^+)X^T) = 0 \quad \Longrightarrow \quad \gamma = \frac{trL^-}{trL^+} \tag{7}$$

It is obviously that $X(L^- - \gamma L^+)X^T$ is a $D \times D$, sparse, symmetric and positive semidefinite matrix. According to the result of Rayleigh quotient, the optimization problem can be calculated by eigenvalue decomposition.

Denote $P \in \mathbb{R}^{D\times d}$ by $P=[p_1, p_2, \cdots, p_d]$, where $p_i(i = 1, \cdots, d)$ is the d largest eigenvectors corresponding to the d largest eigenvalues of $X(L^- - \gamma L^+)X^T$. $\sum \lambda_i$ is the optimal value of the above optimization problem, where $\lambda_i(i = 1, \cdots, d)$ are the d largest eigenvalues. d is the number of positive eigenvalues and $\mathcal{J}$ reaches the maximum.

We can see that the singularity problem in LPP, LDE does not exist in our approach, meanwhile we find the optimal dimensionality.

To get returns for the query in image retrieval, we project any test image $\overline{\mathrm{x}} \in \mathbb{R}^D$ to $\mathbb{R}^d$ via $\overline{\mathrm{y}} = P^T\overline{\mathrm{x}}$ and will find the nearest neighbors of Euclidean distances. Those images corresponding to the nearest neighbors will be the top-ranking returns.

3 Kernel SDP

As the kernel trick [12] successfully applied to Kernel LPP [13], Kernel LDE [10], we generalize SDP to kernel SDP, in which kernel transformation is applied to handle nonlinear data.

Denote $\Phi : \mathbb{R}^D \to \mathcal{F}$ is a nonlinear mapping, so the image feature vectors in the high-dimensionality feature space is denoted as $\Phi(\mathrm{x}_i), (i = 1, \cdots, n)$. The inner product in $\mathcal{F}$ can be computed by the kernel function. we specify the RBF kernel $k(\mathrm{x}_i, \mathrm{x}_j) = \Phi(\mathrm{x}_i)^T \Phi(\mathrm{x}_j) = exp(-\|\mathrm{x}_i - \mathrm{x}_j\|^2/t)$ in our work. we find the optimal projection $V, (\mathrm{v}_i, i = 1, \cdots, d)$ in $\mathcal{F}$, the v_i is spanned by $\Phi(\mathrm{x}_i), i = 1, \cdots, n$, and assume v_i is the linear combination of $\Phi(\mathrm{x}_i)$ in the projected space $\mathcal{F}$:

$$\mathrm{v}_i = \sum_{j=1}^{n} \alpha_{ij} \Phi(\mathrm{x}_j) = \Phi(\mathrm{X})\alpha_i \tag{8}$$

we have:

$$(\mathrm{V}^T \Phi(\mathrm{X}))_{ij} = \mathrm{v}_i^T \Phi(\mathrm{x}_j) = (AK)_{ij} \tag{9}$$

where $A = [\alpha_i, \cdots, \alpha_n]^T$ denotes the linear coefficient in vector, $K_{ij} = k(\mathrm{x}_i, \mathrm{x}_j)$ is kernel matrix.

Replacing X with $\Phi(\mathrm{X})$, we rewrite (5), and consider the kernel-based optimization problem:

$$\begin{aligned} A = &\quad \arg \max_{A^T A = I} tr\{AK(L^- - \gamma L^+)KA\} \\ \Rightarrow A = &\quad \mathrm{eig}(K(L^- - \gamma L^+)K) \end{aligned} \tag{10}$$

where the constraint $A^T A = I$ avoids trivial solution, I is the $d \times d$ identity matrix.

Analogously according to the result of Rayleigh quotient, the optimization problem can be calculated by eigenvalue decomposition. We select the d largest eigenvalues of $K(L^- - \gamma L^+)K$, where d is the number of positive eigenvalues. Our approach doesn't suffer from the singularity problem, and get the optimal dimensionality.

To get returns for the query in image retrieval, we map any test image $\overline{\mathrm{x}}$ to by $\overline{\mathrm{y}} = V^T \overline{\mathrm{x}}$ with the ith dimensionality computed by $\overline{\mathrm{y}}_i = \mathrm{v}_i^T \overline{\mathrm{x}} = \sum_{i=1}^{n} \alpha_{ij} k(\mathrm{x}_j, \overline{\mathrm{x}})$, and find the nearest neighbors of Euclidean distances. Those images corresponding to the nearest neighbors will be the top-ranking returns.

4 Experimental Results

In this section, we experimentally evaluate the performance of SDP on COREL dataset and compare with LPP, LDE in order to demonstrate effectiveness of our approach for image retrieval. The COREL dataset is widely used in many CBIR systems [18]. In our experiments, the dataset consists of 2500 color images, including 25 categories, each category contains 100 samples. Those images in the same category share the same semantic concept, but they have their individual varieties. Images from the same category are considered relevant, and otherwise irrelevant.

In our experiments, we only consider these queries which don't exist in the training images. Five-fold cross validation is used to evaluate the retrieval performance. We pick one set as the testing images, and leave the other four sets as the training images. Table 1 shows the features of which the dimensionality is 145.

Precision-Recall curve (PRC) is widely used as a performance evaluation metrics for image retrieval [19]. In many cases, PRC is overlapped in high recall,

Table 1. Image features used in experiment

Name	Description &Dimension
Normalized Color Histogram [14]	HSV Space 64 dimension
Normalized LBP Histogram [15]	HSV Space 54 dimension
Normalized Color Moments [16]	HSV Space 9 dimension
Normalized Wavelet Moments [17]	RGB Space 18 dimension

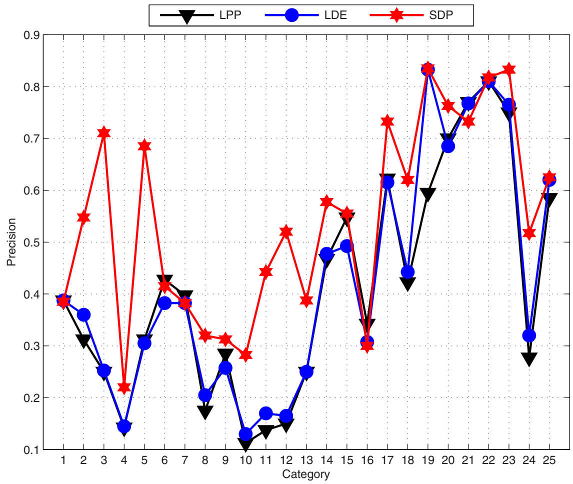

Fig. 1. The plots of precision versus category for SDP, LDE, and LPP

moreover, it is unreasonable to calculate the mean average precision among different categories. We alternatively adopt the precision-category. Given a specified number N, we define the precision as following:

$$precision = \frac{\text{The number of relevant images in top } N \text{ returns}}{N}$$

Users are usually interested in the first screen results just like Google$^{\text{TM}}$ Image. We have N=20 in our experiments.

4.1 Comparisons with LPP, LDE

Model Selection is very important in many subspace learning algorithms. In our experiments, it is not very sensitive to tuning parameter k. we set $k = 10$. We adopt Gaussian heat kernel to compute the W_{ij}, $W_{ij} = exp(-\|x_i - x_j\|^2/c)$, where c is a positive scalar. the aforementioned W_{ij} is superior to 0/1 and is not sensitive to c.

In this experiment, we compare SDP with LPP, LDE. LPP and LDE have the limitation of singularity problem due to compute inverse matrix. Both LPP and LDE adopt PCA to overcome the limitation, retaining the 98% principal components [13]. The optimal dimensionality of SDP is 64, as shown in Fig. 1. SDP achieves much better retrieval performance than other approaches. As a matter of fact, we gain much more discriminating image representation by SDP. Next, we give the experiment result of Kernel SDP in Fig. 2, the optimal dimensionality of Kernel SDP is 145, Except for the category 15, Kernel LDE performs marginally better than Kernel SDP. We can conclude that Kernel SDP outperforms other approaches.

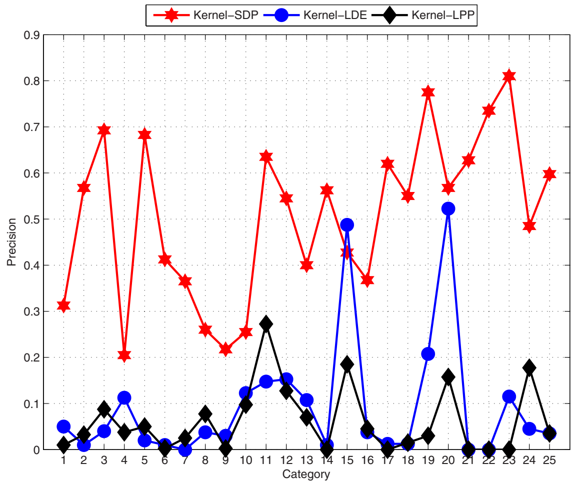

Fig. 2. The plots of precision versus category for Kernel SDP, Kernel LDE, and Kernel LPP

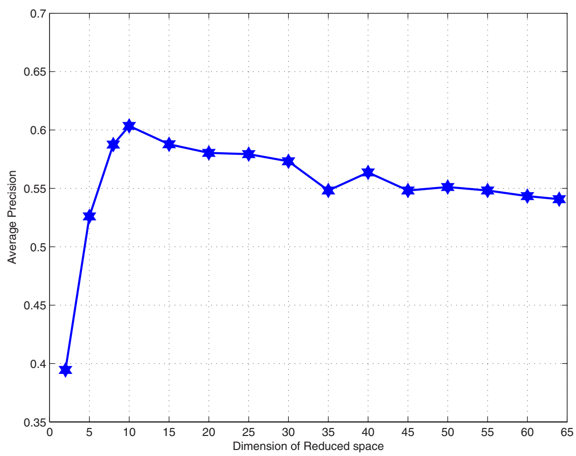

Fig. 3. Average Precision in reduced space with different dimension for SDP

4.2 Reduced Dimensionality

Even though our approach can determine the optimal dimensionality, the dimensionality of reduced space is important tradeoff between retrieval precision and the computational complexity. In this experiment, we investigate the relation between dimension and precision. The precision of SDP reaches its peak at 10 dimensions in Fig. 3. The precision of Kernel SDP converges rapidly from 2 to 10 dimensions, and then achieves reliable results from 10 to 64 dimensions. As shown in Fig. 4, The precision reaches its peak at 30 dimensions, converges

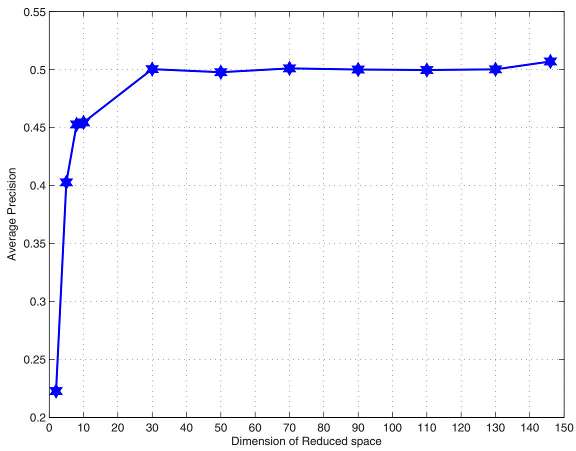

Fig. 4. Average Precision in reduced space with different dimension for Kernel SDP

rapidly from 2 to 30 dimensions, and then achieves smooth results from 30 to 145 dimensions. As shown in Fig. 3 and Fig. 4, we can gain the lower dimensionality while the precision is stable, even relatively higher.

5 Conclusions and Future Work

We have introduced a subspace learning approach SDP for image retrieval. The image structure is approximated by the adjacent graph integrating descriptive and discriminative information of the original image space. SDP focuses on the improvement of the discriminative performance of image representation. As previous work neglect the singularity problem and optimal dimensionality, SDP avoid the singularity problem and determine the optimal dimensionality simultaneously. Moreover, we extend our approach and present kernel SDP. Experimental results have revealed the effectiveness of our approach.

Owing to the effectiveness of SDP, further work will be investigated as following:

1. Feature selection is an open issue in CBIR. In this work, we only adopt global features, which only describe the respective overall statistics for a whole image. A great of previous work has applied local features to CBIR [20,21,22]. We augment local features to improve retrieval performance.
2. Representing an image as a matrix intrinsically, and extending the subspace learning algorithm with tensor representation [23,24].
3. Utilizing the user's interaction, A possible extension of our work is to incorporate the feedback information to update the affinity matrix, might achieve higher precision [25,26].

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 60572078) and by Guangdong Natural Science Foundation (Grant No. 05006349).

References

1. Rui, Y., Huang, T.S., Chang, S.F.: Image retrieval: current techniques, promising directions and open issues. Journal of Visual Communication and Image Representation 10(4), 39–62 (1999)
2. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12), 1349–1380 (2000)
3. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys 39(20), 1–77 (2007)
4. Liu, Y., Zhang, D.S., Lu, G.J., Ma, W.Y.: Asurvey of content-based image retrieval with high-level semantics. Pattern Recognition 40(1), 262–282 (2007)
5. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290(5500), 2319–2323 (2000)
6. Saul, L.K., Roweis, S.T.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
7. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. Advances in Neural Information Processing Systems 14, 585–591 (2002)
8. Cox, T., Cox, M.: Multidimensional Scalling (1994)
9. He, X.F.: Laplacian Eigenmap for Image Retrieval. Master's thesis, Computer Science Department, The University of Chicago (2002)
10. Chen, H.T., Chang, H.W., Liu, T.L.: Local discriminant embedding and its variants. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 846–853 (2005)
11. Nie, F.P., Xiang, S.M., Song, Y.Q., Zhang, C.S.: Optimal Dimensionality Discriminant Analysis and Its Application to Image Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, June17-22, pp. 1–8 (2007)
12. Scholkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge (2001)
13. He, X.F.: Locality Preserving Projections. PhD thesis, Computer Science Department, the University of Chicago (2005)
14. Swain, M.J., Ballard, D.H.: Color indexing. International Journal of Computer Vision 7(1), 11–32 (1991)
15. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
16. Stricker, M., Orengo, M.: Similarity of color images. In: Proceedings SPIE Storage and Retrieval for Image and Video Databases, San Jose, CA, USA, vol. 2420, pp. 381–392 (1995)
17. Smith, J.R., Chang, S.F.: Transform features for texture classification and discrimination inlarge image databases. In: IEEE International Conference Image Processing, vol. 3, pp. 407–411 (1994)

18. Muller, H., Marchand-Maillet, S., Pun, T.: The truth about corel-evaluation in image retrieval. In: Proceedings of the International Conference on Image and Video Retrieval, pp. 38–49. Springer, Heidelberg (2002)
19. Smith, J.R., Chang, S.F.: Tools and techniques for color image retrieval. In: Storage & Retrieval for Image and Video Databases IV, vol. 2670, pp. 426–437 (1996)
20. Tuytelaars, T., Van Gool, L., et al.: Content-based image retrieval based on local affinely invariant regions. In: Int. Conf. on Visual Information Systems, pp. 493–500. Springer, Heidelberg (1999)
21. Jing, F., Li, M., Zhang, H.J., Zhang, B.: An effective region-based image retrieval framework. In: Proceedings of the tenth ACM international conference on Multimedia, pp. 456–465. ACM Press, New York (2002)
22. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
23. Xia, J., Yeung, D.Y., Dai, G.: Local Discriminant Embedding with Tensor representation. In: IEEE International Conference on Image Processing, pp. 929–932 (2006)
24. He, X.F., Cai, D., Niyogi, P.: Tensor subspace analysis. Advances in Neural Information Processing Systems 18 (2005)
25. He, X.F.: Incremental semi-supervised subspace learning for image retrieval. In: Proceedings of the 12th annual ACM international conference on Multimedia, pp. 2–8. ACM Press, New York (2004)
26. Lin, Y.Y., Liu, T.L., Chen, H.T.: Semantic manifold learning for image retrieval. In: Proceedings of the 13th annual ACM international conference on Multimedia, pp. 249–258. ACM Press, New York (2005)