A Semantic Subspace Learning Method to Exploit Relevance Feedback Log Data for Image Retrieval

Lining Zhang
School of EEE
Nanyang Technological University
Singapore, 639798
Email: zhan0327@ntu.edu.sg

Lipo Wang
School of EEE
Nanyang Technological University
Singapore, 639798
Email: elpwang@ntu.edu.sg

Weisi Lin
School of Computer Engineering
Nanyang Technological University
Singapore, 639798
Email: wslin@ntu.edu.sg

***Abstract*—Conventional content-based image retrieval (CBIR) systems with the Euclidean distance metric in a high-dimensional visual feature space usually cannot achieve satisfactory performance due to the semantic gap. Relevance feedback (RF) has been introduced as a powerful tool to involve the user in the system to improve the performance of CBIR. Despite the success, an on-line learning task can be tedious and boring for the user. Various schemes have been proposed to exploit the RF log data to further enhance the performance of CBIR. In this paper, we propose a semantic subspace learning (SSL) method to exploit the RF log data with contextual information for an image retrieval task. Different from conventional subspace learning approaches, our method can directly learn a semantic concept subspace from the RF log data with contextual information without using any class label information. We show that the performance of the image retrieval task can be significantly improved in the low-dimensional semantic concept subspace. Extensive experiments on a real-world image database demonstrate the effectiveness of the proposed scheme in improving the performance of CBIR by exploiting the RF log data.**

I. Introduction

Content-based image retrieval (CBIR) has attracted much attention during the past decades [1], [2], [3]. Conventional CBIR systems usually adopt the Euclidean distance metric in a high-dimensional low-level visual feature space to measure the similarity between the query image and the images in the database [1], [2], [3]. However, the Euclidean distance metric in the original high-dimensional space is often not very effective due to the semantic gap between low-level visual features and high-level semantic concepts.

To narrow down this semantic gap, relevance feedback (RF) has been widely designed as a powerful tool to involve the user in the system by letting the user label semantically similar and dissimilar images with the query image, and thus to define a more effective similarity metric for image retrieval [4], [5], [6]. During the last decade, various RF approaches have been developed based on different assumptions for the positive and negative feedback samples. Despite the success, an on-line learning task is usually boring and tedious for the user in RF. Given the difficulties in capturing the user preferences, multiple rounds of RF are usually required to achieve satisfactory results for an image retrieval task, which will significantly limit the capability of conventional RF methods for real-world applications [4].

Beyond conventional RF methods, several promising approaches have been emerging to attack this semantic gap in the CBIR community. For instance, image annotation techniques intend to directly acquire the semantic concepts from the low-level visual features of an image [7]. However, major challenges still remain regarding these image annotation techniques. Recently, a large number of studies have attempted to narrow down this semantic gap by exploiting the RF log data with contextual information [8], [9], [10], [11]. In these studies, the system can accumulate the RF information provided by a number of users in image retrieval, which can be regarded as the RF log data with contextual information (e.g. similar and dissimilar constraints). As a consequence, different from conventional CBIR tasks, besides low-level visual features, each image in the log database can also be associated with a set of similar and dissimilar pairwise constraints judged by users. From a long-term perspective, the RF log data with contextual information is an important and useful resource to further enhance the image retrieval task. This new paradigm of image retrieval can alleviate the aforementioned major overhead on users in image retrieval by leveraging the log data accumulated by conventional CBIR systems over a long period of time.

During the past several years, various methods have been widely conducted to exploit the RF log data with contextual information; however, little work has been made to explicitly evaluate subspace learning approaches to narrow down this semantic gap between low-level visual features and high-level semantic concepts by exploiting the RF log data for image retrieval. Let us first use an easy example to show the importance of semantic subspaces in measuring the similarity between images in the CBIR community. Figs. 1 (a) and (b) show two easy images with different semantic concepts (e.g., color, texture, shape, etc) and their associated high-dimensional low-level visual features, respectively. With an assumption that different semantic concepts live in different subspaces and each image can live in many different semantic concept subspaces [4], it is not appropriate to measure the similarity between the two images based on the Euclidean distance metric in the original high-dimensional multiple semantic concept space (e.g., color, texture, and shape). This is mainly because there are many different semantic concept subspaces in the original high-dimensional visual feature space and the

978-1-4673-5895-8/13/$31.00

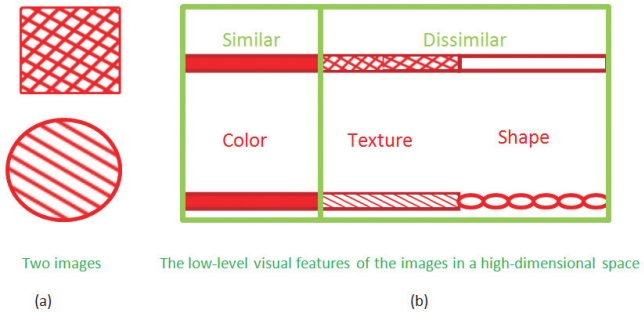


Fig. 1. Two synthetic images and the associated low-level visual features in a high dimensional space, for concept illustration.

two images are only similar in one semantic concept subspace (e.g. color) but different with each other in other semantic concept subspaces (e.g., texture and shape). Therefore, it is more reasonable to measure the similarity between the two images in a low-dimensional semantic subspace (e.g., color) than in the original high-dimensional visual feature space. Subspace learning approaches are powerful tools for various tasks in computer vision. Most of the conventional subspace learning approaches normally need to acquire explicit class label information. However, the explicit class information for each image is usually expensive to obtain in many real-world applications. Compared with the explicit class label information of each image, it is much easier to acquire the contextual information (i.e., similar and dissimilar pairwise constraints) when the RF log data accumulated by conventional CBIR systems are available [8], [9], [12], [10], [11]. Therefore, it is more attractive and useful to directly learn a low-dimensional semantic subspace from the RF log data without using any class label information. Recently, learning distance metrics with contextual information has been actively studied in both the CBIR community and the machine learning community [13], [14], [11]. Despite the active research efforts during the past a few years, most of these approaches in this category have involved a high computational burden when dealing with high-dimensional images and also cannot give the explicit image semantic representations, which is not appropriate and will significantly limit their potential applications to the RF log data with contextual information.

In this paper, we propose a novel semantic subspace learning (SSL) method to attack the semantic gap between low-level visual features and high-level semantic concepts by exploiting the RF log data with contextual information for image retrieval. The proposed SSL method can effectively learn a reliable semantic subspace from the RF log data by incorporating the discriminative information and the geometric information together. Compared with the previous distance metric learning with contextual information, our method can also learn a distance metric but perform more effectively when dealing with high-dimensional images.

This paper is organized as follows: the SSL method is detailed in Section II; in Section III, we first give the experimental results, and then show some analysis to the important parameters in SSL; Section IV concludes this paper.

II. Semantic Subspace Learning with Contextual Information

In this paper, we aim to find a semantic subspace to reflect the similar relation between a pair of images and reduce the semantic gap by exploiting the log data judged by users in RF iterations, and thus to enhance the performance of CBIR. We use a linear mapping matrix U to approximate this semantic concept subspace and then the images in this subspace can be represented as $Y = U^T X = [y_1, y_2, \ldots, y_n] \in R^{l\times n}(l < h)$ with $y_i \in R^l$ for image $x_i \in R^h$. Therefore, in this reduced semantic concept subspace, improved retrieval performance is expected.

In this subsection, we present a SSL with contextual information method to learn such a mapping matrix U by exploiting the log data for an image retrieval task. Especially, the SSL method can effectively integrate the discriminative information of labeled log images, the geometric information of labeled log images together.

Given images with contextual information, a popular principle for learning a distance metric is to minimize the distances between samples with similar constraints and to maximize the distances between samples with dissimilar constraints simultaneously, which can be referred to as a min-max principle [13]. Following this principle, we try to minimize the average squared distances between each image x_i and its associated k_1 nearest images with similar constraints; meanwhile, we also try to maximize the average squared distances between each image x_i and its associated k_2 nearest images with dissimilar constraints. Especially, for the new semantic representations of the images, we expect that the loss function between k_1 nearest images with similar constraints and k_2 nearest images with dissimilar constraints will be minimized as much as possible, i.e.,

$$\begin{aligned} f(y_i) &= \min \sum_{j=1}^{k_1} ||y_i - y_{i,j}||^2 \frac{1}{k_1} - \gamma \sum_{j=k_1+1}^{k_1+k_2} ||y_i - y_{i,j}||^2 \frac{1}{k_2} \\ &= \min \sum_{j=1}^{k_1+k_2} h_{i,j} ||y_i - y_{i,j}||^2 \end{aligned} \tag{1}$$

where the parameter γ is used to balance the two squared distances; and the weight coefficient $h_{i,j}$ encodes both the similar and dissimilar constraints, i.e.,

$$h_{i,j} = \begin{cases} 1/k_1, \text{if} x_i \text{and} x_j \text{aresimilar}, \\ -\gamma/k_2, \text{if} x_i \text{and} x_j \text{aredissimilar}. \end{cases} \tag{2}$$

Although the discriminative loss information for each labeled log image can capture the discriminative informative well, it is empirically known that the geometric information of log images can help to find the intrinsic semantic concept subspace. In particular, for each image x_i, we assume that each image can be reconstructed through the nearest samples with similar constraints [15]. Thus, x_i can be linearly reconstructed from its nearest k_1 samples $x_{i,j}, j = 1, \ldots, k_1$ as:

$$x_i = \sum_{j=1}^{k_1} c_{i,j} x_j + \varepsilon_i \tag{3}$$

where ε_i is the reconstruction error for x_i and ε_i is obtained through minimizing $||\varepsilon_i||^2$, i.e.,

$$c_{i,j} = \arg\min_{c_{i,j}} ||\varepsilon_i||^2 = \arg\min_{c_{i,j}} ||x_i - \sum_{i=1}^{k_1} c_{i,j} x_{i,j}||^2 \tag{4}$$

By imposing $\sum_{j=1}^{k_1} c_{i,j} = 1$ on the above function, we have $c_{i,j} = \sum_{p=1}^{k_1} G_{jp}^{-1} / (\sum_{s=1}^{k_1} \sum_{t=1}^{k_1} G_{st}^{-1})$ with a local gram matrix $G_{jp} = (x_i - x_{i_j})^T (x_i - x_{i_j})$ as described in [15].

In SSL, $c_{i,j}$ reconstructs x_i from y_i in the low-dimensional space, so we have

$$g(y_i) = \min ||y_i - \sum_{i=1}^{k_1} c_{i,j} y_{i,j}||^2 \tag{5}$$

By combining the discriminative loss function and the geometric regularization term together, we have

$$\begin{aligned} y_i &= \arg\min_{y_i, 1 \le i \le n} \sum_{i=1}^{n} f(y_i) + \beta_1 \sum_{i=1}^{n} g(y_i) \\ &= \arg\min_{y_i, 1 \le i \le n} \sum_{i=1}^{n} \sum_{j=1}^{k_1+k_2} h_{i,j} ||y_i - y_{i,j}||^2 + \beta_1 \sum_{i=1}^{n} ||y_i - \sum_{i=1}^{k_1} c_{i,j} y_{i,j}|| \end{aligned} \tag{6}$$

Based on a series of matrix operations, we can obtain the linear mapping matrix U according to

$$\begin{aligned} U^* &= \arg\min_{U \in R^{h \times l}} \left\{ \begin{array}{l} \sum_{i=1}^{n} \sum_{j=1}^{k_1+k_2} h_{i,j} ||U^T x_i - U^T x_{i,j}||^2 \\ +\beta_1 \sum_{i=1}^{n} ||U^T x_i - \sum_{i=1}^{k_1} c_{i,j} U^T x_{i,j}||^2 \end{array} \right\} \\ &= \arg\min_{U \in R^{h \times l}} \left\{ \begin{array}{l} tr\left(U^T X (D_h - {W_h}^T)(D_h - {W_h}^T)^T X^T U\right) \\ +\beta_1 tr\left(U^T X (I - {W_c}^T)(I - {W_c}^T)^T X^T U\right) \end{array} \right\} \\ &= \arg\min_{U \in R^{h \times l}} tr\left(U^T X (F + \beta_1 G) X^T U\right) \end{aligned} \tag{7}$$

where $W_h = [h_{i,j}] \in R^{n \times n}, W_c = [c_{i,j}] \in R^{n \times n}$, and $D_h \in R^{n \times n}$ is a diagonal matrix and its ith matrix is $\sum_{j=1}^{n} h_{i,j}$; F encodes the discriminative information and $F = (D_h - W_h^T)(D_h - W_h^T)^T$; G encodes the geometric information and $G = (I - W_c^T)(I - W_c^T)^T$; $\beta_1 > 0$ is the tuning parameter, which is used to trade off the contributions of these two different terms.

By imposing the constraint $U^T U = I$, we can avoid the trivial solutions of this problem and solve this problem by conducting a standard Eigenvalue decomposition, and the mapping matrix U is formed by the l eigenvectors associated with the first l smallest eigenvalues.

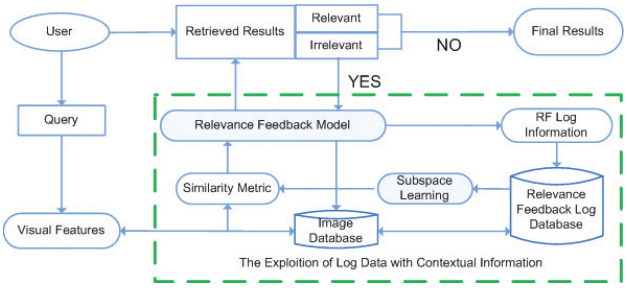


Fig. 2. The flowchart of our CBIR system

III. Experimental Results

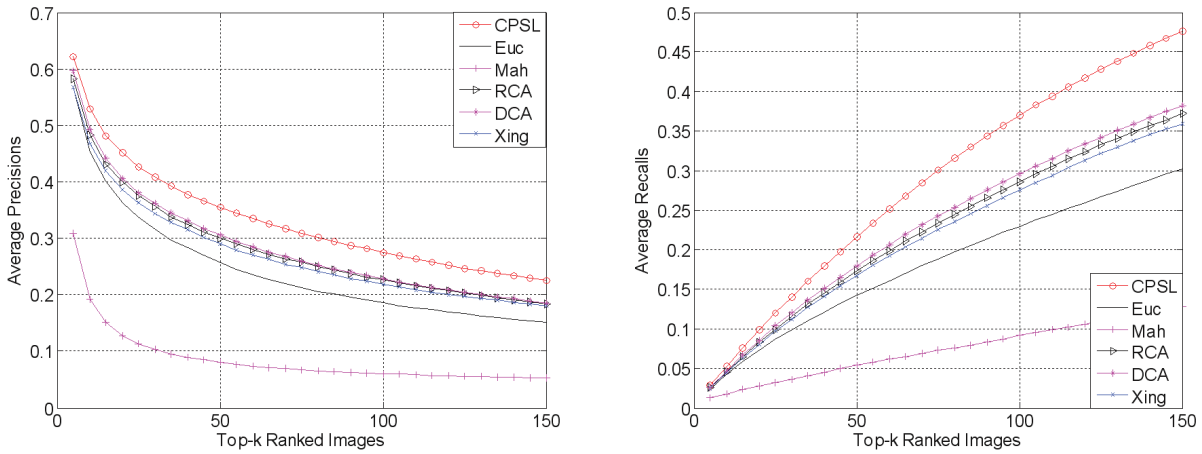
In this section, we evaluate the performance of the proposed method in exploiting the log data with contextual informaiton for an image retrieval task. Firstly, we investigate the effectiveness of the proposed method by comparing it with some representative metric learning methods. And then, we show the sensitivity of important parameters of the proposed method.

A. Experimental Log Data with Contextual Information

Collecting the log data with contextual information is a very important step for performance evaluation of the proposed method. However, to our best knowledge, there is no datasets for the application of exploiting the log data with contextual information for image retrieval. It is not difficult to build a log database based on the existing real-world database, e.g., the Corel image database. Here, we first randomly select 30 classes according to the ground truth of the images from the Corel image database to form a log database, which contains 3,568 real-world images. And then, different from supervised learning, we divide each class of the database into two groups with an equal size. Therefore, the log data database comprises 60 groups with 30 different concepts. We randomly select 10 images uniformly from each group, and thus we can gather a labeled log database. The similar constraints are imposed on the images within the same group, while the dissimilar constraints are imposed on the images with different concepts. Finally, we can obtain a log database with 600 images. To represent images, we utilize the color histogram [16], Webber's law descriptors [17], and the edge directional histogram [18] from Y component in YCrCb space, each of which can describe the semantic content of images from different aspects. All of these features are combined into a feature vector, which results in a vector with 510 values, and then we normalize each feature to a normal distribution. Fig. 2 shows the flowchart of our CBIR system.

B. Performance Evaluation by Exploiting RF Log Data

In this subsection, we aim to examine if the proposed method is comparable with or better than representative metrics learning with contextual information. We compare the proposed method with two major distance metrics (i.e., the Euclidean distance metric and the Mahalanobis distance metric), and three representative distance metrics learning



(a) (b)

Fig. 3. (a) AP curves and (b) AR curves of the six compared methods on the log database.

with contextual information (i.e., relevant component analysis (RCA) [14], discriminative component analysis (DCA) [19], and Xing [13]). Moreover, we do not compare the proposed method with supervised learning techniques since they often require explicit class label information, which are not suitable to exploit the log data with contextual information. Parameters in each method were determined to achieve its best performance in this paper. The parameter sensitivity of the proposed method will be carefully analyzed in the next subsection. All of the compared algorithms are implemented on the log database as given in subsection, i.e., the log database with 600 images. For the two parameters k_1 and k_2, we set them as 4 in experiments. In experiments, 500 query images are first randomly selected from the database and the image retrieval procedure is automatically conducted. We use average precision (AP) and average recall (AR) to evaluate the performance of the compared algorithms.

Fig.3 shows the experimental results of the compared algorithms on the log database. As we can see from Fig.3, directly using the Euclidean distance metric in a high-dimensional visual feature space is not appropriate due to the semantic gap. Moreover, a simple Mahalanobis distance metric does not outperform the Euclidean distance metric because of the matrix singular problem, i.e., the number of the log images is much less than the dimension of the visual feature space. And then, all of the metric learning methods (i.e., RCA, DCA, Xing, and SSL) can perform better than the Euclidean distance metric by exploiting the log data with pairwise constraints. In experiments, the optimal distance metric learned by RCA is computed as the inverse of the average covariance matrix of the chunklets. Similar to the Mahalanobis distance metric, the RCA will also encounter a singular covariance matrix when dealing with high-dimensional images. In this work, the RCA is preceded by constraints based LDA, which can reduce the dimension to that of SSL. By doing this, we notice that the RCA can show much better performance than the Euclidean distance metric by exploiting the log data with similar pairwise constraints. The DCA can effectively utilize the dissimilar constraints and was formulated into a trace ratio problem. However, much discriminative information in the null space of the dissimilar scatter has been discarded in solving this problem. Although the DCA incorporates the dissimilar pairwise constraints into the RCA, the performance of DCA has been significantly degraded due to the numerical computation in handing the trace ratio. The SSL can learn a distance metric by resorting to the mapping matrix U and solve this problem with a standard Eigenvalue decomposition, which is much effective and efficient when dealing with high dimensional images and also does not encounter the said problem of numerical computation. From the results, we can see that the proposed SSL significantly outperforms the two major existing distance metrics and three compared metric learning approaches for overall evaluation.

C. Parameter Sensitivity

In this subsection, we will study the parameter sensitivity of the SSL method in exploiting the log data with contextual information for an image retrieval task. The analysis is performed based on the experiments conducted on the log database. In experiments, we analyze the trade-off parameter β_1 for balancing the loss function and the regularization term, and the dimension of the projected features for the SSL method. Firstly, 500 query images are randomly selected from the database, and then the image retrieval procedure is automatically conducted. The APs in top 50 results are utilized for overall performance evaluation.

1) Evaluation on the geometric regularization parameter β_1: Empirically, geometric information is useful for finding a semantic subspace. In this part, we investigate the influence of the trade-off parameter β_1 for balancing the discriminative loss function and the geometric regularization term. A small

β_1 reflects less importance of separating the dissimilar images from the similar ones, i.e., the SSL method focuses on the discriminative information but ignores the geometric information. Fig.4 shows the performance of SSL with different β_1, from which we can notice that when β_1 is small, e.g., $\beta_1 = 0$, the performance is unsatisfactory. This is mainly because in this situation the discriminative information is preserved while important geometric information in labeled log images with similar pairwise constraints is less considered. The performance of SSL increases when β_1 increases and reaches the optimal value, i.e., 7. And then, APs decrease when β_1 is larger than this best setup, in which case the geometric information dominates the objective function and the important discriminative information is ignored.

Consequently, both the discriminative loss function and the geometric regularization term can reflect the important information contained in log data from different aspects for complimentary. A suitable combination of them is essential to achieve good performance for SSL.

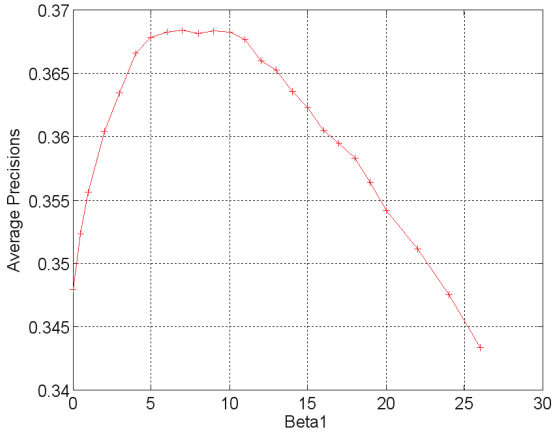


Fig. 4. Performance of SSL with different β_1 for the log database.

2) Evaluation on the dimension of projected subspaces: Different from distance metric learning methods, the proposed method intends to learn a mapping matrix, which can find a low-dimensional subspace from the the original high-dimensional space. To find out an appropriate dimension of the projected semantic concept subspace, we have investigated the influence of the dimension in the following experiments. In Fig.5 we have shown the performance of the SSL method with features projected onto the subspaces with different dimensions. From Fig.5, we can notice that when the projected dimension is too low, the reduced subspace is insufficient to encode the semantic concepts of images, which makes the performance of the system poor. When the dimension equals or closes to that of the original high-dimensional space, no or less benefit can be obtained from semantic concept subspace learning. From the experimental results, we can notice that the proposed method achieves its best performance with the dimension of 23 for this log database. Moreover, low-dimensional data can lead to a less computational burden than high-dimensional data for an image retrieval task.

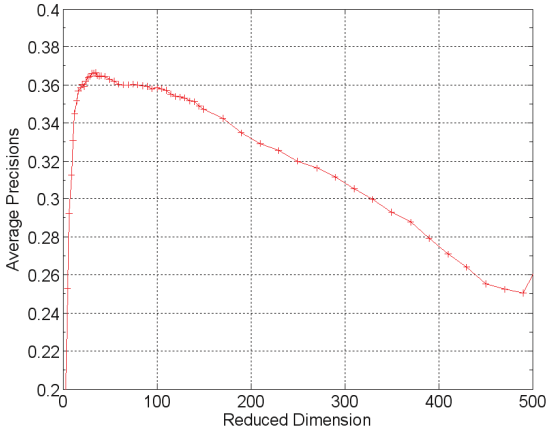


Fig. 5. Performance of SSL with features projected onto the subspaces with different dimensions for the log database.

IV. Conclusion

Subspace learning, a typical computational intelligence technique, sheds light on various tasks from computer vision to data mining. In light of the idea that different semantic concepts live in different subspaces and each image can live in many different semantic concept subspaces, in this paper, we have proposed to learn an effective semantic concept subspace from the RF log data with contextual information for an image retrieval task. The proposed method can directly learn a reliable semantic concept subspace from the log data without using any class label information, and this is more appropriate, realistic, and useful in many real-world applications. We compare the proposed method with two standard metrics and several recent relevant methods. The experimental results have shown that the effectiveness of the proposed method in exploiting the RF log data to improve the performance of CBIR.

Acknowledgment

The authors would like to acknowledge the Ph.D. grant from the Institute for Media Innovation, Nanyang Technological University, Singapore. This work was partially supported by the SINGAPORE MINISTRY OF EDUCATION Academic Research Fund (AcRF) Tier 2, Grant Number: T208B1218.

References

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349 –1380, Dec. 2000.

[2] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 1, pp. 39–62, 1999.

[3] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–60, May 2008.

[4] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Systems*, vol. 8, no. 6, pp. 536–544, Apr. 2003.
[5] L. Zhang, L. Wang, and W. Lin, "Semisupervised biased maximum margin analysis for interactive image retrieval," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2294–2308, 2012.
[6] L. Zhang, L. Wang, and W. Lin, "Generalized biased discriminant analysis for content-based image retrieval," *IEEE Transactions on Systems, Man, Cybernetics-Part B: Cybernetics*, vol. 42, no. 1, pp. 282–290, Feb. 2012.
[7] J. Li and J. Z. Wang, "Real-time computerized annotation of pictures," *IEEE Transactions on Pattern Analysis and Machine Intellitence*, vol. 30, no. 6, pp. 985–1002, June 2008.
[8] J. He, M. Li, H. Zhang, H. Tong, and C. Zhang, "Manifold-ranking based image retrieval," in *Proceedings of the 12th ACM International Conference on Multimedia*, 2004.
[9] C. H. Hoi and M. R. Lyu, "A novel log-based relevance feedback technique in content-based image retrieval," in *Proceedings of the 12th ACM International Conference on Multimedia*, 2004, pp. 24–31.
[10] L. Zhang, L. Wang, and W. Lin, "Conjunctive patches subspace learning with side information for colaborative image retrieval," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3707–3720, 2012.
[11] C. H. Hoi, W. Liu, and S. F. Chang, "Semi-supervised distance metric learning for collaborative image retrieval and clustering," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 6, no. 3, pp. 1–26, 2010.
[12] C. H. Hoi, M. R. Lyu, and R. Jin, "A unified log-based relevance feedback scheme for image retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 4, pp. 509 – 524, 2006.
[13] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning with application to clustering with side-information," *Advances in Neural Information Processing Systems*, pp. 521–528, 2003.
[14] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning a mahalanobis metric from equivalence constraints," *Journal of Machine Learning Research*, vol. 6, no. 1, pp. 937–965, 2006.
[15] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323, 2000.
[16] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.
[17] J. Chen, S. Shan, C. He, G. Zhao, P. Matti, X. Chen, and W. Gao, "Wld: A robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705 –1720, 2010.
[18] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, no. 8, pp. 1233 – 1244, 1996.
[19] C. H. Hoi, W. Liu, M. R. Lyu, and W. Ma, "Learning Distance Metrics with Contextual Constraints for Image Retrieval," in *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2072–2078.