Semi-Supervised Biased Maximum Margin Analysis for Interactive Image Retrieval

Lining Zhang[1,2], *Student Member, IEEE,* Lipo Wang[1], *Senior Member, IEEE* and Weisi Lin[3], *Senior Member, IEEE*

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798
[2]Institute for Media Innovation, Nanyang Technological University, Singapore, 637553
[3]School of Computer Engineering, Nanyang Technological University, Singapore, 639798

***Abstract*— With many potential practical applications, Content-Based Image Retrieval (CBIR) has attracted substantial attention during the past few years. A variety of Relevance Feedback (RF) schemes have been developed as a powerful tool to bridge the semantic gap between low-level visual features and high-level semantic concepts and thus to improve the performance of CBIR systems. Among various RF approaches, Support Vector Machine (SVM) based RF is one of the most popular techniques in CBIR. Despite the success, directly using SVM as a RF scheme has two main drawbacks. First, it treats the positive and negative feedbacks equally, which is not appropriate since the two groups of training feedbacks have distinct properties. Second, most of the SVM based RF techniques do not take into account the unlabelled samples although they are very helpful in constructing a good classifier. To explore solutions to overcome these two drawbacks, in this work, we propose a Biased Maximum Margin Analysis (BMMA) and a Semi-Supervised Biased Maximum Analysis (SemiBMMA), for integrating the distinct properties of feedbacks and utilizing the information of unlabeled samples for SVM based RF schemes. The BMMA differentiates positive feedbacks from negative ones based on local analysis, while the SemiBMMA can effectively integrate information of unlabelled samples by introducing a Laplacian regularizer to the BMMA. We formulate the problem into a general subspace learning task and then propose an automatic approach of determining the dimensionality of the embedded subspace for RF. Extensive experiments on a large real world image database demonstrate that the proposed scheme combined with the SVM RF can significantly improve the performance of CBIR systems.**

***Index Terms*— Support Vector Machine, Relevance Feedback, Graph Embedding, Content-Based Image Retrieval**

I. Introduction

During the past few years, Content-Based Image Retrieval (CBIR) has gained more attention for its potential application in multimedia management [1, 2]. It is motivated by the explosive growth of image records and online accessibility of remotely stored images. An effective search scheme is urgently required to manage the huge image database. Different from the traditional search engine, in CBIR, an image query is described using one or more example images and low-level visual features (e.g., color [3-5], texture [5-7], shape [8-10], etc.) are automatically extracted to represent the images in the database. However, the low-level features captured from the images may not accurately characterize the high-level semantic concepts [1, 2].

To narrow down the so called semantic gap, Relevance Feedback (RF) was introduced as a powerful tool to enhance the performance of CBIR [11, 12]. Huang et al introduced both the query movement and re-weighting techniques [13, 14]. Self Organizing Map was used to construct the RF algorithms [15]. In [16], one-class Support Vector Machine (SVM) estimated the density of positive feedback samples. Derived from one-class SVM, a biased SVM inherited the merits of one-class SVM but incorporated the negative feedback samples [17]. Considering the geometry structure of image low-level visual features, [18, 19] proposed manifold learning based approaches to find intrinsic structure of images and improve the retrieval performance. With the observation that *"all positive examples are alike; each negative example is negative in its own way"*, RF was formulated as a biased subspace learning problem, in which there is an unknown number of classes, but the user is only concerned about the positive class [20,21,22]. However, all of these methods have some limitations. For example, the method in [13, 14] is heuristically based, the density estimation method in [16] ignores any information contained in the negative feedback samples, and the discriminant subspace learning techniques in [20, 22] often suffer from the so-called "Small Sample Size" problem. Regarding the positive and negative feedbacks as two different groups, classification-based RFs [23, 24, 25] have become a popular technique in the CBIR community. However, RF is very different from the traditional classification problem because the feedbacks provided by the user are often limited in real-world image retrieval systems. Therefore, small sample learning methods are most promising for RF.

Two-class SVM is one of the popular small sample learning methods widely used in recent years and obtains the state of the art performance in classification for its good generalization ability [24-28]. The SVM can achieve a minimal structural risk by minimizing the Vapnik-Chervonenkis dimensions [27]. Guo et al developed a constrained similarity measure for image

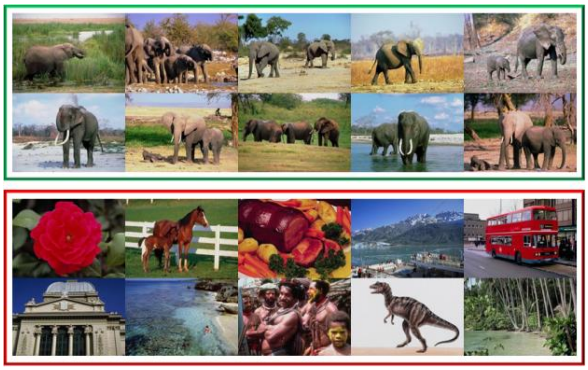


Fig.1 A typical set of positive and negative feedback samples in a relevance feedback iteration

retrieval [26], which learns a boundary that divides the images into two groups and samples inside the boundary are ranked by their Euclidean distance to the query image. The SVM active learning method selects samples close to the boundary as the most informative samples for the user to label [28]. Random sampling techniques were applied to alleviate unstable, biased and overfitting problems in SVM RF [25]. Li et al proposed a multitraining SVM method by adapting a co-training technique and a random sampling method [29]. Nevertheless, most of the SVM RF approaches ignore the basic difference between the two distinct groups of feedbacks, that is, all positive feedbacks share a similar concept while each negative feedback usually varies with different concepts. For instance, a typical set of feedback samples in RF iteration are shown in Fig.1. All the samples labeled as positive feedbacks share a common concept (i.e., elephant), while each sample labeled as negative feedback varies with diverse concepts (i.e., flower, horse, banquet, hill, etc.). Traditional SVM RF techniques treat positive and negative feedbacks equally [24, 25, 26, 28, 29]. Directly using the SVM as a RF scheme is potentially damaging to the performance of CBIR systems. One problem stems from the fact that different semantic concepts live in different subspaces and each image can live in many different subspaces, and it is the goal of RF schemes to figure out "which one" [20]. However, it will be a burden for traditional SVM based RF schemes to tune the internal parameters to adapt to the changes of the subspace. Such difficulties have severely degraded the effectiveness of traditional SVM RF approaches for CBIR. Additionally, it is problematic to incorporate the information of unlabelled samples into traditional SVM based RF schemes for CBIR, although unlabelled samples are very helpful in constructing the optimal classifier, alleviating noise and enhancing the performance of the system.

To explore solutions to these two aforementioned problems in the current technology, we propose a Biased Maximum Margin Analysis (BMMA) and a Semi-Supervised Biased Maximum Margin Analysis (SemiBMMA) for the traditional SVM RF schemes, based on the graph embedding framework [30]. The proposed scheme is mainly based on (a) the effectiveness of treating positive examples and negative examples unequally [20, 21, 22]; (b) the significance of the optimal subspace or feature subset in interactive CBIR; (c) the success of graph embedding in characterizing intrinsic geometric properties of the data set in high-dimensional space [30, 31, 32]; and (d) the convenience of the graph embedding framework in constructing semi-supervised learning techniques. With the incorporation of BMMA, labeled positive feedbacks are mapped as close as possible, while labeled negative feedbacks are separated from labeled positive feedbacks by a maximum margin in the reduced subspace. The traditional SVM combined with BMMA can better model the relevance feedback process and reduce the performance degradation caused by distinct properties of the two groups of feedbacks. The SemiBMMA can incorporate the information of unlabelled samples into the relevance feedback and effectively alleviate the over fitting problem caused by the small size of labeled training samples. To show the effectiveness of the proposed scheme combined with the SVM RF, we will compare it with the traditional SVM RF and some other relevant existing techniques for RF on a real world image collection. Experimental results demonstrate that the proposed scheme can significantly improve the performance of the SVM RF for image retrieval.

The rest of this paper is organized as follows: in Section II, the related previous work, i.e., the principle of SVM RF for CBIR and the graph embedding framework, are briefly reviewed; in Section III, we introduce the BMMA and the SemiBMMA for SVM RF; an image retrieval system is given in Section IV; a large number of experiments which validate the effectiveness of the proposed scheme are given in Section V; conclusion and future work are presented in Section VI.

II. Related Previous Work

A. The principle of SVM RF for CBIR

In this section, we briefly introduce the principle of the traditional SVM based RF for CBIR. The SVM implements the structure risk minimization by minimizing Vapnik-Chervonenkis dimensions [27]. Consider a linearly separable binary classification problem as follows:

$$\{(x_1, y_1), \ldots, (x_N, y_N)\} \text{ and } y_{i=1,\ldots,N} = \{+1, -1\} \quad (1)$$

where x_i denotes a h-dimensional vector, N is the number of training samples and y_i is the label of the class that the vector belongs to. The objective function of SVM aims to find an optimal hyperplane to separate the two classes, i.e.,

$$w^T x + b = 0 \quad (2)$$

where x is an input vector, w is a weight vector and b is a bias. The SVM attempts to find the two parameters w and b for the optimal hyperplane by maximizing the geometric margin $2/\|w\|$, subject to:

$$y_i (w^T x_i + b) \geq 1 \quad (3)$$

The solution of the objective function can be found through a Wolf dual problem with the Lagrangian multiplied by α_i :

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)/2 \quad (4)$$

subject to $\alpha_i \geq 0$ and $\sum_{i=1}^{N} \alpha_i y_i = 0$.

In general, in the dual problem data points appear only in the inner product, which can often be replaced with a positive definite kernel function for better performance.

$$x_i \cdot x_j \to \Phi(x_i) \cdot \Phi(x_j) = K(x_i, x_j) \quad (5)$$

where $K(\cdot)$ is a kernel function. The kernel version of the Wolfe dual problem is

$$Q(\alpha) = \sum_{i=1}^{N} \boldsymbol{\alpha}_i - \sum_{i,j=1}^{N} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_j y_i y_j K(x_i \cdot x_j)/2 \quad (6)$$

Thus, for a given kernel function, the SVM classifier is given by

$$F(x) = \mathrm{sgn}(f(x)) \quad (7)$$

where $f(x) = \sum_{i=1}^{s} \boldsymbol{\alpha}_i y_i K(x_i, x) + b$ is the output hyperplane decision function of SVM and s is the number of support vectors.

Generally, the output of SVM (i.e., $f(x)$), is usually used to measure the similarity between a given pattern and the query image in the traditional SVM RF for CBIR. The performance of a SVM classifier depends mainly on the number of support vectors. Orthogonal Complement Component Analysis (OCCA) decreases the number of support vectors by finding a subspace, in which all the positive feedbacks are merged [33]. However, it still totally ignores the information contained in negative feedbacks, which is very helpful in finding a homogeneous subspace. Intuitively, good separation is achieved by the hyperplane that has the largest distance to the nearest training samples, since in general, the larger the margin, the lower the generalization error of the classifier.

B. Graph embedding framework

In order to describe our proposed approach clearly, we firstly review the graph embedding framework introduced in [30]. Generally, for a classification problem, the sample set can be represented as a matrix $X = [x_1, x_2, \ldots, x_n] \in R^{h \times n}$, where n indicates the total number of the samples and h is the feature dimension. Let $G = \{X, W\}$ be an undirected similarity graph, which is called an intrinsic graph, with vertices set X and similarity matrix $W \in R^{n^* n}$. The similarity matrix W is real and symmetric, and measures the similarity between a pair of vertices; W can be formed using various similarity criteria. The corresponding diagonal matrix D and the Laplacian matrix L of the graph G can be defined as follows:

$$L = D - W, D_{ii} = \sum_{j \neq i} W_{ij}, \forall\ i = 1, \ldots, n \quad (8)$$

Graph embedding of the graph G is defined as an algorithm to determine the low-dimensional vector representations $Y = [y_1, y_2 \ldots, y_n] \in R^{l \times n}$ of the vertex set X, where l is lower than h for dimensionality. The column vector y_i is the embedding vector for the vertex x_i, which preserves the similarities between pairs of vertices in the original high-dimensional space. Then in order to characterize the difference between pairs of vertices in the original high-dimensional space, a penalty graph $G^p = \{X, W^p\}$ is also defined, where the vertices X are the same as those of G, but the edge weight matrix W^p corresponds to the similarity characteristics that are to be suppressed in the low-dimensional feature space. For a dimensionality reduction problem, direct graph embedding requires an intrinsic graph G, while a penalty graph G^p is not a necessary input. Then the similarities among vertex pairs can be maintained according to the graph preserving criterion as follows:

$$y^* = \arg\min_{tr(YBY^T)=c} \sum_{i \neq j} \| y_i - y_j \|^2 W_{ij} = \arg\min_{tr(YBY^T)=c} tr(YLY^T) \quad (9)$$

where $tr(\cdot)$ is the trace of an arbitrary square matrix; c is a constant; B is the constraint matrix. B may typically be a diagonal matrix for scale normalization or express more general constraints among vertices in a penalty graph G^p, and it describes the similarities between vertices that should be avoided; B or L^p is the Laplacian matrix of G^p, similarly to Equation (8), which can also be defined as follows:

$$L^p = D^p - W^p, D^p_{ii} = \sum_{j \neq i} W^p_{ij}, \forall\ i = 1, \ldots, n \quad (10)$$

where W^p is the similarity matrix of penalty graph G^p to measure the difference between a pair of vertices in G^p.

The graph embedding framework preserves the intrinsic property of the samples in two ways: For larger similarity between samples x_i and x_j, the distance between y_i and y_j should be smaller to minimize the objective function. Conversely, smaller similarity between x_i and x_j should lead to larger distance between y_i and y_j. Hence, through the intrinsic graph G and penalty graph G^P, the similarities and differences among vertex pairs in a graph G can be preserved in the embedding.

In [30], based on the graph embedding framework, Equation (9) can be resolved by converting it into the following trace ratio formulation:

$$Y^* = \arg\min_{Y} \frac{tr(YLY^T)}{tr(YBY^T)} \quad (11)$$

Generally, if the constraint matrix represents only scale normalization, then this ratio formulation can be directly solved by eigenvalue decomposition. However, for a more general constraint matrix, it can be approximately solved with Generalized Eigenvalue Decomposition by transforming the objective function into a more tractable approximate form $\arg\min_{Y} Tr((YBY^T)^{-1}(YLY^T))$.

With the assumption that the low-dimensional vector representations of the vertices can be obtained from a linear projection, i.e., $y_i = \alpha^T x_i$, where α is the projection matrix, then the objective function (11) can be changed to

$$\alpha^* = \arg\min_{\alpha} \frac{tr(\alpha^T XLX^T \alpha)}{tr(\alpha^T XBX^T \alpha)} \quad (12)$$

During the past few years, a number of manifold learning based feature extraction methods have been proposed to capture the intrinsic geometry property [31, 32, 34,35,36,37], In [30], Yan et al. claimed that all of the mentioned manifold learning algorithms can be mathematically unified within the graph embedding framework described in this subsection. They also

proposed Marginal Fisher Analysis which takes both the manifold geometry and the class information into consideration. However, Marginal Fisher Analysis still suffers from the "*Small Sample Size"* problem when the training samples are insufficient, which is always the case in image retrieval.

III. BMMA AND SEMIBMMA FOR SVM RF IN CBIR

With the observation that *"all positive examples are alike; each negative example is negative in its own way"*, the two groups of feedbacks have distinct properties for CBIR [20]. However, the traditional SVM RF treats the positive and negative feedbacks equally.

To alleviate the performance degradation when using the SVM as a RF scheme for CBIR, we explore solutions based on the argument that different semantic concepts lie in different subspaces and each image can lie in many different concept subspaces [20]. We formally formulate this problem into a general subspace learning problem and propose a BMMA for the SVM RF scheme. In the reduced subspace, the negative feedbacks, which differ in diverse concepts with the query sample, are separated by a maximum margin from the positive feedbacks, which share a similar concept with the query sample. Therefore, we can easily map the positive and negative feedbacks onto a semantic subspace in accordance with human perception of the image contents.

To utilize the information of unlabelled samples in the database, we introduced a Laplacian regularizer to the BMMA, which will lead to SemiBMMA for the SVM RF. The resultant Laplacian regularizer is largely based on the notion of local consistency which was inspired by the recently emerging manifold learning community and can effectively depict the weak similarity relationship between unlabeled samples pairs.

Then, the remaining images in the database are projected onto this resultant semantic subspace and a similarity measure is applied to sort the images based on the new representations. For the SVM based RFs, the distance to the hyperplane of the classifier is the criterion to discriminate the query relevant samples from the query irrelevant samples. After the projection step, all positive feedbacks are clustered together while negative feedbacks are well separated from positive feedbacks by a maximum margin. Therefore, the resultant SVM classifier hyperplane in this subspace will be much simpler and better than in the original high dimensional feature space.

Different from the classical subspace learning methods, e.g., PCA and LDA, which can only see the linear global Euclidean structure of samples, BMMA aims to learn a projection matrix α such that in the projected space, the positive samples have high local within-class similarity, but the samples with different labels have high local between-class separability. To describe the algorithm clearly, first we introduce some notations of this approach.

In each round of feedback iteration, there are n samples $X = \{x_1, x_2, \ldots, x_n\} \in R^h$. For simplicity, we assume that the first n^+ samples are positive feedbacks $x_i (1 \le i \le n^+)$, the next n^- samples are negative feedbacks $x_i (n^+ + 1 \le i \le n^+ + n^-)$, and all the others are unlabelled samples $x_i (n^+ + n^- + 1 \le i \le n)$.

Let $l(x_i)$ be the class label of sample x_i , we denote $l(x_i) = 1$ for positive feedbacks, $l(x_i) = -1$ for negative feedbacks and $l(x_i) = 0$ for unlabelled samples. To better show the relationship between the proposed approaches and the graph embedding framework, we use the similar notations and equations in the original graph embedding framework, which provides us a general platform to develop various new approaches for dimensionality reduction. Firstly, two different graphs are formed: 1) the intrinsic graph G , which characterizes the local similarity of the feedback samples; 2) the penalty graph G^p , which characterizes the local discriminant structure of the feedback samples.

For all the positive feedbacks, we first compute the pair-wise distance between each pair of positive feedbacks. Then for each positive feedback x_i , we find its k_1 nearest neighborhood positive feedbacks, which can be represented as a sample set N_i^s , and put an edge between x_i and its neighborhood positive feedbacks. Then the intrinsic graph is characterized as follows:

$$\begin{aligned} S_I &= \sum_i \sum_{j:\, j \in \mathbb{N}_i^s or i \in \mathbb{N}_j^s} \| \alpha^T x_i - \alpha^T x_j \|^2 * W_{ij} \\ &= 2tr[\alpha^T X (D - W) X^T \alpha] \end{aligned} \tag{13}$$

$$W_{ij} = \begin{cases} 1/|\mathbb{N}^s|, \ if \ \ l(i) = 1 and \ \ l(j) = 1, \ i \in \mathbb{N}_j^s or \ j \in \mathbb{N}_i^s \\ 0, \ else \end{cases} \tag{14}$$

where D is a diagonal matrix whose diagonal elements are calculated by $D_{ii} = \sum_j W_{ij}$; $|\mathbb{N}^s|$ denotes the total number of k_1 nearest neighborhood positive sample pairs for each positive feedback. Basically, the intrinsic graph measures the total average distance of the $|\mathbb{N}^s|$ nearest neighborhood sample pairs, and is used to characterize the local within-class compactness for all the positive feedbacks.

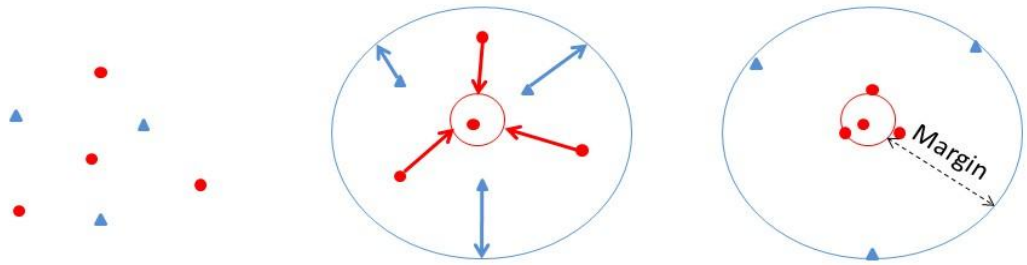
For the penalty graph G^p , its similarity matrix W_{ij}^p represents geometric or statistical properties to be avoided and is used as a constraint matrix in the graph embedding framework. In the BMMA, the penalty graph G^p is constructed to represent the local separability between the positive class and the negative class. More strictly speaking, we expect that the total average margin between the sample pairs with different labels should be as large as possible.

For each feedback sample, we find its k_2 neighbor feedbacks with different labels and put edges between corresponding pairs of feedback samples with weights W_{ij}^p .

Then, the penalty graph can be formed as follows:

$$\begin{aligned} S_p &= \sum_i \sum_{j:\, j \in \mathbb{N}_i^p or i \in \mathbb{N}_j^p} \| \alpha^T x_i - \alpha^T x_j \|^2 * W_{ij}^p \\ &= 2tr[\alpha^T X (D^p - W^p) X^T \alpha] \end{aligned} \tag{15}$$

$$W_{ij}^p = \begin{cases} 1/|\mathbb{N}^p|, \ if \ \ l(i) = 1 and \ \ l(j) = -1, \ i \in \mathbb{N}_j^p or \ j \in \mathbb{N}_i^p \\ 0, \ else \end{cases} \tag{16}$$



(a) (b) (c)

Fig.2 (a) red dots are positive samples and blue dots are negative samples in the original space (b) the BMMA approach (c) the positive samples and negative samples in the maximum margin subspace

where D^p is a diagonal matrix whose diagonal elements are calculated by $D_{ii}^p = \sum_j W_{ij}^p$; $|\mathbb{N}^p|$ denotes the total number of k_2 neighborhood sample pairs with different labels. Similarly, the penalty graph measures the total average distance of the $|\mathbb{N}^p|$ nearest neighbor sample pairs in different class, and is used to characterize the local between-class separability.

In the following, we describe how to utilize the graph embedding framework to develop algorithms based on the designed intrinsic and penalty graphs. Different from the original formulation of the graph embedding framework in [30], the BMMA algorithm optimizes the objective function in a trace difference form instead, i.e.,

$$\begin{aligned}\alpha^* &= \arg\max_{\alpha} 2tr[\alpha^T X(D^p - W^p)X^T\alpha] \\ &\quad - 2tr[\alpha^T X(D-W)X^T\alpha] \\ &= \arg\max_{\alpha} tr[\alpha^T X(D^p - W^p)X^T\alpha] \\ &\quad - tr[\alpha^T X(D-W)X^T\alpha] \\ &= \arg\max_{\alpha} tr(\alpha^T XBX^T\alpha) - tr(\alpha^T XLX^T\alpha) \\ &= \arg\max_{\alpha} tr[\alpha^T X(B-L)X^T\alpha]\end{aligned} \tag{17}$$

As given in Equation (17), we can notice that the objective function works in two ways, which tries to maximize $tr(\alpha^T XBX^T\alpha)$ and at the same time minimize $tr(\alpha^T XLX^T\alpha)$. Intuitively, we can analyze the meaning of the objective function in (17) geometrically. By formulating the objective function as a trace difference form, we can regard it as the total average local margin between positive samples and negative samples. Therefore, Equation (17) can be used as a criterion to discriminate the different classes. In [38], the maximum margin criterion (MMC) was presented as an objective function with a similar difference form. The differences between BMMA and MMC are the definitions of the interclass separability and intraclass compactness. In MMC, both of the interclass separability and intraclass compactness are defined as the same in LDA, which treats the two different classes equally, and MMC can only see the linear global Euclidean structure. In BMMA, the intraclass compactness is constructed by only considering one class (e.g., positive feedbacks) and characterized by a sum of the distances between each positive sample and its k_1 nearest neighbors in the same class. The interclass separability is defined by resorting to interclass marginal samples instead of the mean vectors of different classes as in MMC. Without prior information on data distributions, BMMA can find more reliable low-dimensional representations of the data compared to the MMC [38] and also follow the original assumption in BDA [20] (i.e., all positive examples are alike; each negative example is negative in its own way). It should be noted that previous methods [39, 40] that followed MMC cannot be directly used for the SVM RF in image retrieval because these methods treat samples in different classes equally.

In order to remove an arbitrary scaling factor in the projection, we additionally require that α is constituted by the unit vectors, i.e., $\alpha_k^T\alpha_k = 1, k = 1, 2\ldots, l$. This means that we need to solve the following constraint optimization.

$$\begin{aligned}&\max_{\alpha} tr(\alpha^T X(B-L)X^T\alpha) \\ &\quad = \sum_{k=1}^{l} \alpha_k^T X(B-L)X^T\alpha_k \\ &s.t.\ \ \alpha_k^T\alpha_k - 1 = 0,\ k = 1, 2, \ldots, l\end{aligned} \tag{18}$$

Note that we may also use other constraints instead. For example, we may require $tr(\alpha^T XBX^T\alpha) = 1$ and then minimize $tr(\alpha^T XLX^T\alpha)$. It is easy to check that the above maximum margin approach with such a constraint in fact results in the traditional Marginal Fisher Analysis (MFA) [30]. The only difference is that Equation (18) involves a constrained optimization problem, whereas the traditional MFA solves an unconstrained optimization problem. The motivation for using the constraint $\alpha_k^T\alpha_k = 1,\ k = 1, 2, \ldots, l$ is to avoid calculating the inverse of XBX^T , which leads to the potential "*Small Sample Size"* problem. In order to solve the above constraint optimization problem, we introduce a Lagrangian

$$L(\alpha_k, \lambda_k) = \sum_{k=1}^{l} \alpha_k^T X(B-L)X^T\alpha_k - \lambda_k(\alpha_k^T\alpha_k - 1) \tag{19}$$

with the multipliers λ_k . The Lagrangian L should be maximized with respect to both λ_k and α_k . The condition that at the stationary point, the derivatives of L respect to α_k must vanish, i.e.,

$$\frac{\partial L(\alpha_k, \lambda_k)}{\partial \alpha_k} = (X(B-L)X^T - \lambda_k I)\alpha_k = 0,\ k = 1, 2, \ldots, l \tag{20}$$

and therefore,

$$X(B-L)X^T\alpha_k = \lambda_k\alpha_k,\ k=1,2,\ldots,l \tag{21}$$

which means that the λ_k's are the eigenvalues of $X(B-L)X^T$ and α_k's are the corresponding eigenvectors. Thus, we have

$$J(\alpha)=\sum_{k=1}^{l}\alpha_k^T X(B-L)X^T\alpha_k=\sum_{k=1}^{l}\lambda_k\alpha_k^T\alpha_k=\sum_{k=1}^{l}\lambda_k \tag{22}$$

Therefore, the objective function is maximized when α is composed of the largest eigenvectors of $X(B-L)X^T$. Here, by imposing constraint $\alpha_k^T\alpha_k=1, k=1,2,\ldots,l$, we need not calculate the inverse of XBX^T, and this allows us to avoid the "*Small Sample Size"* problem easily.

The BMMA can be illustrated in Fig.2.

In the previous subsection, we have formulated the BMMA algorithm and shown that the optimal projection matrix α can be obtained by Generalized Eigenvalue Decomposition on a matrix. Then the problem is how to determine an optimal dimensionality for RF, i.e., the projected subspace. To achieve such a goal, we give the detail of determining the optimal dimensionality.

In general,

$$\max_{\alpha} tr(\alpha^T X(B-L)X^T\alpha)=\sum_{i=1}^{l}\lambda_i \tag{23}$$

where λ_i's are the associated eigenvalues and we have

$$\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_{d-1}\geq 0\geq\lambda_d\geq\ldots\geq\lambda_l \tag{24}$$

To maximize the margin between the positive samples and negative samples, we should preserve all the eigenvectors associated with the positive eigenvalues. However, as indicated in [33], for image retrieval the orthogonal complement components are essential to capture the same concept shared by all positive samples. Based on this observation, we should also preserve the components associated with zero eigenvalues although they do not contribute to maximize the margin. This technique can effectively preserve more geometry properties of the feedbacks in the original high-dimensional feature space. Therefore, the optimal dimensionality of the projected subspace just corresponds to the number of nonnegative eigenvalues of the matrix. Therefore, compared to the original formulation of the graph embedding framework in [30], the new formulation (18) can easily avoid the intrinsic "Small Sample Size" problem and also provide us with a simple way to determine the optimal dimensionality for this subspace learning problem.

Biased Discriminant Analysis (BDA) and its kernel version BiasMap [20] were first proposed to address the asymmetry between the positive and negative samples in interactive image retrieval. However, to use BDA, the "*Small Sample Size*" problem and Gaussian assumption for positive feedbacks are two major challenges. While the kernel method BiasMap cannot exert its normal capability since the feature dimensions are much higher than the number of training samples. Additionally, it is still problematic to determine the optimal dimensionality of BDA and BiasMap for CBIR. Different from the original BDA, our BMMA algorithm is a local discriminant analysis approach, which does not make any assumption on the distribution of the samples. Biased

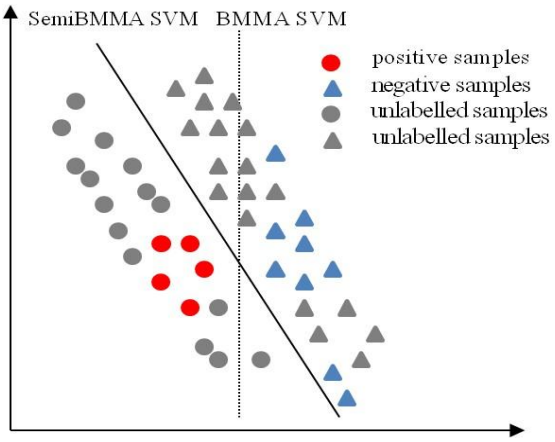


Fig.3 An illustration of the SVM hyperplance comparison between BMMA SVM and SemiBMMA SVM for two classes of feedbacks

towards the positive samples, maximizing the objective function in the projected space can push the nearby negative samples away from the positive samples while pulling the nearby positive samples towards the positive samples. Therefore, the definition in Equation (18) can maximize the overall average margin between the positive samples and negative samples. In such a way, each sample in the original space is mapped onto a low-dimensional local maximum margin subspace in accordance with human perception of the image contents.

Since the graph embedding technique is an effective way to capture the intrinsic geometry structure in the original feature space, we propose a way to incorporate the unlabelled samples based on the intrinsic graph, which is helpful in capturing the manifold structure of samples and alleviating the over fitting problem. In the following, we design a regularization term based on intrinsic graph for the unlabelled samples in the image database.

For each unlabelled sample $x_i (n^+ + n^- + 1 \leq i \leq n)$, we expect that the nearby unlabelled samples are likely to have the similar low-dimensional representations. Specifically, for each unlabelled sample, we find its k_1 nearest neighborhood unlabelled samples, which can be represented as a sample set N_i^u, and put an edge between the unlabelled x_i and its neighborhood unlabelled samples. Then the intrinsic graph for the unlabelled samples is characterized as follows:

$$\begin{aligned} S_U &= \frac{1}{2}*\sum_i \sum_{j:\, j\in\mathbb{N}_i^s or i\in\mathbb{N}_j^s} \| \alpha^T x_i - \alpha^T x_j \|^2 * W_{ij}^u \\ &= tr[\alpha^T X(D^u - W^u)X^T\alpha] \\ &= tr[\alpha^T XUX^T\alpha] \end{aligned} \tag{25}$$

$$W_{ij}^u = \begin{cases} \frac{1}{|\mathbb{N}|^n}\exp(-\| x_i - x_j \|^2/\delta^2), & if\ \ l(i)=l(j)=0, i\in\mathbb{N}_j^u or\ j\in\mathbb{N}_i^u \\ 0, & else \end{cases} \tag{26}$$

which reflects the affinity of the sample pairs; D^u is a diagonal matrix whose diagonal elements are calculated by $D_{ii}^u=\sum_j W_{ij}^u$; $|D^u|$ denotes the total number of k_1 nearest neighborhood unlabelled sample pairs for each unlabelled sample. $L^u = D^u - W^u$ can be known as a Laplacian matrix.

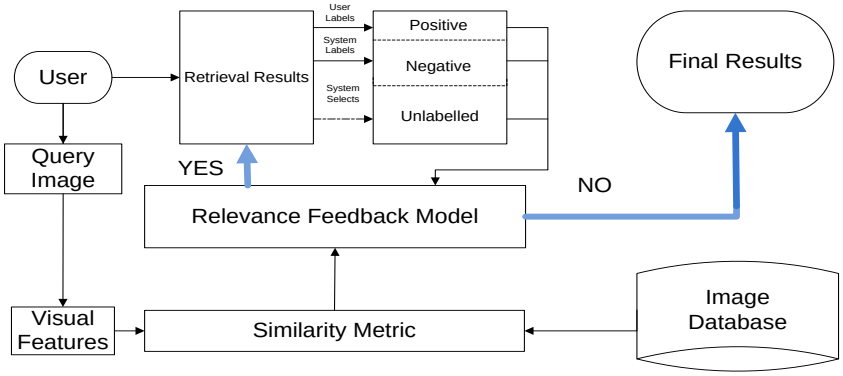


Fig.4 The flowchart of the Content-Based Image Retrieval system

Hence, we call this term as a Laplacian regularizer.

The motivation for introducing this term is inspired by the regularization principle, which is the key to enhancing the generalization and robust performance of the approach in practical applications. There are a lot of possible ways to choose a regularizer for the proposed BMMA. In this work we chose the Laplacian regularizer, which is largely inspired by the recently emerging manifold learning community. Actually, this scheme can preserve weak (probably correct) similarities between all unlabeled sample pairs and thus effectively integrate the similarity information of unlabeled samples into the BMMA. By integrating the Laplacian regularizer into the supervised BMMA, we can easily obtain the SemiBMMA for the SVM RF, i.e.,

$$\alpha^* = \arg\max_{\alpha} tr[\alpha^T X(B-L-\beta*U)X^T\alpha] \qquad (27)$$

where β is used to trade off the contributions of the labeled samples and unlabelled samples. Similarly, the solution of (27) is obtained by conducting the Generalized Eigenvalue Decomposition and α is calculated as a set of eigenvectors.

The difference between BMMA SVM and SemiBMMA SVM for two classes of feedbacks can be shown in Fig.3. The SemiBMMA algorithm is illustrated in Table 1.

Table 1 Semi-Supervised Biased Maximum Margin Analysis

Input: $X=\{x_1,x_2,\ldots x_n\}\in R^h$ stand for all the feedback samples and unlabelled samples, which include the positive sample set X^+, the negative sample set X^-, and unlabelled samples.

1) Construct the supervised intrinsic graph G, according to the formulation (14) and calculate the matrix value XLX^T.

2) Construct the supervised penalty graph G^p, according to the formulation (16) and calculate the matrix value XBX^T.

3) Construct the Laplacian regularizer according to the formulation (26) and calculate the matrix value XUX^T.

4) Calculate the projection matrix α^* according to Generalized Eigenvalue Decomposition on the matrix $X(B-L-\beta*U)X^T$.

5) Calculate the new representations: project all positive, negative and remaining samples in the database onto the reduced subspace respectively, i.e., $Y^+=\alpha^{*T}X^+$, $Y^-=\alpha^{*T}X^-$.

Output: Positive and negative samples, Y^+ and Y^-, in this reduced subspace.

Then all the unlabeled samples in the database are projected onto this subspace. After the projection, the traditional SVM RF is executed on the new representations. Finally, similar to the traditional SVM RF, we can measure the degree of relevance through the output of SVM, i.e., $|f(x)|$.

IV. The Content-Based Image Retrieval System

In experiments, we use a subset of the Corel Photo Gallery as the test data to evaluate the performance of the proposed scheme. The original Corel Photo Gallery includes plenty of semantic categories, each of which contains 100 or more images. However, some of the categories are not suitable for image retrieval, since some images with different concepts are in the same category and many images with the same concept are in different categories. Therefore, the existing categories in the original Corel Photo Gallery are ignored and reorganized into 80 conceptual classes based on the ground truth, such as lion, castle, aviation, train, dog, autumn, cloud, tiger, etc. Finally, the test database comprises totally 10,763 real-world images.

Given a query image by the user, the CBIR system is expected to feed back more semantically relevant images after each feedback iteration [2]. However, during RF, the number of the relevant images is usually very small because of the semantic gap. At the same time, the user would not like to label a large number of samples. The user also expects to obtain more relevant images with only a few rounds of RF iterations. Keeping the size of labeled relevant images small and the relevance feedback iterations few are two key issues in designing the image retrieval system. Therefore, we devise the following CBIR framework accordingly to evaluate the RF algorithms.

From the flowchart in Fig.4, we can notice that when a query image is provided by the user, the image retrieval system first extracts the low-level features. Then all the images in the database are sorted based on a similarity metric, i.e. Euclidean distance. 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 retrieval results. Then she/he

will label the most semantically relevant images as positive feedbacks in top retrieval results. All of the remaining images in top results are automatically labeled by the system as the negative feedbacks. Based on the small size positive and negative feedbacks, the RF model can be trained based on various existing techniques. Then all the images in the database are resorted based on a new similarity metric. After each round of retrieval, the user will check whether the results are satisfied. If the user is satisfied with the results, then the process is ended; otherwise, the feedback process repeats until the user is satisfied with the retrieval results.

Generally, the image representation is a crucial problem in CBIR. The images are usually represented by low level features, such as color [3-5], texture [5-7] and shape [8-10], each of which can capture the content of an image to some extent.

For color, we extracted three moments: color mean, color variance, and color skewness in each color channel (L, U, V) respectively. Thus, a 9-dimensional color moment is employed as the color features in our experiments to represent the color information. Then a 256-dimensional (8*8*4) HSV color histogram is calculated. Both hue and saturation are quantized into 8 bins and the values are quantized into 4 bins. These two kinds of visual features are formed as color features.

Comparing with the classical global texture descriptors (e.g., Gabor features, wavelet features), the local dense features show good performance in describing the content of an image. The Webber Local Descriptors (WLD) [41] are adopted as feature descriptors which are mainly based on the mechanism of the human perception of a pattern. The WLD local descriptor results in a feature vector of 240 values.

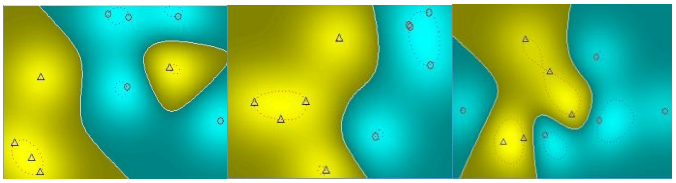
We employ the edge directional histogram [8] from the Y component in YCrCb space to capture the spatial distribution of edges. The edge direction histogram is quantized into five categories including horizontal, 45° diagonal, vertical, 135° diagonal and isotropic directions to represent the edge features.

Generally, these features are combined into a feature vector, which results in a vector with 510 values (i.e., 9+256+240+5=510). Then all feature components are normalized to normal distributions with zero mean and one standard deviation to represent the images.

V. Experimental Results on a Real World Image Database

A The intrinsic problems in the traditional SVM RF

An image is usually represented as a high-dimensional feature vector in CBIR. However, one key issue in RF is that which subset of features can reflect the basic properties of different groups of feedback samples and benefit the construction of the optimal classifier. This problem can be illustrated from some real-world data in relevance feedback. There are five positive samples and five negative feedback samples. We randomly select two features to construct the optimal SVM hyperplane for three times. As shown in Fig.5,

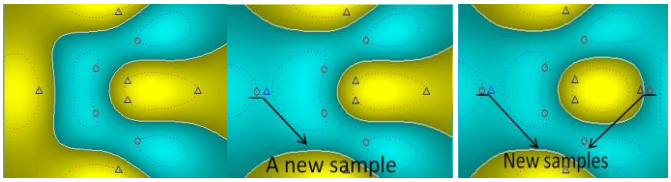


(a) (b) (c)

Fig.5 The SVM hyper plane is diverse for different combinations of features.

we can see that the resultant SVM classifiers are diverse with different combinations of features.

It is essential to obtain a satisfactory classifier when the number of available feedback samples is small, which is always the case in RF, especially in the first few rounds of feedbacks. Therefore, we first show a simple example to simulate the unstable problem of SVM when dealing with a small number of training samples. The open circles in Fig.6 indicate the positive feedback samples and the plus points indicate the negative samples in relevance feedback. The Fig.6 (a) shows an optimal hyperplane, which is trained by the original training samples. Fig.6 (b) and (c) show a different optimal hyperplane, which are trained by the original training set with only one and two incremental positive sample respectively. From Fig.6, we can see that the hyperplane of the SVM classifier changes sharply when a new incremental sample is integrated into the original training set. Additionally, we can also note that the optimal hyper planes of SVM are much complex when the feedbacks have a complicated distribution.



(a) (b) (c)

Fig. 6 The SVM hyper plane is unstable and complex when dealing with small size of training set

Note that the similar results have been indicated in the previous research [25]. However, in this section, we have shown slightly different problems in the traditional SVM RF, that is, distinct property of feedback samples in RF and unstable and complex hyper planes of the traditional SVM in the first few rounds of feedbacks.

B Features extraction based on different methods

Six experiments are conducted for comparing the BMMA with the traditional LDA, BDA method and a Graph embedding approach MFA, in finding the most discriminative directions. We plot the directions which correspond to the largest eigenvalue of the decomposed matrices for LDA, BDA, MFA and BMMA respectively. From these examples, we can clearly notice that LDA can find the best discriminative direction when the data from each class are distributed as Gaussian with similar covariance matrices, as shown in Fig.7 (a) and (d), but it may confuse when the data distribution is more complicated, as given in Fig.7(b), (c), (e) and (f). Biased towards the positive samples, BDA can find the direction that

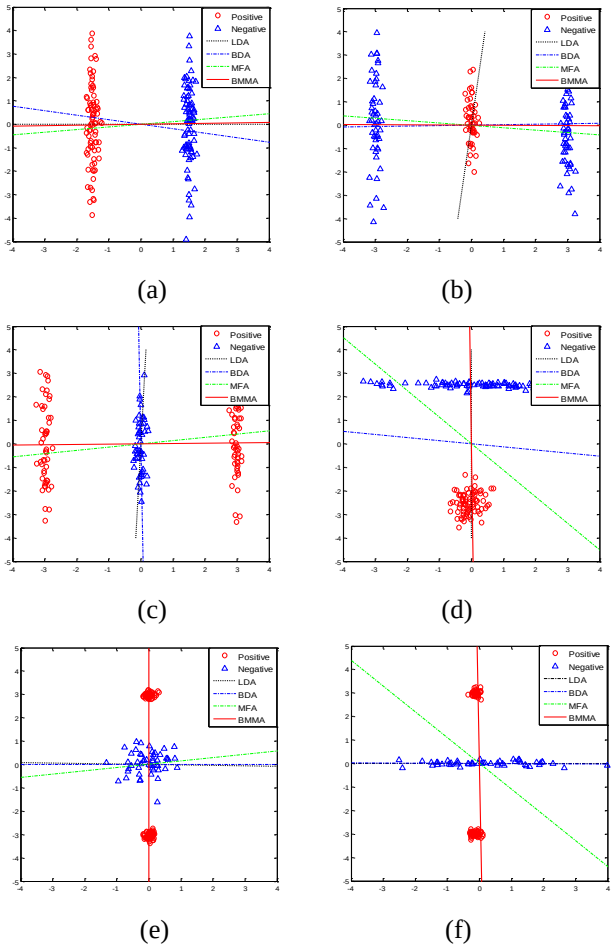


Fig.7 Four feature extraction methods (i.e., LDA, BDA, MFA and BMMA) for two classes of samples (i.e., positive samples and negative samples) with different distribution

the positive samples are well separated with the negative samples when the positive samples have Gaussian distribution, but it may also confuse when the distribution of the positive samples is more complicated. For instance, in Fig.7 (b), BDA can find the direction for distinguishing positive samples from negative ones. However, in Fig.7 (c), (e) and (f), when the positive samples have more complicated distribution, the BDA algorithm obviously fails. MFA can also find the discriminative direction when the distribution of negative feedbacks is simple, as shown in Fig.7 (a), (b), (c). But when the negative samples pose a more complicated distribution, MFA will fails as in Fig.7 (d), (e), (f). Biased towards positive samples, the BMMA method can find the most discriminative direction for all the 6 experiments based on local analysis, since it doesn't make any assumptions on the distributions of the positive and negative samples. It should be noted that BMMA is a linear method and therefore, we only gave the comparison results of linear methods above.

C Statistical experimental results

In this section, we evaluate the performance of the proposed scheme on a real world image database. We use precision-scope curve, precision rate and standard deviation to evaluate the effectiveness of the image retrieval algorithms. The scope is specified by the number N of top-ranked images presented to the user. The precision is the major evaluation criterion, which evaluates the effectiveness of the algorithms. The precision-scope curve describes the precision with various scopes and can give the overall performance evaluation of the approaches. Precision rate is the ratio of the number of relevant images retrieved to the top N retrieved images, which emphasizes the precision at a particular value of scope. Standard deviation describes the stability of different algorithms. Therefore, the precision evaluates the effectiveness of a given algorithm and the corresponding standard deviation evaluates the robustness of the algorithm. We empirically select the parameters $k_1, k_2 = 4$ according to manifold learning approaches. Considering the computable efficiency, we randomly select 300 unlabeled samples in each round of feedback iteration. For the trade off parameter between labeled samples and unlabeled samples, we simply set $\beta = 1$. For all the SVM-based algorithms, we choose the Gaussian kernel:

$$K(x, y) = e^{-\rho|x-y|^2}, \rho = 0.001 \quad (28)$$

Note that, the kernel parameters and kernel type can significantly affect the performance of retrieval. For different image database, we should tune the kernel parameters and kernel type carefully. In our experiments, we determine the kernel parameters from a series of values according to the performance. Moreover, much better performance can be achieved by tuning the kernel parameters further for different queries.

1) Experiments on a small size image database

Fig.8 Example categories used in the small size image database

In order to show how efficient the proposed BMMA combined with SVM in dealing with the asymmetry properties of feedback samples, the first evaluation experiment is executed on a small size database, which includes 3899 images with 30 different categories. We use all 3899 the images in 30 categories as queries. Some example categories used in experiments are shown in Fig.8. To avoid the potential problem caused by the asymmetry amount of positive and negative feedbacks [25], we selected equal number of positive and negative feedbacks in this subsection. In practice, the first 5 query relevant images and first 5 irrelevant images in the top 20 retrieved images in the previous iterations were automatically selected as positive and negative feedbacks respectively.

In [33], OCCA was proposed to only analyze the positive feedbacks for SVM RF in a retrieval task. Hence, we compared the RF performance of BMMA combined with SVM (BMMA SVM), OCCA combined with SVM (OCCA SVM) and the traditional SVM in this subsection.

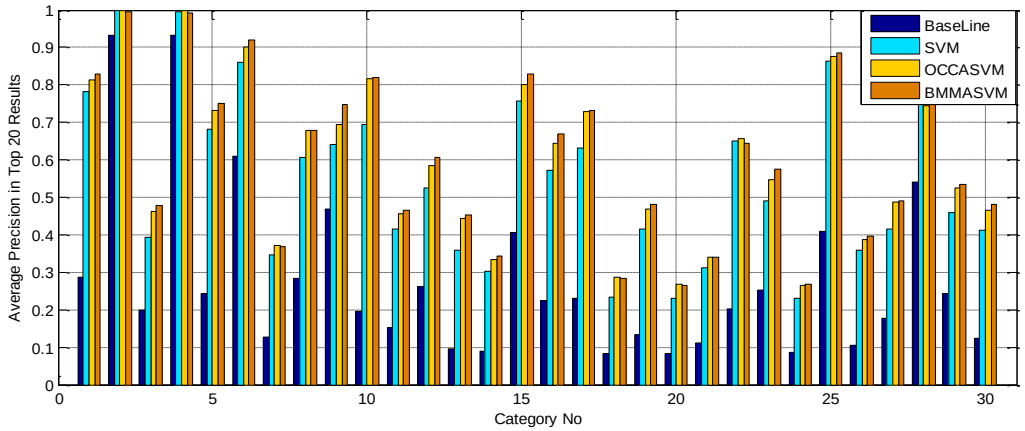
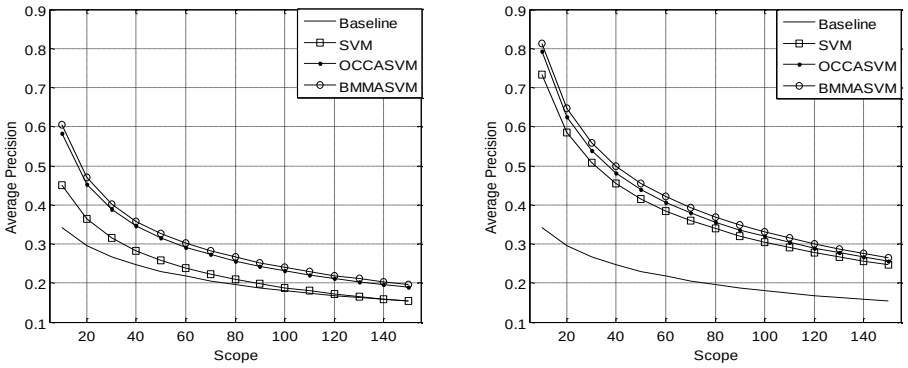


Fig.9 Average precisions in the top 20 results of SVM, OCCA SVM and BMMA SVM after 2 rounds of feedback



(a) (b)

Fig.10 The precision-scope curves after the 1st feedback and 2nd feedback for SVM, OCCA SVM and BMMA SVM

In real world, it is not practical to require the user to label many samples. Therefore, small size of training samples will cause the severe unstable problem in SVM RF (as shown in Section V. A). Fig.9 shows the precisions in the top 20 after the 2nd round of feedback iteration for all the 30 categories. The baseline curve describes the initial retrieval results without any feedback information. Specially, at the beginning of retrieval, the Euclidean distances in the original high dimensional space are used to rank the images in the database. After the user provides relevance feedbacks, the traditional SVM, BMMA SVM and OCCA SVM algorithms are then applied to sort the images in the database. As can be seen, the retrieval performance of these algorithms varies with different categories. For some easy categories, all the algorithms can perform well (for Categories 2, 4 even the baseline can achieve over 90% for precision). For some hard categories, all the algorithms perform poorly (e.g., Categories 18, 20, 24). After two rounds of feedbacks, all the algorithms are significantly better than the baseline, and this indicates that the relevance feedbacks provided by the user are very helpful in improving the retrieval performance.

Fig.10 shows the average precision-scope curves of the algorithms for the 1st and 2nd iterations. We can notice that both the BMMA SVM and the OCCA SVM can perform much better the traditional SVM on the entire scope, especially the 1st round of feedback. The main reason is that in the first round of feedback iteration, the number of training samples is especially small (usually 8-10 training samples totally), and this will make SVM perform extremely poorly. The BMMA SVM algorithm and the OCCA SVM algorithm can significantly improve the performance of SVM by treating the positive and negative feedbacks unequally. Therefore, we can conclude that the technique, which asymmetrically treats the feedback samples (i.e., biased towards the positive feedbacks), can significantly improve the performance of SVM RF which treats the feedback samples equivalently. As shown in Fig.10 (b), with the number of feedbacks increasing, the performance difference between the enhanced algorithms and the traditional SVM gets small. Generally, by iteratively adding the user's feedbacks, more samples will be fed back as training samples, and will make the performance of SVM much more stable. Meanwhile, the dimension of the BMMA and OCCA decreases with the increasing of the positive feedbacks. Consequently, the performance of BMMA SVM and OCCA SVM will be degraded by over fitting. Therefore, the performance difference between the enhanced algorithms and the traditional SVM gets small. However, the performance of the first a few rounds of feedbacks is usually most important, since the user would not like to provide more rounds of feedback iteration. In the first a few rounds of feedbacks, the classifier trained based on few labeled training samples is not reliable, but its performance can

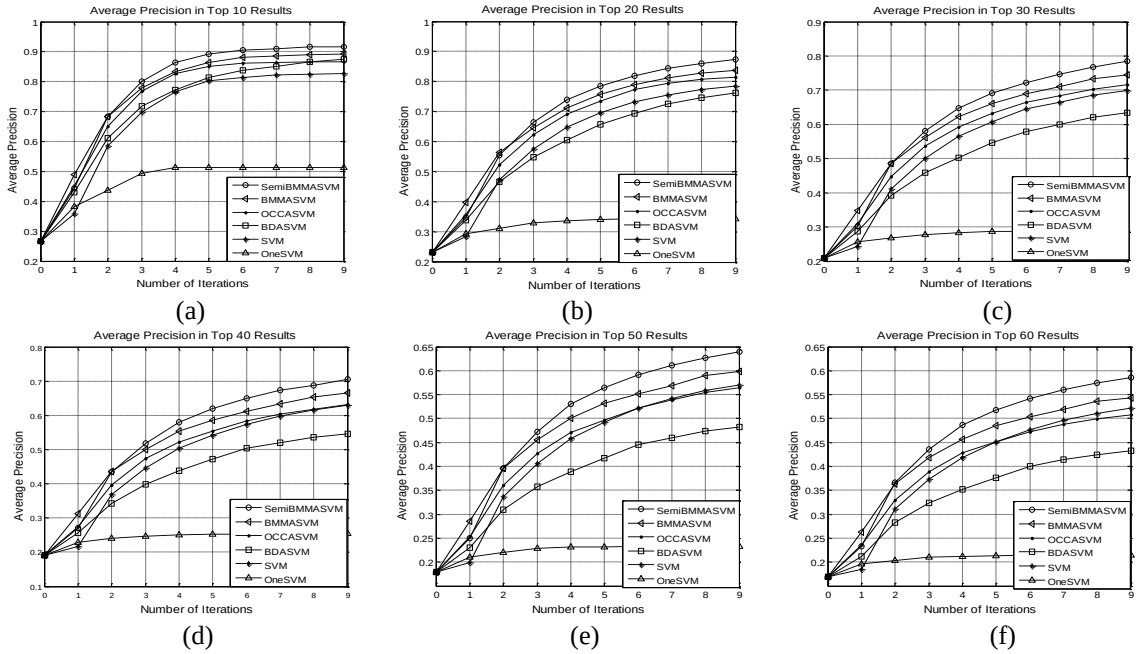


Fig.11 The average precisions in top 10- top 60 results of the six approaches from the fivefold cross validation.

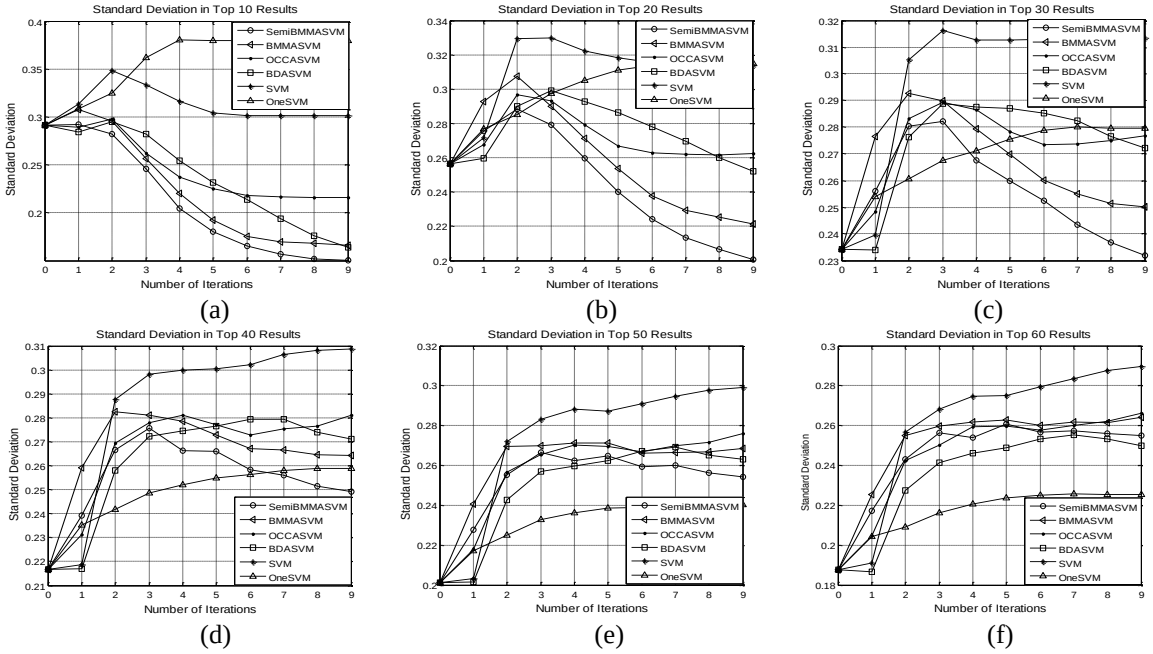


Fig.12 The standard deviations in top 10- top 60 results of the six approaches from the fivefold cross validation.

be improved when more images are labeled by the user in the subsequent feedback iterations. Both BMMA and OCCA can significantly improve the performance of the traditional SVM in the first two iterations. Therefore, we can conclude that BMMA can effectively integrate the distinct properties of two groups of feedback samples into the retrieval process and thus enhance the performance.

2) Experiments on a large scale image database

We designed a slightly different feedback scheme to model the real world retrieval process. In a real image retrieval system, a query image is usually not in the image database. To simulate such an environment, we use fivefold cross validation to evaluate the algorithms. More precisely, we divide the whole image database into five subsets of equal size. Thus, there are 20 percent images per category in each subset. At each run of cross validation, one subset is selected as the query set, and the other four subsets are used as the database for retrieval. Then 400 query samples are randomly selected from the query subset and the relevance feedback is automatically implemented by the system. For each query image, the system retrieves and ranks the images in the database and 9 RF iterations are automatically executed.

At each iteration of the relevance feedback process, top 20 images are picked from the database and labeled as relevant

Table 2 Average precisions in top *N* results of the six algorithms after the 9th feedback iteration (mean ± standard deviation)

Method	SemiBMMASVM	BMMASVM	OCCASVM	BDASVM	SVM	OneSVM
Top10	**0.9165** ±0.1505	0.8914 ±0.1663	0.8658 ±0.2162	0.8747 ±0.1638	0.8265 ±0.3011	0.5122 ±0.3795
Top20	**0.8727** ±0.2008	0.8363 ±0.2213	0.8154 ±0.2622	0.7612 ±0.2520	0.7856 ±0.3138	0.3428 ±0.3149
Top30	**0.7840** ±0.2319	0.7452 ±0.2502	0.7151 ±0.2767	0.6339 ±0.2722	0.6996 ±0.3134	0.2878 ±0.2797
Top40	**0.7058** ±0.2491	0.6645 ±0.2642	0.6315 ±0.2811	0.5465 ±0.2712	0.6290 ±0.3088	0.2546 ±0.2588
Top50	**0.6396** ±0.2542	0.5982 ±0.2687	0.5638 ±0.2761	0.4829 ±0.2631	0.5706 ±0.2993	0.2331 ±0.2404
Top60	**0.5854** ±0.2548	0.5431 ±0.2640	0.5078 ±0.2660	0.4319 ±0.2497	0.5216 ±0.2847	0.2147 ±0.2255
Top70	**0.5416** ±0.2523	0.4964 ±0.2549	0.4629 ±0.2544	0.3925 ±0.2362	0.4806 ±0.2783	0.2000 ±0.2115
Top80	**0.5025** ±0.2460	0.4564 ±0.2430	0.4246 ±0.2407	0.3596 ±0.2229	0.4462 ±0.2687	0.1886 ±0.2000
Top90	**0.4865** ±0.2387	0.4224 ±0.2298	0.3941 ±0.2287	0.3320 ±0.2102	0.4166 ±0.2581	0.1789 ±0.1890

and irrelevant feedbacks. Generally, in real world retrieval systems, the negative samples usually largely outnumber the positive ones. To simulate such a case in the retrieval system, the first 3 relevant images are labeled as positive feedbacks and all the other irrelevant images in top 20 results are automatically marked as negative feedbacks. Note that, the images which have been selected at the previous iterations are excluded from later selections. The experimental results are shown in Fig.11 and Fig.12. The average precision and standard deviation are computed from the fivefold cross validation.

To demonstrate the effectiveness of the proposed scheme, we compare them with the traditional SVM, the OCCA SVM, BDA SVM and one-class SVM (OneSVM). The traditional SVM regards RF as a strict two-class classification problem, with equal treatments on both positive and negative samples. The OCCA tries to find a subspace, in which all positive samples are merged, and then the traditional SVM are implemented to retrieve the relevant images to the query image. For BDA, we select all the eigenvectors with the eigenvalues larger than one percent of the maximum eivenvalues and then the traditional SVM are used to classify the relevant and irrelevant images, which is the common way to select the dimension of the subspace. OneSVM assumes RF as a one-class classification problem and estimates the distribution of the target images in the feature space.

Fig.11 and Fig.12 show the average precision and the standard deviation curves of different algorithms respectively. SemiBMMA SVM outperforms all the other algorithms on the entire scope. Both BMMA and OCCA can improve the performance of the traditional SVM RF, as shown in Fig.11 (a), (b), (c) and (d). Comparing with OCCA SVM, the BMMA SVM performs much better for all the top results, since BMMA takes both the positive and negative feedbacks into consideration. However, both BMMA and OCCA will encounter the over fitting problem, i.e., both of them combined with SVM will degrade the performance of SVM after a few rounds of feedbacks although they can improve the performance of SVM in the first a few rounds of feedback. As can be seen in Fig.11 (d) (e) (f), with the increase of rounds of feedbacks, the OCCA SVM performs poorly in comparison with the traditional SVM. At the same time, the performance difference between BMMA SVM and the traditional SVM gets smaller. The SemiBMMA combined with SVM can significantly improve the performance of the traditional SVM, since it can effectively utilize the basic property of the different groups of the feedback samples and integrate the information of the unlabelled samples into the construction of the classifier. For top 10 results, the BDA SVM can achieve better result than the traditional SVM RF. However, the BDA algorithm still discards much information contained in the orthogonal complement components of the positive samples. Therefore, BDA combined with traditional SVM performs much worse than the traditional SVM RF, as shown in Fig.11 (b), (c), (d), (e), (f). Although OneSVM tries to estimate the distribution of the target images in the feature space, it cannot work well without the help of negative samples.

Considering the stability of the algorithms, we can also notice that SemiBMMA SVM and BMMA SVM perform best among all the algorithms for top10, top 20 and top 30 results. Although OneSVM shows good stability for top 40, top 50 and top 60 results, its average precision for retrieval is too low.

We should indicate that the performance difference of algorithms between experiments in Subsection V.C 1) and experiments in this subsection is mainly caused by the different experimental setting. Because the number of positive and negative feedbacks is equal in Subsection V.C 1), while negative feedbacks largely outnumber positive feedbacks in subsection 2). Additionally, the performance of SemiBMMA SVM does not perform better comparing with BMMA SVM in the first two rounds of feedback iterations for most of the results. This is mainly because the maximum margin between different classes is essentially important when the number of training samples is extremely small.

The detailed results of all the algorithms after the 9th feedback are shown in Table 2.As can be seen, SemiBMMA combined with SVM integrates all the available information into relevance feedback iteration and achieves much better performance comparing with other approaches for all the top results. The BMMA SVM still obtains satisfactory performance comparing with the traditional SVM and OCCA SVM. Therefore, we can conclude that the proposed BMMA and SemiBMMA combined with the SVM RF have shown much better performance than the traditional SVM RF (i.e., directly using the SVM as a RF scheme) for CBIR.

3) Visualization of the retrieval results

In the previous subsections, we have presented some statistic quantitative results of the proposed scheme. In this

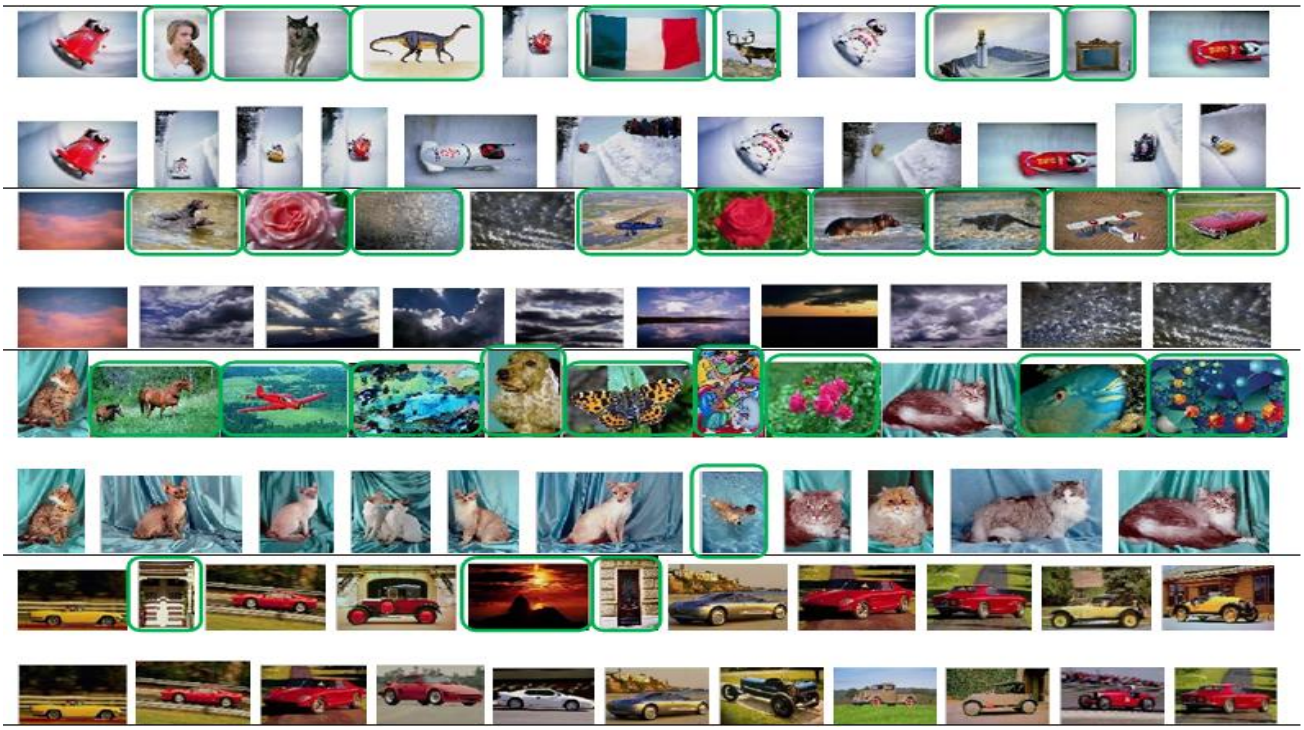


Fig.13 Top 10 results for 4 different query images based on initial results and SemiBMMA SVM after 4 rounds of feedback iteration. Incorrect results are highlighted by green boxes.

subsection, we show the visualization of retrieval results. In experiments, we randomly select some images (e.g., bobsled, cloud, cat and car) as the queries and perform the relevance feedback process based on the ground truth. For each query image, we do 4 RF iterations. For each RF iteration, we randomly select some relevant and irrelevant images as positive and negative feedbacks from the first screen, which contain 20 images in total. The number of selected positive and negative feedbacks is about 4 respectively. We choose them according to the ground truth of the images, i.e., whether they share the same concept with the query image or not. Fig.13 shows the experimental results. The query images are given as the first image of each row. We show the top 1 to top 10 images of initial results without feedback and SemiBMMA SVM after 4 feedback iterations respectively. And incorrect results are highlighted by green boxes. From the results, we can notice that our proposed scheme can significantly improve the performance of the system. For the 1st 2nd and 4th query images, our system produce 10 relevant images out of the top 10 retrieved images. For the 3rd query image, our system produces 9 relevant images out of the top 10 retrieved images. Therefore, SemiBMMA SVM can effectively detect the homogeneous concept shared by the positive samples and hence improve the performance of the retrieval system.

VI. Conclusion And Future Work

Support Vector Machine (SVM) based Relevance Feedback (RF) has been widely used to bridge the semantic gap and enhance the performance of CBIR systems. However, directly using the SVM as a RF scheme has two main drawbacks. First, it treats the positive and negative feedbacks equally although this assumption is not appropriate since all positive feedbacks share a common concept while each negative feedback differs in diverse concepts. Second, it does not take into account the unlabelled samples although they are very helpful in constructing a good classifier. In this paper, we have explored solutions based on the argument that different semantic concepts live in different subspaces and each image can live in many different subspaces. We have designed a Biased Maximum Margin Analysis and a Semi-Supervised Biased Maximum Margin Analysis to alleviate the two drawbacks in the traditional SVM RF. The novel approaches can distinguish the positive feedbacks and negative feedbacks by maximizing the local margin and integrity the information of unlabeled sample by introducing a Laplacian regularizer. Extensive experiments on a large real world Corel image database have shown that the proposed scheme combined with the traditional SVM RF can significantly improve the performance of CBIR systems.

Despite the promising results, several questions remain to be investigated in our future work: First, this approach involves dense matrices eigen decomposition which can be computationally expensive both in time and memory. Therefore, an effective technique for computation is required to alleviate the drawback. Second, theoretic questions need to be investigated regarding how the proposed scheme affects the generalization error of classification models. More specifically, we expect to get a better tradeoff between the integration of the distinct properties of feedbacks and the generalization error of the classifier.

Acknowledgement

The authors would like to thank Prof. James Z. Wang (with the College of Information Sciences and Technology in Pennsylvania State University) for his kindly providing the Corel Image Gallery. The authors would like to thank the

handling Associate Editor and three anonymous reviewers for their constructive comments on this manuscript in the three rounds of review.

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 Trans. Pattern Anal. Mach. Intell.*, Vol. 22, No. 1, pp. 1349–1380, Jan.2000.

[2] 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.

[3] M. J. Swain and D. H. Ballard, Color Indexing, *International. Journal of Compututer. Vision.*, Vol.7, No. 1, pp. 11–32, 1991.

[4] G. Pass, R. Zabih, and J. Miller, Comparing Images Using Color Coherence Vectors, *In Proc. ACM Multimedia*, pp. 65–73. 1996.

[5] 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.

[6] H. Tamura, S. Mori, and T.Yamawaki, Texture Features Corresponding to Visual Perception, *IEEE Trans. Syst., Man, Cybern.*, Vol. 8, No.6, pp. 460–473, Jun. 1978.

[7] J. Mao and A. Jain, Texture Classification and Segmentation using Multiresolution Simultaneous Autoregressive Models, *Pattern Recognit.*,Vol. 25, No. 2, pp. 173–188, 1992.

[8] A. Jain and A.Vailaya, Image Retrieval Using Color and Shape, *Pattern Recognit.*, Vol. 29, No. 8, pp. 1233–1244, Aug. 1996.

[9] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubino, The QBICproject: querying images by content using color, texture, and shape, In *Proc. SPIE, Storage and Retrieval for Images and Video Databases*, pp. 173–181. Feb. 1993.

[10] A. Jain and A. Vailaya, Shape-Based Retrieval: a Case Study with Trademark Image Databases, *Pattern Recognit.*, Vol. 31, No. 9, pp.1369–1390, 1998.

[11] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra, Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval, *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 8, No. 5, pp. 644-655, Sep. 1998.

[12] X. Zhou and T. Huang, Relevance Feedback for Image Retrieval: A Comprehensive Review, *ACM Multimedia Systems J.*, Vol. 8, pp. 536-544, 2003.

[13] Y. Rui, T. S. Huang, and S. Mehrotra, Content-Based Image Retrieval with Relevance Feedback in MARS, in *Proc. IEEE Int. Conf. Image Process.*, 1997, Vol. 2, pp. 815–818.

[14] T. S. Huang and X. S. Zhou. Image Retrieval by Relevance Feedback: from Heuristic Weight Adjustment to Optimal Learning Methods. In *Proc. IEEE International Conference on Image Processing (ICIP'01)*, Thessaloniki,Greece, Oct. 2001.

[15] J. Laaksonen, M. Koskela, and E. Oja. Picsom: Self-Organizing Maps for Content-Based Image Retrieval. *In Proceedings of International Joint Conference on Neural Networks (IJCNN'99)*, Washington, DC, USA, 1999.

[16] Y. Chen, X.-S. Zhou, and T.-S. Huang, One-class SVM for Learning in Image Retrieval, *In Proceedings of IEEE Int. Conf. Image Processing (ICIP'01)*, pp. 815–818, 2001.

[17] C. Hoi, C. Chan, K. Huang, M.R. Lyu and I.King, Biased Support Vector Machine for Relevance Feedback in Image Retrieval. *In Proceedings of International Joint Conference on Neural Networks (IJCNN'04)*, Budapest, Hungary, pp. 3189-3194, 25-29 July, 2004.

[18] X. He, D. Cai, and J. Han Learning a Maximum Margin Subspace for Image Retrieval. *IEEE Trans. on Knowledge and Data Engineering*, Vol. 20, No. 2, Feb. 2008.

[19] W. Bian, 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

[20] X. Zhou and T. Huang, Small Sample Learning during Multimedia Retrieval Using Biasmap, *In Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*,Vol.1, pp.11-17,2001.

[21] 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. Multimedia*, Vol.8, No.4, pp.716-724, August, 2006

[22] 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.

[23] D. Cai, X. He, and J. Han. Regularized Regression on Image Manifold for Retrieval, *In Proceedings of 9th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR'07),* Augsburg, Germany, Sep. 2007.

[24] P. Hong, Q. Tian, and T. S. Huang. Incorporate support vector machines to content-based image retrieval with relevant feedback. In *Proc. IEEE International Conference on Image Processing (ICIP'00)*, Vancouver,BC, Canada, 2000.

[25] D. Tao, X. Tang, X. Li, and X. Wu, Asymmetric Bagging and Random Subspace for Support Vector Machines-Based Relevance Feedback in Image Retrieval, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 28, No. 7, pp. 1088–1099, Jul. 2007.

[26] G. Guo, A.K. Jain, W. Ma, and H. Zhang, Learning Similarity Measure for Natural Image Retrieval with Relevance Feedback, *IEEE Trans. Neural Networks*, Vol. 12, No. 4, pp. 811-820, 2002.

[27] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[28] S. Tong and E. Chang, Support Vector Machine Active Learning for Image Retrieval, *In Proc. ACM Multimedia*, 2001, pp.107–118.

[29] J. Li, N. Allinson, D. Tao and X. Li, Multitraining Support Vector Machine for Image Retrieval. *IEEE Trans. On Image Processing*. Vol.15 No.11, Nov.2006.

[30] 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. Pattern Anal. Mach. Intell.*, Vol. 29, No. 1, pp.40–51, Jan. 2007.

[31] S. T. Roweis, and L. K. Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science*, Vol 290, No.22, pp. 2323-2326, Dec. 2000.

[32] J. Tenenbaum, V. Silva, and J. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, Vol. 290, No. 22, pp. 2319-2323, Dec. 2000.

[33] D. Tao, X. Tang and X. Li, Which Components Are Important for Interactive Image Searching? *IEEE Trans. Circuits System Video Technology*, Vol. 18, No.1, Jan. 2008.

[34] M. Belkin and P. Niyogi. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. *In Advances in Neural Information Processing Systems (NIPS 2001)*, Vol.14, 2001.

[35] X. He, D. Cai, S. Yan, and H.Zhang, Neighborhood Preserving Embedding, *Proc. IEEE Int'l Conf.Computer Vision (ICCV'05)*, pp. 1208-1213, 2005.

[36] X. He and P. Niyogi, Locality Preserving Projections, *In Advances in Neural Information Processing Systems (NIPS 2004),*, Vol. 16, 2004.

[37] D. Cai, X. He, and J. Han Isometric Projection. *In Proceedings of 22nd Conference on Artifical Intelligence (AAAI'07),* Vancouver, Canada, July 2007.

[38]H. Li, T. Jiang and K. Zhang, Efficient and Robust Feature Extraction by Maximum Margin Criterion, *Advances in Neural Information Processing Systems (NIPS 2004),* pp.157-165, 2004.

[39]W. Yang, C. Sun, H. S. Du, J. Yang: Feature Extraction Using Laplacian Maximum Margin Criterion. *Neural Processing Letters* Vol. 33, No. 1, pp. 99-110, 2011

[40] F. Wang, C. Zhang, Feature Extraction by Maximizing the Average Neighborhood Margin, *In Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR'07)*, pp.1-8, 2007.

[41] J. Chen, S. Shan, G. Zhao, X. Chen, W. Gao, Matti Pietikainen, WLD: A Robust Descriptor based on Weber's Law, *IEEE Trans. Pattern Anal. Mach. Intell*., Vol. 32, No.9, pp. 1705–1720, Sep. 2010.

Lining Zhang (S'11) received the B.Eng. degree and the M. Eng. degree in electronic engineering from Xidian University, Xi'an, China, in 2006 and 2009, respectively.

He is currently working towards the Ph.D. degree at the Nanyang Technological University, Singapore. His research interests include computer vision, machine learning, multimedia information retrieval, data mining and computational intelligence. He is a student member of the IEEE.

Lipo Wang (M'97-SM'98) received the B.S. degree from the National University of Defense Technology, Changsha, China, in 1983, and the Ph.D. degree from Louisiana State University, Baton Rouge, in 1988.

He is currently with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interest is computational intelligence with applications to bioinformatics, data mining, optimization, and image processing. He is (co-)author of over 200 papers (of which 80+ are in journals). He holds a U.S. patent in neural networks. He has co-authored 2 monographs and (co-)edited 15 books. He was/will be keynote/panel speaker for several international conferences.

He is/was Associate Editor/Editorial Board Member of 20 international journals, including *IEEE Transactions on Neural Networks*, *IEEE Transactions on Knowledge and Data Engineering*, and *IEEE Transactions on Evolutionary Computation*. He is an elected member of the AdCom (Board of Governors, 2010-2012) of the IEEE Computational Intelligence Society (CIS) and served as IEEE CIS Vice President for Technical Activities (2006-2007) and Chair of Emergent Technologies Technical Committee (2004-2005). He is an elected member of the Board of Governors of the International Neural Network Society (2011-2013) and a CIS Representative to the AdCom of the IEEE Biometrics Council. He was President of the Asia-Pacific Neural Network Assembly (APNNA) in 2002/2003 and received the 2007 APNNA Excellent Service Award. He was Founding Chair of both the IEEE Engineering in Medicine and Biology Singapore Chapter and IEEE Computational Intelligence Singapore Chapter. He serves/served as IEEE EMBC 2011 & 2010 Theme Co-Chair, IJCNN 2010 Technical Co-Chair, CEC 2007 Program Co-Chair, IJCNN 2006 Program Chair, as well as on the steering/advisory/organizing/program committees of over 180 international conferences.

Weisi Lin (M'92–SM'98) received the B.Sc. degree in electronics and the M.Sc. degree in digital signal processing from Zhongshan University, Guangzhou, China, in 1982 and 1985, respectively, and the Ph.D. degree in computer vision from King's College, London University, London, U.K., in 1992.

He taught and conducted research at Zhongshan University, Shantou University (China), Bath University (U.K.), the National University of Singapore, the Institute of Microelectronics (Singapore), and the Institute for Infocomm Research (Singapore). He has been the Project Leader of 13 major successfully-delivered projects in digital multimedia technology development. He also served as the Lab Head, Visual Processing, and the Acting Department Manager, Media Processing, for the Institute for Infocomm Research. Currently, he is an Associate Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. His areas of expertise include image processing, perceptual modeling, video compression, multimedia communication and computer vision. He holds ten patents, edited one book, authored one book and five book chapters, and has published over 180 refereed papers in international journals and conferences.

Dr. Lin is a fellow of Institution of Engineering Technology. He is also a Chartered Engineer (U.K.). He organized special sessions in IEEE International Conference on Multimedia and Expo (ICME 2006), IEEE International Workshop on Multimedia Analysis and Processing (2007), IEEE International Symposium on Circuits and Systems (ISCAS 2010), Pacific-Rim Conference on Multimedia (PCM 2009), SPIE Visual Communications and Image Processing (VCIP 2010), Asia Pacific Signal and Information Processing Association (APSIPA 2011), MobiMedia 2011 and ICME 2012. He gave invited/keynote/panelist talks in International Workshop on Video Processing and Quality Metrics (2006), IEEE International Conference on Computer Communications and Networks (2007), SPIE VCIP 2010, and IEEE Multimedia Communication Technical Committee (MMTC) Interest Group of Quality of Experience for Multimedia Communications (2011), and tutorials in PCM 2007, PCM 2009, IEEE ISCAS 2008, IEEE ICME 2009, APSIPA 2010, and IEEE International Conference on Image Processing (2010). He is currently on the editorial boards of IEEE Trans. on Multimedia, IEEE SIGNAL PROCESSING LETTERS and *Journal of Visual Communication and Image Representation*, and four IEEE Technical Committees. He co-chairs the IEEE MMTC Special Interest Group on Quality of Experience. He has been on Technical Program Committees and/or Organizing Committees of a number of international conferences.