

Laplacian Regularized Subspace Learning for Interactive Image Reranking

Lining Zhang

School of Electrical and Electronic
Engineering
Nanyang Technological University
Singapore 639798
zhan0327@ntu.edu.sg

Lipo Wang

School of Electrical and Electronic
Engineering
Nanyang Technological University
Singapore 639798
elpwang@ntu.edu.sg

Weisi Lin

School of Computer Engineering
Nanyang Technological University
Singapore 639798
wslin@ntu.edu.sg

Abstract—Content-based image retrieval (CBIR) has attracted substantial attention during the past few years for its potential application. To bridge the gap between the low level visual features and the high level semantic concepts, various relevance feedback (RF) or interactive reranking schemes have been designed to improve the performance of a CBIR system. In this paper, we propose a novel subspace learning based IR scheme by using a graph embedding framework, termed Laplacian Regularized Subspace Learning (LRSL). The LRSL method can model both the within-class compactness and the between-class separation by specially designing an intrinsic graph and a penalty graph in the graph embedding framework, respectively. In addition, LRSL can share the popular assumption of the biased discriminant analysis (BDA) for IR but avoid the singular problem in BDA. Extensive experimental results have shown that the proposed LRSL method is effective for reducing the semantic gap and targeting the intentions of users for an image retrieval task.

Keywords—content based image retrieval; image reranking; subspace learning; graph embedding

I. INTRODUCTION

During the past few years, image retrieval has attracted much attention in the multimedia society [1]. It is impractical for users to manually give all images precise textual descriptions. Moreover, to clearly describe an image using limited words is infeasible. On the other hand, to automatically annotate an image record is actually far beyond the current technology. Therefore, content based image retrieval (CBIR) has gained tremendous interests both from the academia and the industry.

However, the gap between low-level visual features and high-level semantic concepts usually lead to the poor performance of a CBIR system. In addition, different users even the same user at a different time may have different perception on the same image [1]. Therefore, interaction schemes between the user and the system, e.g., relevance feedback (RF) or interactive reranking (IR), are actually required to improve the performance of a CBIR system [2] [3].

In a CBIR system, images are usually represented by low-level visual features in a high dimensional space. It is reasonable to assume that different semantic concepts live in different subspaces and each image can live in many different subspaces. Consequently, it is necessary to reduce the

dimensionality of visual features to acquire an efficient and effective semantic subspace before conducting an image retrieval task. Based on a popular assumption that all positive samples are alike while each negative example is negative in its own way, biased discriminant analysis (BDA) [3] was proposed for subspace learning in image retrieval, in which only positive samples are required to be clustered, and negative feedback samples are required to be separated from the centroid of positive feedback samples. The kernel version, termed kernel BDA (i.e., KBDA for short), was also devised to facilitate nonlinear biased discrimination. To address the singular problem of the positive within scatter in BDA, DBDA [4] was proposed by utilizing the direct method. However, this approach still discards the null space of the negative scatter with respect to the positive centroid, which contains important discriminative features. As a variant of marginal Fisher analysis, marginal biased analysis (MBA) [5] was also introduced to construct an RF approach and has shown better performance than BDA; however, it still suffers from the intrinsic singular problem in the original BDA. By precisely parametrizing positive feedbacks, negative feedbacks and unlabelled samples, Bian and Tao proposed an RF method, which can find the intrinsic coordinate of image low-level visual features [6]. Nevertheless, it is generally believed that more samples are actually required to model the exquisite geometry structure in a high dimensional space.

Various subspace learning methods have been proposed to capture the intrinsic geometry property of samples in a high dimensional space during the past few years. In [7], Yan et al. has shown that most of popular methods for subspace learning or dimensionality reduction methods can be mathematically unified into a graph embedding framework. For a subspace learning problem, the graph embedding framework requires an intrinsic graph, which is used to describe the similarity relation of sample pairs within a same class, and a penalty graph, which characterizes the difference relation of sample pairs between different classes.

In this paper, we proposed a Laplacian Regularized Subspace Learning (LRSL) method based on the graph embedding framework [7] for IR in CBIR. The new algorithm can not only extract the most discriminative information from the nearest neighbor samples in different classes but also share the popular assumption (i.e., all positive examples are alike; each negative example is negative in its own way) used in

BDA [3], DBDA [4] and MBA [5]. In addition, by introducing a Laplacian regularization term [8], a locally smooth transformation can also be learned, which significantly reduce the overfitting problem. Finally, the new algorithm can never meet the singular problem encountered by BDA and MBA. Extensive experiments on a subset of Corel Image Database have shown the effectiveness of the proposed LRSL for IR in comparing with representative methods.

II. LAPLACIAN REGULARIZED SUBSPACE LEARNING FOR INTERACTIVE IMAGE RERANKING

Let us denote the high-dimensional space as R^h and the low-dimensional space as R^l . In each round of feedback iteration, there are n feedback samples $X = \{x_1, x_2, \dots, x_n\} \in R^h$. For simplicity, we assume that the first n^+ samples are positive samples $x_i (1 \leq i \leq n^+)$, the next n^- samples are negative samples $x_i (n^+ + 1 \leq i \leq n^+ + n^-)$. Let $l(x_i)$ be the class label of sample x_i , we denote $l(x_i) = 1$ for positive samples, $l(x_i) = -1$ for negative samples. For simplicity, we restrict the embedding transform to be linear, which can be defined by a transformation matrix $\alpha \in R^{h \times l} (l < h)$.

With the observation that “all positive example are alike; each negative sample is negative in its own way” [3], two different graphs should be formed accordingly [7]: 1) an intrinsic graph G , which characterizes the similarity of positive samples; 2) a penalty graph G^p , which characterizes the difference between positive and negative samples.

We first construct an intrinsic graph to characterize the similarity relation of positive samples in the positive class. Since all positive samples are alike, for each positive sample x_i , we find all other positive feedback samples in the positive class, which always share a common concept with the positive sample x_i , and put an edge between x_i and all other positive samples. The intrinsic graph of the positive class is characterized as follows:

$$\begin{aligned} \alpha &= \arg \min \left(\sum_i \sum_j \|\alpha^T x_i - \alpha^T x_j\|^2 * W_{ij} \right) \\ &= \arg \min \left(\text{tr}[\alpha^T X (D - W) X^T \alpha] \right) \\ &= \arg \min \left(\text{tr}[\alpha^T X L_1 X^T \alpha] \right) \end{aligned} \quad (1)$$

$$W_{ij} = \begin{cases} 1/|N|, & \text{if } l(i) = 1 \text{ and } l(j) = 1, i \neq j \\ 0, & \text{if } l(i) = 1 \text{ and } l(j) = 1, i = j \end{cases} \quad (2)$$

where D is a diagonal matrix whose diagonal elements are calculated by $D_{ii} = \sum_j W_{ij}$; $|N|$ is the total number of pairs of samples in the positive class. Basically, the intrinsic graph measures the total average distance of the positive sample pairs, and is used to characterize the within-class compactness for all the positive feedback samples.

And then, we construct a penalty graph to characterize the difference relation between positive and negative samples. For all feedback samples, we first compute the pair-wise distances between each pair of samples in different classes. More strictly speaking, we expect that the total average distance between the sample pairs with different labels should be as large as possible. For each feedback sample, we find its k_1 nearest neighbor feedbacks with different labels and corresponding pairs of feedback samples with weights W_{ij}^p . The penalty graph can be formed as follows:

$$\begin{aligned} \alpha &= \arg \max \left(\sum_i \sum_{j: j \in \mathbb{N}_i^p \text{ or } i \in \mathbb{N}_j^p} \|\alpha^T x_i - \alpha^T x_j\|^2 * W_{ij}^p \right) \\ &= \arg \max \left(\text{tr}[\alpha^T X (D^p - W^p) X^T \alpha] \right) \end{aligned} \quad (3)$$

$$\begin{aligned} &= \arg \max \left(\text{tr}[\alpha^T X B X^T \alpha] \right) \\ W_{ij}^p &= \begin{cases} 1/|N^p|, & \text{if } l(i) = 1 \text{ and } l(j) = -1, i \in \mathbb{N}_j^p \text{ or } j \in \mathbb{N}_i^p \\ 0, & \text{else} \end{cases} \end{aligned} \quad (4)$$

where D^p is a diagonal matrix whose diagonal elements are calculated by $D_{ii}^p = \sum_j W_{ij}^p$; $|N^p|$ denotes the total number of k_1 nearest neighbor sample pairs with different labels. Similarly, the penalty graph measures the total average distance of the $|N^p|$ nearest neighbor sample pairs in different classes, and is used to characterize the between-class separation.

In the following, we describe how to utilize the graph embedding framework to develop algorithms based on the designed intrinsic and penalty graphs [7]. Different from the original trace ratio formulation of the graph embedding framework in [7], the new algorithm optimizes the objective function in a trace difference form instead, i.e.,

$$\begin{aligned} \alpha^* &= \arg \max_{\alpha} \left(\text{tr}(\alpha^T X B X^T \alpha) - \lambda * \text{tr}(\alpha^T X L_1 X^T \alpha) \right) \\ &= \arg \max_{\alpha} \text{tr}[\alpha^T X (B - \lambda * L_1) X^T \alpha] \end{aligned} \quad (5)$$

where λ is a nonnegative tuning parameter and is used to trade off the between-class separation and within-class compactness. As given in Equation (5), we can notice that the objective function works in two ways, which tries to maximize $\text{tr}(\alpha^T X B X^T \alpha)$ and at the same time minimize $\text{tr}(\alpha^T X L_1 X^T \alpha)$. By formulating the objective function as a trace difference form, we can regard it as the total average margin between the positive and negative class. Therefore, Equation (5) can be used as a criterion to discriminate the different classes.

To avoid trivial solutions, in this paper, we impose a constraint, i.e., $\alpha^T \alpha = I$, on Equation (5). And thus, this problem can be solved by conducting the standard Eigenvalue decomposition and the mapping matrix α is formed by the l eigenvectors associated with the first l largest eigenvalues.

Moreover, it is reasonable to expect that integrating the essential manifold structure [8] of the positive samples will

further improve the performance of the IR algorithm for an

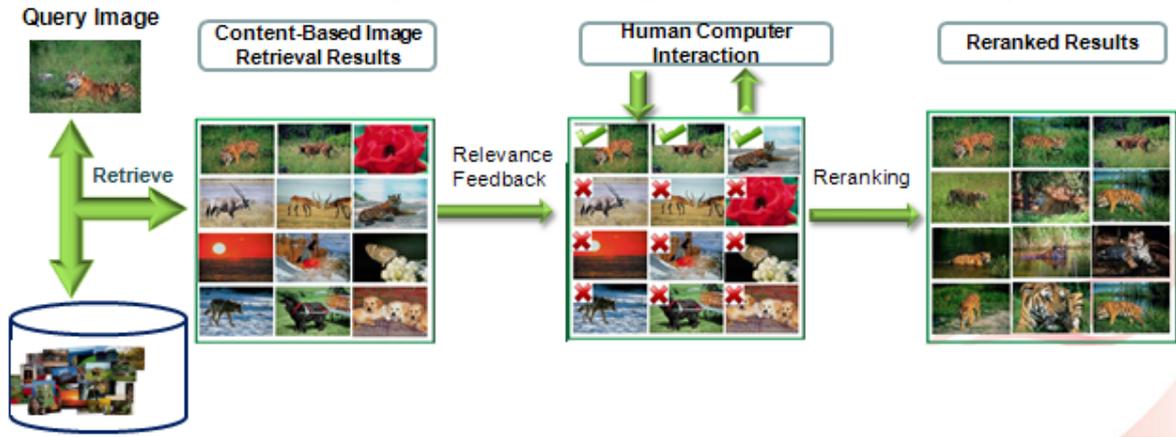


Fig.1 the framework of the CBIR system using the Laplacian Regularized Subspace Learning for IR

design another intrinsic graph to characterize the local consistence property of positive samples, which can be used to regularize the separability of different classes. The regularization term implements the local consistency principle by preserving the local similarity among the positive neighborhood samples. For each positive sample x_i , we find its k_2 nearest neighborhood positive samples, which can be represented as a sample set N_i^s , and put an edge between x_i and its neighborhood positive samples. Then the local consistency for the positive samples can be characterized as follows:

$$\begin{aligned} \alpha &= \arg \min \left(\sum_i \sum_{j: j \in N_i^s \text{ or } i \in N_j^s} \|\alpha^T x_i - \alpha^T x_j\|^2 * W_{ij}^u \right) \\ &= \arg \min (\text{tr}[\alpha^T X (D^L - W^L) X^T \alpha]) \\ &= \arg \min (\text{tr}[\alpha^T X L_2 X^T \alpha]) \end{aligned} \quad (6)$$

$$W_{ij}^L = \begin{cases} \frac{1}{|N^L|} \exp(-\|x_i - x_j\|^2 / \delta^2), & \text{if } l(i) = l(j) = 0, i \in N_j^s \text{ or } j \in N_i^s \\ 0, & \text{else} \end{cases} \quad (7)$$

where $\omega_{ij} = \exp(-\|x_i^p - x_j^p\|^2 / \delta^2)$ is the heat kernel according to Laplacian Eigenmaps [8], which reflects the affinity of the sample pairs; $W = \{w_{ij}\} \in R^{n^+ * n^+}$ is a symmetric matrix, which reflects the local geometry of the positive samples in a high dimensional space; and D_w is a diagonal matrix and its i -th entry is $\sum_{j=1}^{n^+} w_{ij}$; $S(i)$ is the set of indices of the neighboring samples in the positive class for the positive sample x_i ; $|N^L|$ is the total number of k_2 nearest positive samples for all positive samples.

According to [8] [9], a definition in Equation (11) corresponds to the approximation of $\int_M \|\nabla f(x)\|^2$, the manifold on which the positive samples reside. Minimizing the objective function can encourage the consistent output for the positive samples in the high dimensional space and this will result in transforming with high local smoothness and best

image retrieval task. As a consequence, we can further

local preservation. Hence, a smooth transformation that is expected to be less likely to over fit the training samples can be learned by this Laplacian regularization term. By integrating the Laplacian regularizer into the separability term, we can easily obtain the LRSL for IR, i.e.,

$$\alpha^* = \arg \max (\text{tr}[\alpha^T X (B - \lambda * L_1 - \beta * L_2) X^T \alpha]) \quad (8)$$

The value of λ is used to balance the between-class separability and the within-class compactness, which can be set as 1 in experiments for simplicity. And the value of β trades off the separability of different classes and the local consistency of positive samples in the positive class. In the following experiments, we present the sensitivity of LRSL in relation to the regularization parameter β , and then select the value that shows the best performance.

III. CONTENT-BASED IMAGE RETRIEVAL SYSTEM

Given a query image by the user, the CBIR system is expected to feedback more semantically relevant images after each feedback iteration [1]. We have implemented a CBIR system to evaluate the referred IR algorithms. The framework of the CBIR system is illustrated in Figure 1.

From Figure 1, we can notice that when a query image is provided by the user, the system first extracts the low level features of the query. The all the images in the database are sorted based on a certain similarity metric, i.e., Euclidean distance metric. If the user is satisfied with the results, the retrieval process is ended, and the results are presented to the user. However, because of the semantic gap, most of the time, the user is not satisfied with the first retrieved results. Then she/he will label the most semantically relevant and irrelevant images as positive and negative feedbacks in top retrieved results, respectively. Based on these feedback samples, an RF model can be trained based on certain machine learning algorithms. The similarity metric can be updated as well as the RF model. Then, all the images are resorted based on the recalculated similarity metric. If the user is not satisfied with

the retrieved results, the IR process will be performed iteratively.

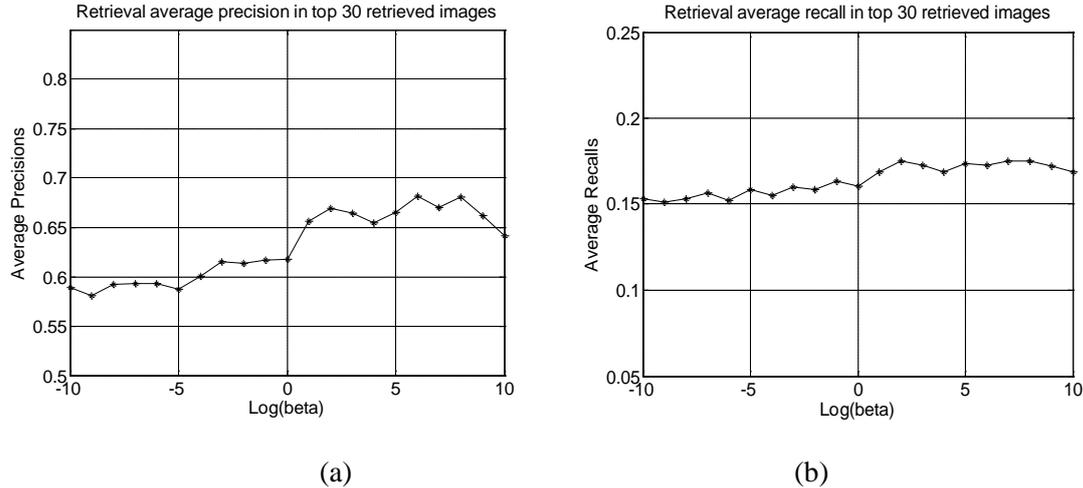


Fig.2 Performance evaluation at the top 30 reranked images for different value of β (a) average precisions. (b) average recalls

IV. EXPERIMENTAL RESULTS

All experiments are implemented on a Corel Image Database, which includes 10,763 images with 80 different conceptual groups, e.g., tiger, castle, cloud, dog, elephant, iceberg, train and waterfall. To represent images, we utilize the color histogram [10], Webber's law descriptors [11] and the edge directional histogram from Y component in YCrCb space [12], each of which can describe the semantic content of images from different aspects, respectively. 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.

In experiments, 500 queries are randomly selected from the image database, and then RF is automatically done by a computer. At each feedback iteration, the top 20 retrieved result are serially examined from the top; the first 5 query relevant images are labeled as positive feedbacks and the first 5 query irrelevant images are marked as negative feedbacks unless fewer such images are found among the top 20 images, in which case the fewer number of samples found are used as the feedbacks. All the labeled images in the feedback iterations are used to train an IR model.

In this paper, average precision, standard deviation, and average recall are utilized to evaluate the performance of algorithms. In experiments, the parameter k_1 and k_2 are empirically set as 4 according to manifold learning approaches [7] [8].

A. Sensitivity in relation to the regularization parameter

This experiment shows the sensitivity of LRSL with regard to different values of the regularization parameter β . In experiment, we empirically set the parameter β as a value in a sequence, i.e., $\{2^i, i = -10, -9, \dots, 9, 10\}$. Fig.2 shows the average precision and average recall curves in top 30 results of the 7-th feedback iterations with different β values based on 500 independent experiments, respectively. In experiments,

we find that the regularization parameter can significantly affect the results. However, it is still open problem how to tune the regularization parameter to balance the separability term and the local consistency term. From the results, we can see that for this problem, the algorithm achieves best performance when β is set as 2^6 . Therefore, in the following experiments, we empirically set the tradeoff parameter $\beta = 2^6$. It is convinced that the parameter β can be further tuned to achieve better performance. This analysis above also indicates the important role of local consistency for improving the generalization ability.

B. Statistical Experimental Results

In this part, we compare the proposed LRSL with some representative methods, i.e., BDA [3], DBDA [4], MBA [5]. The figures in Fig.3 show the average precisions and standard deviations of the 500 experiments for top 10, 30 and 50 results. We can see that the LRSL can achieve better performance comparing with the original BDA.

The unreasonable Gaussian distribution assumption for the positive samples in the original BDA and the regularization method used are the main reasons which usually result in poor performance of IR for CBIR. The DBDA algorithm solves the singular problem by the direct method and can achieve better performance than the original BDA. However, much discriminative information contained in the null space of S_b is discarded. The MBA algorithm effectively extracts the discriminative information from the marginal samples; however, it still suffers from the singular problem, which causes serious stability problems for MBA. The proposed LRSL method can extract the most discriminative information from the nearest neighborhood samples in different classes, but never encounters the singular problem. By introducing the Laplacian regularization term, a locally smooth transformation can also be learned which is expected to be less vulnerable to over fit the training samples and have good generalization properties. The experimental results have shown that the

proposed LRSL consistently outperforms the BDA, DBDA and MBA.

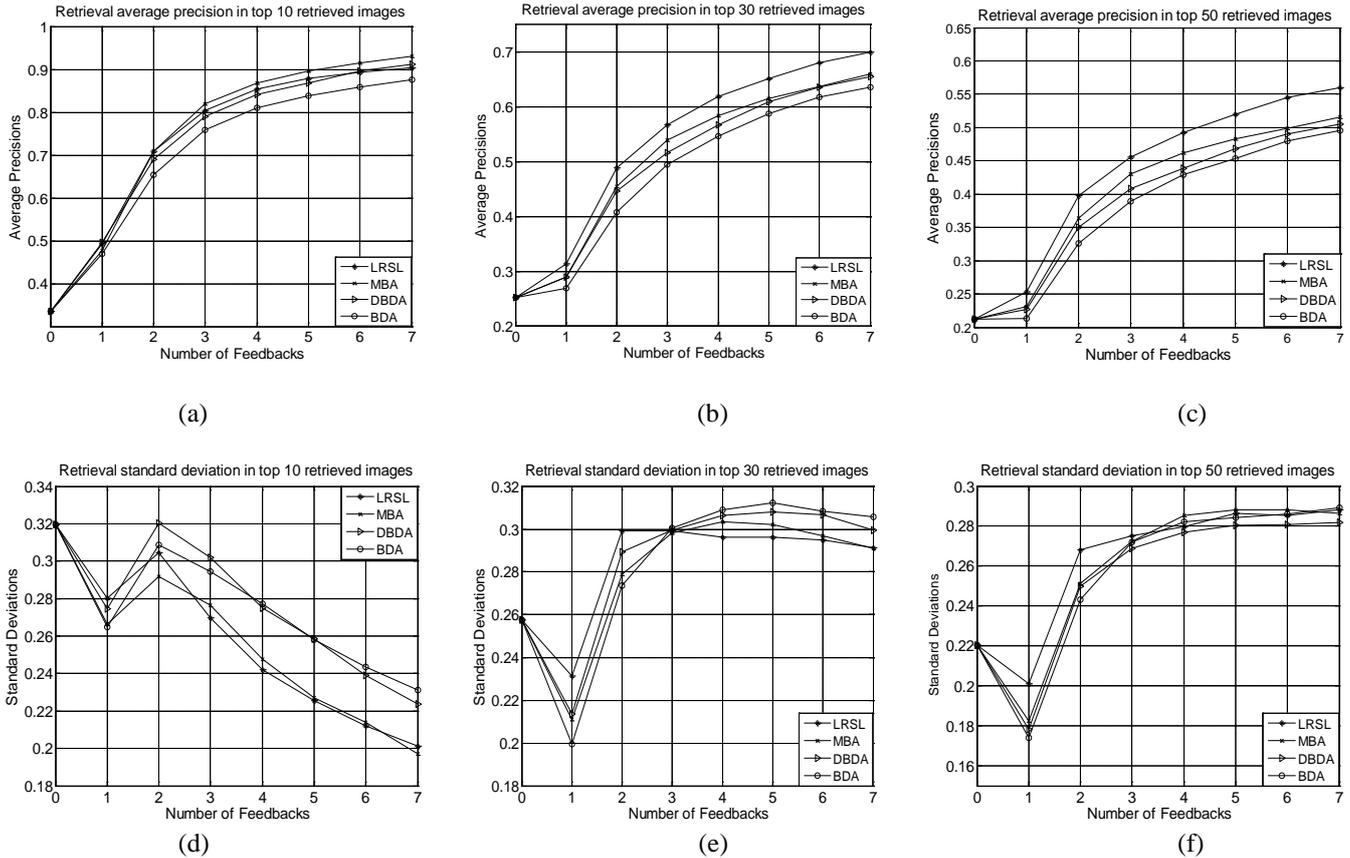


Fig.3 Evaluation experimental results based on the Corel database with 10,768 images with 300 queries. (a), (b), and (c) display the retrieval average precision in top 10, 30, and 50 retrieved images, respectively. (d), (e), and (f) display the corresponding standard deviation of the precision curve in top 10, 30, and 50 retrieved images, respectively.

V. CONCLUSIONS

In this paper, we have studied a new subspace learning algorithm, Laplacian Regularized Subspace Learning (LRSL) for Interactive Reranking (IR) in CBIR. The proposed LRSL algorithm is under the graph embedding framework, with a particular designed intrinsic graph and a penalty graph to simultaneously capture the within-class compactness and the between-class separation, respectively. By introducing a Laplacian regularization term, a smooth and locally consistent transformation can also be learned for IR to effectively reduce the risk of over fitting. Extensive experimental results on a large real world image database have shown that the proposed LRSL method has much improved performance than other representative methods.

REFERENCE

- [1] R. Datta, D. Joshi, Jia Li, J.Z. Wang, "Image retrieval: ideas, influences, and trends of the new age". *ACM Computing Surveys*, Vol. 40, No.2, pp.1–60, Apr., 2008.
- [2] X. Zhou and T. Huang, "Relevance feedback for image retrieval: a comprehensive review", *ACM Multimedia Systems J.*, Vol. 8, pp. 536-544, 2003.
- [3] X. Zhou and T. Huang, "Small sample learning during multimedia retrieval using biasmap", *In Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp.11-17,2001.
- [4] D. Tao, X. Tang, X. Li, and Y. Rui, "Kernel direct biased discriminant analysis: a new content-based image retrieval relevance feedback algorithm", *IEEE Trans.on Multimedia*, Vol.8, No.4, pp.716-724, Aug. 2006.
- [5] D. Xu, S. Yan, D. Tao, S. Lin, H. Zhang, "Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval", *IEEE Trans. on Image Processing* Vol.16, No.11, pp.2811-2821, Nov. 2007.
- [6] W. Bian and D. Tao, "Biased Discriminant Euclidean Embedding for Content-Based Image Retrieval", *IEEE Trans. on Image Processing* Vol.19, No.2, pp.545-554, Feb. 2010.
- [7] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, Graph embedding and extensions: A general framework for dimensionality reduction, *IEEE Trans.on Pattern Anal. Mach. Intell.*, Vol. 29, No. 1, pp.40–51, Jan. 2007.
- [8] X. He, S. Yan, Y.Hu, P. Niyogi and H. Zhang, "Face recognition using laplacian faces", *IEEE Trans.on Pattern Anal. Mach. Intell.*, Vol. 27, No. 3, pp.1624–1637, Dec. 2005.
- [9] X. He and P. Niyogi, "Locality preserving projections", *In Advances in Neural Information Processing Systems*, Vol. 16, 2004.
- [10] M. J. Swain and D. H. Ballard, "Color indexing", *International. Journal of Computer. Vision.*, Vol.7, No. 1, pp. 11–32, 1991.
- [11] J. Chen, S. Shan, G. Zhao, X. Chen, W. Gao, Matti Pietikainen, "WLD: A Robust Descriptor based on Weber's Law", *IEEE Trans.on Pattern Anal. Mach. Intell.*, Vol. 32, No.9, pp. 1705–1720, Sep. 2010.

[12] Y. Rubner, J. Puzicha, C. Tomasi, J. M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture", *Computer*

Vision and Image Understanding, Vol.84, pp.25-43. 2001.