Conjunctive Patches Subspace Learning with Side Information for Collaborative Image Retrieval

Lining Zhang, *Student Member, IEEE,* Lipo Wang, *Senior Member, IEEE,* and Weisi Lin*, *Senior Member, IEEE*

***Abstract*—Content-Based Image Retrieval (CBIR) has attracted substantial attention during the past few years for its potential practical applications to image management. A variety of Relevance Feedback (RF) schemes have been designed to bridge the semantic gap between the low-level visual features and the high-level semantic concepts for an image retrieval task. Various Collaborative Image Retrieval (CIR) schemes aim to utilize the user historical feedback log data with similar and dissimilar pairwise constraints to improve the performance of a CBIR system. However, existing subspace learning approaches with explicit label information cannot be applied for a CIR task, although the subspace learning techniques play a key role in various computer vision tasks, e.g., face recognition and image classification. In this paper, we propose a novel subspace learning framework, i.e., Conjunctive Patches Subspace Learning (CPSL) with side information, for learning an effective semantic subspace by exploiting the user historical feedback log data for a CIR task. The CPSL can effectively integrate the discriminative information of labeled log images, the geometrical information of labeled log images and the weakly similar information of unlabeled images together to learn a reliable subspace. We formally formulate this problem into a constrained optimization problem and then present a new subspace learning technique to exploit the user historical feedback log data. Extensive experiments on both synthetic data sets and a real-world image database demonstrate the effectiveness of the proposed scheme in improving the performance of a CBIR system by exploiting the user historical feedback log data.**

***Index Terms*—collaborative image retrieval, log data, side information, subspace learning.**

I. Introduction

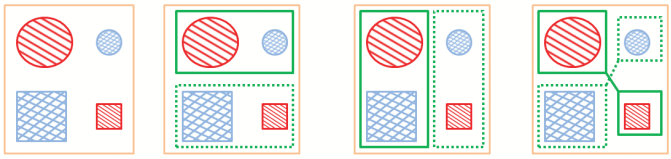
CONTENT-Based Image Retrieval (CBIR) has attracted much attention during the past decades [1], [2], [3]. However, the gap between the low-level visual features and the high-level semantic concepts usually leads to poor performance for CBIR. Although substantial research has been conducted, CBIR is still an open research topic mainly due to difficulties in bridging the semantic gap [1], [2], [3].

Relevance Feedback (RF) is one of the most powerful tools to narrow down this semantic gap and thus to improve the performance of a CBIR system [4], [5]. In general, RF focuses on the interactions between a user and a search engine by requiring the user to label semantically similar or dissimilar images with the query image [4], which are positive and negative feedbacks, respectively. During the last decade, various RF techniques have been proposed to involve the user in the loop to enhance the performance of CBIR [5]. Feature selection based methods adjust weights associated with various dimensions of the feature space to adapt to the user preferences [4], [6]. Support Vector Machine (SVM) based methods either estimate the density of positive feedbacks or regard the RF as a strict two-class on-line classification problem [7], [8]. Traditional discriminant analysis based methods aim to find a low dimensional subspace of the feature space, so that positive feedbacks and negative feedbacks are well separated after projecting onto this subspace. Moreover, Biased Discriminant Analysis (BDA) techniques define a (1+x) class problem and find a subspace within which to separate the one positive class from the unknown number of negative classes [9], [10], [11], [12].

Despite the broad interest in constructing RF approaches, an on-line learning task can be tedious and boring for a user. Given the difficulty in capturing the user preferences, multiple rounds of RF are actually required to achieve satisfactory results for an image retrieval task, which can significantly limit the capability of RF for real-world applications. Recently, a number of studies have attempted to address the challenges encountered by traditional RF approaches by resorting to the user historical feedback log data [13], [14], [15], [16], [17], [18], [19]. In these studies, the system can accumulate RF information provided by a number of users, which can be regarded as the user historical feedback log data. Therefore, besides the low-level visual features, each pair of images can also be associated with a set of similar or dissimilar pairwise constraints judged by users. This new paradigm of utilizing user feedback log data for image retrieval can be referred to as "Collaborative Image Retrieval (CIR)". During the past several years, a lot of research work has been done regarding this new paradigm for image retrieval. In [13], [14], manifold learning algorithms were applied to learn an exquisite manifold structure from the log data, which can better reflect the semantic relation among different images. In [15], Muller et al suggested a weighting scheme by exploiting the user historical feedback log data for CBIR. In [17], Hoi et al proposed a log-based RF technique with the SVM by engaging the user feedback log data in a regular on-line RF task. In [19], the authors proposed a distance metric learning technique by exploiting the user historical feedback log data with pairwise constraints and showed the effectiveness of the proposed scheme comparing with some representative distance metric learning techniques for image retrieval. To sum up, we can notice that the key issue for CIR is to design an effective scheme to fully exploit the user historical feedback log data and to utilize the acquired information to enhance the

L. Zhang and L. Wang are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Email: zhan0327@ntu.edu.sg, elpwang@ntu.edu.sg.

* Corresponding author: W. Lin is with the School of Computer Engineering, Nanyang Technological University, Singapore. Email: wslin@ntu.edu.sg.



(a) (b) (c) (d)

Fig. 1. Different similar relation between pairs of images based on different concept subspaces in a multi-dimensional low level visual feature space. (a) four images with low level visual features (b) similar in the shape subspace, (c) similar in the size subspace, (d) similar in the texture subspace

performance of a CBIR system.

Various methods and schemes have been investigated for CIR; however, there is still little work on explicitly evaluating the subspace learning approaches in exploiting the user historical feedback log data, although the subspace learning techniques play a vital role in many multimedia retrieval tasks. Let us first use a toy example to show the importance of subspace learning approaches in defining the similar relation between a pair of images, which is usually the key issue in exploiting the user historical feedback log data. For an image retrieval task, the images are usually represented by a set of low level visual features with various semantic concepts (e.g., color, shape, texture, etc) in a high dimensional space. With an assumption that different semantic concepts live in different subspaces and each image can live in many subspaces, Fig.1 (a) shows four images, each of which is associated with a number of semantic concepts (i.e., color, shape, texture and size). However, for CIR, it is problematic for a user to determine the similar relation between a pair of images in the original multi-dimensional space (i.e., color, shape, texture and size) due to the semantic gap. By selecting one-dimensional semantic subspace, defining the similar relation between a pair of images will be easy and obvious. Fig.1 (b), (c) and (d) show three different kinds of similar relation in three different semantic subspaces, respectively (i.e., Fig.1 (b) in the shape subspace, Fig.1 (c) in the size subspace and Fig.1 (d) in the texture subspace).

Subspace learning approaches [20] are powerful tools for various tasks in computer vision [21], [10], [22], e.g., face recognition [23], image retrieval [9] and gait recognition [24]. However, most of these traditional subspace learning techniques (e.g., Linear Discriminant Analysis (LDA)) normally need to acquire explicit class labels [20]. For CIR, explicit class labels for each image might be too expensive to be obtained. Compared with explicit class labels of each image, the similar or dissimilar pairwise constraints between a pair of images can be acquired more easier when the user historical feedback log data is available [15]. Therefore, it is more attractive to learn a semantic concept subspace directly from the similar or dissimilar pairwise constraints without using explicit class labels. Recently, learning distance metrics with similar and dissimilar pairwise constraints (or side information [25]) has been actively studied [25], [26], [19], [27] in the machine learning community. Despite the active research efforts during the past few years, most of these approaches in this group have involved a high computational burden when dealing with high dimensional images, which significantly limits their potential applications to CIR.

In this paper, we propose a novel framework of subspace learning when the training images are associated with only similar and dissimilar pairwise constraints, i.e., Conjunctive Patches Subspace Learning (CPSL) with side information, to explicitly exploit the user historical feedback log data for CIR. The proposed CPSL method can effectively learn a reliable subspace both from labeled and unlabeled images through a regularized learning framework in exploiting the user historical feedback log data. Specially, we formally formulate this method into a constrained optimization problem and then present an efficient algorithm to solve this task with closed-form solutions. Compared with the previous metric learning techniques with side information [25], [26], [19], which usually involve a convex optimization procedure or a semidefinite programming procedure, our method can also learn a distance metric but perform more effectively and efficiently when dealing with high dimensional images.

The rest of this paper is organized as follows: Section II reviews the related work; the CPSL with side information framework is detailed in Section III; a CIR system is introduced in Section IV; in Section V, we first give the experimental results on both of synthetic datasets and a real-world image database, and then show some analysis to the important parameters in CPSL; Section VI concludes this paper.

II. Related Work

To describe our method clearly, let us first review two areas of research that are closely related to our work in this paper, i.e., (1) CIR and (2) subspace learning and distance metric learning.

A. Review on CIR

During the past years, various advanced on-line RF schemes have been constructed. However, it is still a big problem to effectively bridge the semantic gap between the low-level visual features and the high-level semantic concepts.

Besides on-line RF paradigms, there are some emerging research interests in exploiting the user historical feedback log data [16], [17] for image retrieval. In [17], Hoi et al proposed a log based RF scheme with the SVM by engaging the user feedback log data in a traditional on-line RF task. In this scheme, the user first labels some similar and dissimilar images in a few rounds of RF iterations, and then the images in the database that are similar to the current labeled images are included in the pool of labeled data for training some regular RF models, e.g., SVM RF. Besides the SVM approaches with log data, some other efforts are also investigated in exploiting the user historical feedback log data. For instance, manifold learning techniques expect to learn an exquisite manifold structure based on the user historical feedback log data [13], [14]. In [15], Muller et al proposed a feature weighting scheme by exploiting the user historical relevance judgements for a CBIR task. Moreover, some distance metric learning techniques have also been widely investigated to learn a good Mahalaninos distance metric by exploiting the user historical

similar and dissimilar judgements on the feedback images for image retrieval [18], [19].

B. Review on subspace learning and distance metric learning

In view of the close relation between subspace learning techniques and distance metric learning techniques, we briefly classify the two groups of studies into three categories within a unified framework, i.e., unsupervised learning, supervised learning with explicit class labels and weakly supervised learning with pairwise constraints (or side information [25]).

Unsupervised learning methods do not use any class label information and usually exploit the intrinsic distribution or the manifold structure of the data. Examples in this category include the well-known algorithms, such as Principal Component Analysis (PCA) [20] and Multi-Dimensional Scaling (MDS) [28]. Moreover, there are also some recent manifold learning based techniques, which are Locally Linear Embedding (LLE) [29], ISOMAP [30], Laplacian Eigenmaps (LE) [31], Locality Preserving Projections (LPP) [32], etc.

Supervised learning approaches can effectively explore some collections of training data with explicit class labels. Well-known techniques in this category include fisher's LDA [20], Marginal Fisher Analysis (MFA) [33], and some recently proposed methods, such as Neighborhood Component Analysis (NCA) [34], Large Margin Nearest Neighbor classification (LMNN) algorithm [35] and Maximally Collapsing Metric Learning (MCML) [36].

Our work is closely related to the third category of research. Let us briefly introduce several representative algorithms below.

Most of the weakly supervised learning approaches can only learn a Mahalanobis distance metric from the training data that are presented in the forms of pairwise constraints (or side information [25]), in which each pairwise constraint indicates whether the corresponding two samples are similar or dissimilar for a particular task. In [25], Xing et al proposed a distance metric learning approach (called Xing hereafter) and formulated the task into a convex optimization problem, which can be solved by an iterative projection algorithm. And then, a series of research work has been done with regard to this category of studies. In [26], a Relevant Component Analysis (RCA) technique was proposed to exploit only similar pairwise constraints for distance metric learning. In details, given pairwise constraints, RCA first forms a set of "chunklets", each of which is defined as a group of samples linked together by similar pairwise constraints. The optimal distance metric learned by RCA can be computed as the inverse of the average covariance matrix of the chunklets. RCA is simple to calculate, but ignores the dissimilar pairwise constraints. Discriminative Component Analysis (DCA) was proposed to incorporate the dissimilar pairwise constraints [27], which can show slightly better discriminative performance compared to RCA for some datasets. Lately, an Information-Theoretic Metric Learning (ITML) approach was proposed to express the weakly supervised learning problem as a Bregman optimization problem [37]. To effectively exploit the unlabeled samples, Hoi et al proposed a Laplacian Regularized Metric Learning (LRML) approach and then applied the generated solution to image retrieval and clustering [19]. In [38], Wu et al proposed to learn a Bergman distance function with side information and showed the approach can learn nonlinear distance functions for a semi-supervised clustering task.

III. Conjunctive Patches Subspace Learning with Side Information for CIR

In this section, we propose a novel framework of weakly supervised subspace learning, i.e., Conjunctive Patches Subspace Learning (CPSL) with side information, to explicitly exploit the user historical feedback log data for CIR. The proposed CPSL can learn a semantic subspace directly from the similar and dissimilar pairwise constraints without using any class labels, which is more practical for CIR, since explicit class labels for each image might be too expensive to obtain for a real-world image retrieval task.

A. Problem Definition

To facilitate the discussion, let us first introduce some necessary notations. Assume that we are given a set of N images in a H dimensional visual feature space $X = \{x_i\}_{i=1}^{N} \in R^H$, and two sets of similar and dissimilar pairwise constraints among these images:

$$S = \{(i,j) |\ x_i \ and \ x_j \ are \ judged \ to \ be \ similar\}$$
$$D = \{(i,j) |\ x_i \ and \ x_j \ are \ judged \ to \ be \ dissimilar\}$$

where S is the set of similar pairwise constraints and D is the set of dissimilar pairwise constraints. Each pairwise constraint (i, j) indicates if two images x_i and x_j are similar or dissimilar judged by users in RF iterations. It should be noted that it is not necessary for all the images in X to be involved in S or D.

In this paper, we use the low-level visual features in a high dimensional space to represent images. Although the low-level visual features of images are embedded in a high dimensional space, the semantic concepts of images actually live in a low dimensional subspace. Here, in this paper, the high dimensional space R^H is the low-level visual feature space and the low dimensional subspace R^L is the high-level semantic concept space. Therefore, our objective is to find a mapping function F to select an effective semantic concept subspace R^L from R^H for bridging the semantic gap. To learn such a semantic concept subspace, one can assume there is some corresponding linear mapping $W \in R^{H*L}$ for a possible subspace, and then we can obtain the low-dimensional semantic representations as $Y = W^T X \in R^{L*N}$, where each column of Y is $y_i = W^T x_i \in R^L$.

To measure the similarity between two images y_i and y_j in the low dimensional semantic concept subspace, we adopt the Euclidean distance metric because of its simplicity and robustness. The Euclidean distance between two images in the low dimensional semantic subspace can be calculated as follows:

$$\begin{aligned} d(y_i, y_j) &= \sqrt{(W^T x_i - W^T x_j)^T (W^T x_i - W^T x_j)} \\ &= \sqrt{(x_i - x_j)^T W W^T (x_i - x_j)} \end{aligned} \quad (1)$$

Let $M = WW^T$, then,

$$d(y_i, y_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \quad (2)$$

Therefore, learning a mapping matrix W is actually equivalent to learning an efficient Mahalanobis distance metric M in the original high dimensional space, or more concretely, learning a proper Mahalanobis distance metric M in R^H.

During recent years, a variety of techniques have been proposed to learn such an optimal Mahalanobis distance metric M from training data that are given in forms of side information [19], [25], [26], [27], [38], [39]. However, most of these methods are imperfect for a CIR task, since they either require solving a convex optimization problem with gradient decent and iterative projections [25], [26], [38] or involve to solve a semi-definite programming problem [19], [39], which often suffers from large computational cost and limits its potential applications for high dimensional data. Moreover, most of these methods, which can learn Mahalanobis distance metrics from the training data, are unable to explicitly give the new representations of data in the new metric space. Considering this, in this paper we expect to learn a mapping matrix W instead of a Mahalanobis distance metric M. From another point of view, we can also learn a Mahalanobis distance metric M by resorting to the mapping matrix, i.e., $M = WW^T$.

In this paper, we present a novel regularized weakly supervised subspace learning framework to explicitly exploit the user historical feedback log data for a CIR task, i.e.,

$$\begin{aligned} W^* &= \arg\min_{W \in R^{H*L}} f(W, X_l, S, D) \\ &+ \beta_1 g(W, X_l, S) + \beta_2 r(W, X_l, X_u) \end{aligned} \quad (3)$$

where W is the mapping matrix, and $f(\cdot)$ is a loss function defined over the labeled images X_l with the associated constraints S and D to reflect the discriminative information; $g(\cdot)$ is a regularizer defined over the labeled images X_l with the associated similar constraints S, which models the geometry information of labeled image pairs; and $r(\cdot)$ is also a regularizer, which is defined over the labeled images X_l and unlabeled images X_u on the target mapping matrix W; β_1 and β_2 are two trade off parameters, which are used to balance the three terms. The above regularized subspace learning framework is largely inspired by the recent regularization principle in the machine learning community, which is usually the key to enhance the generalization and robustness performance of machine learning techniques. The regularization principle has played a vital role in alleviating the over fitting problem encountered by many machine learning techniques [40]. For instance, the regularization principle is the most critical aspect in ensuring the good generalization performance in SVMs [41], [42], [43]. Similarly, the regularization method is also an effective technique to enhance the stable performance of the fisher's LDA when dealing with small number of samples in a high dimensional space [41].

Given the above weakly supervised subspace learning framework, the key issue to attack this problem is to design one appropriate loss function $f(\cdot)$, two regularizer terms $g(\cdot)$ and $r(\cdot)$, and afterward find an efficient algorithm to solve this problem. In the following subsections, we will study some principles for formulating the reasonable loss function, the regularizer terms and also discuss the solutions to this problem.

B. Conjunctive Patches Subspace Learning with Side Information for CIR

In this paper, the CIR system reduces the semantic gap by exploiting the historical feedback log data judged by users in RF iterations and finding a semantic concept subspace to reflect the similar relation between image pairs, thereby further enhancing the performance of an image retrieval system. We use a linear mapping matrix W to approximate this semantic concept subspace and then the images in this subspace can be represented as $Y = W^T X = [y_1, y_2, \ldots, y_N] \in R^{L \times N} (L < H)$ with $y_i \in R^L$ for image $x_i \in R^H$. Therefore, in this reduced semantic concept subspace, an improved retrieval performance is expected.

In this subsection, we present a Conjunctive Patches Subspace Learning method (CPSL) with side information to learn such a mapping matrix W. Specially, the CPSL can effectively integrate the discriminative information of labeled log images, the geometry information of labeled log images, and the weakly similar information of unlabeled images. This process is conducted by building different kinds of local patches for each image, and then aligning those different kinds of patches together to learn a consistent coordinate through the above regularized learning framework. One patch is a local area, which is formed by one image and its associated neighboring images. Particularly, in CPSL, we build three different kinds of patches, which are: 1) local discriminative patches for labeled log images to represent the discriminative information; 2) local geometry patches for labeled log images to represent the geometry information and the 3) local weakly similar patches for labeled and unlabeled images to represent the weak similarity of unlabeled images.

1) Local Discriminative Patches for Labeled Images: Given images with side information, a popular principle for learning a distance metric M is to minimize the distances between samples with similar pairwise constraints and to maximize the distances between samples with dissimilar pairwise constraints simultaneously, which can be referred to as a min-max principle. In [25], Xing et al formulated the weakly supervised distance metric learning problem as a constrained convex optimization problem, i.e.,

$$\min_{M \succeq 0} \sum_{(x_i, x_j) \in S} ||x_i - x_j||_M^2 s.t. \sum_{(x_i, x_j) \in D} ||x_i - x_j||_M \geq 1 \quad (4)$$

Eq.(4) attempts to find the optimal metric M by minimizing the sum of squared distances between the samples with similar pairwise constraints, and meanwhile enforcing the sum of distances between the samples with dissimilar pairwise constraints larger than or equal to 1. Following this principle, [19] defined two loss functions by minimizing the sum of squared distances between all the samples with similar pairwise constraints and maximizing the sum of squared distances

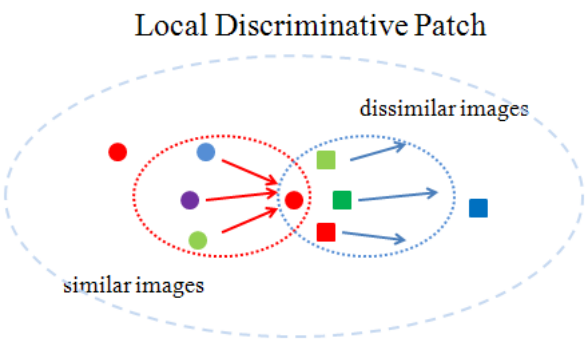


Fig. 2. The illustration of the Local Discriminative Patch for an image and its associated nearby similar and dissimilar images. Circular solid dots and square solid dots denote labeled similar and dissimilar images, respectively. For each given image, minimizing the objective function of the Local Discriminative Patch will pull the nearby similar images towards this image while pushing the nearby dissimilar images away from this image in the reduced subspace.

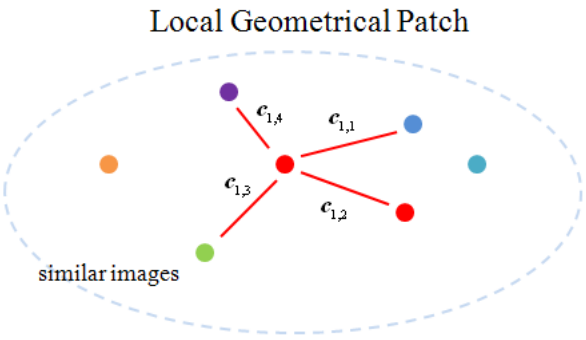


Fig. 3. The illustration of the Local Geometrical Patch for an image and its associated nearby similar images. Circular solid dots denote labeled similar images. For each given image, the Local Geometrical Patch aims to preserve the local geometry of labeled similar images before and after projection. Minimizing the objective function of the Local Geometrical Patch will reconstruct the given image from its associated nearby similar images with a minimal error in the reduced subspace.

between all the samples with dissimilar pairwise constraints. Although the above distance metric learning approaches have been demonstrated to be effective for some test data sets, they are essentially linear global approaches and therefore might fail to find the nonlinear structure hidden in high dimensional visual feature space.

Following this min-max principle, to further exploit the discriminative power, we define a new loss function for discriminative information preservation. Particularly, for each image x_i associated with a discriminative patch $X_{d(i)} = [x_i, x_{i_1}, x_{i_2}, \ldots, x_{i_{k_1}}, x_{i_{k_1+1}}, x_{i_{k_1+2}}, \ldots, x_{i_{k_1+k_2}}]$, wherein $x_{i_1}, x_{i_2}, \ldots, x_{i_{k_1}}$, i.e., the k_1 nearest images of x_i with similar pairwise constraints, and $x_{i_{k_1+1}}, \ldots, x_{i_{k_1+k_2}}$, i.e., the other k_2 nearest images with dissimilar pairwise constraints. We define the discriminative loss function as the average difference between two kinds of squared distances over this patch. That is, the discriminative loss function attempts to minimize the average squared distances between each image x_i and its associated k_1 nearest images with similar constraints; meanwhile, it tries to maximize the average squared distance between each image x_i and its associated k_2 nearest images with dissimilar constraints. A illustration of the local discriminative patch for one image is given in Fig. 2. Specially, for the new representations of each patch, i.e., $y_i, y_{i_1}, y_{i_2}, \ldots, y_{i_{k_1}}, y_{i_{k_1+1}}, y_{i_{k_1+2}}, \ldots, y_{i_{k_1+k_2}}$, we expect that the loss function between k_1 nearest images with similar constraints and k_2 nearest images with dissimilar constraints will be minimized as much as possible, i.e.,

$$f(y_i) = \min \sum_{j=1}^{k_1} ||y_i - y_{i_j}||^2 \frac{1}{k_1} - \gamma \sum_{j=k_1+1}^{k_1+k_2} ||y_i - y_{i_j}||^2 \frac{1}{k_2} \quad (5)$$

To rewrite Eq.(5) in a more compact form,

$$\begin{aligned} f(y_i) &= \min \sum_{j=1}^{k_1} ||y_i - y_{i_j}||^2 \tfrac{1}{k_1} - \gamma \sum_{j=k_1+1}^{k_1+k_2} ||y_i - y_{i_j}||^2 \tfrac{1}{k_2} \\ &= \min tr(Y_{d(i)} \begin{bmatrix} -e_{k_1+k_2}^T \\ I_{k_1+k_2} \end{bmatrix} diag(w_i) \left[-e_{k_1+k_2}, I_{k_1+k_2}\right] Y_{d(i)}^T) \\ &= \min tr(Y_{d(i)} L_{d(i)} Y_{d(i)}^T) \end{aligned} \quad (6)$$

where $w_i = [\overbrace{1/k_1, \ldots, 1/k_1}^{k_1}, \overbrace{-\gamma/k_2, \ldots, -\gamma/k_2}^{k_2}]^T$; the parameter γ is used to balance the two squared distances; $I_{k_1+k_2}$ is a $(k_1+k_2) \times (k_1+k_2)$ identity matrix; $L_{d(i)} = \begin{bmatrix} \sum_{j=1}^{k_1+k_2} (w_i)_j & -w_i^T \\ -w_i & diag(w_i) \end{bmatrix}$; the vector $e_{k_1+k_2} = [1, \ldots, 1]^T \in R^{k_1+k_2}$; $d(i)$ encodes the discriminative information over this local discriminative patch.

2) Local Geometrical Patches for Labeled Images: Although the discriminative loss function for each labeled image can capture the discriminative information well, it is empirically known that the geometrical information of images can help to find the intrinsic semantic concept subspace. In the past few years, various geometry based subspace learning algorithms were proposed to recover the mainfold structure of samples in a high dimensional space. LE [31] minimizes the average of the Laplacian operator over the manifold of samples and LPP [32] is a linearization version of LE. ISOMAP [30] tries to preserve the pairwise geodesic distance, which can also be used to effectively recover the intrinsic structure of samples in a high dimensional space . LLE [29] uses the reconstruction coefficients in a high dimensional space to reconstruct the sample from its neighboring samples in a low dimensional space with a minimal error. In this work, we utilize the LLE technique to preserve the local geometry information for semantic concept subspace learning.

In particular, for each image x_i associated with a geometrical patch $X_{g(i)} = [x_i, x_{i_1}, x_{i_2}, \ldots, x_{i_{k_1}}]$, wherein $x_{i_1}, x_{i_2}, \ldots, x_{i_{k_1}}$, i.e., the k_1 nearest samples of x_i with similar pairwise constraints. As we can see in Fig.3, this work assumes that the new representation y_i of one image x_i can be linearly reconstructed by its k_1 nearest images with similar constraints with a minimal error, i.e.,

$$g(y_i) = \min ||y_i - \sum_{j=1}^{k_1} c_{ij} y_{i_j}||^2 \quad (7)$$

Eq.(7) is used to preserve the local geometry of labeled images with similar constraints before and after projection, and the linear combination coefficient vector c_i is required

to reconstruct x_i from its k_1 nearest similar images with a minimal error, i.e.,

$$\begin{aligned} &\min_{c_i} ||x_i - \sum_{j=1}^{k_1} c_{ij} x_{i_j}||^2 \\ &s.t. \sum_{j=1}^{k_1} c_{i_j} = 1 \end{aligned} \tag{8}$$

To solve this problem, we have $c_{i_j} = \sum_{p=1}^{k_1} G_{jp}^{-1} / (\sum_{s=1}^{k_1} \sum_{t=1}^{k_1} G_{st}^{-1})$ with a local gram matrix $G_{jp} = (x_i - x_{i_j})^T (x_i - x_{i_j})$ as described in [29].

For simplicity, we rewrite Eq.(7) in a more compact form. By attaching k_2 nearest dissimilar images of x_i with the geometrical patch $X_g(i)$, we have

$$\begin{aligned} g(y_i) &= \min ||y_i - \sum_{j=1}^{k_1} c_{ij} y_{ij}||^2 \\ &= \min ||y_i - \sum_{j=1}^{k_1} c_{ij} y_{i_j} - \sum_{j=k_1+1}^{k_1+k_2} 0 \cdot y_{ij}||^2 \\ &= \min ||y_i - \sum_{j=1}^{k_1+k_2} \bar{c}_{i_j} y_{ij}||^2 \\ &= \min tr(Y_{d(i)} \begin{bmatrix} 1 & -\bar{c}_i^T \\ -\bar{c}_i & \bar{c}_i \bar{c}_i^T \end{bmatrix} Y_{d(i)}^T) \\ &= \min tr(Y_{d(i)} L_{g(i)} (Y_{d(i)})^T) \end{aligned} \tag{9}$$

where $Y_{d(i)} = [y_i, y_{i_1}, y_{i_2}, \ldots, y_{i_{k_1}}, y_{i_{k_1+1}}, \ldots, y_{i_{k_1+k_2}}]$; $L_{g(i)} = \begin{bmatrix} 1 & -\bar{c}_i^T \\ -\bar{c}_i & \bar{c}_i \bar{c}_i^T \end{bmatrix}$ with $\bar{c}_i = [c_i^T, \underbrace{0, \ldots, 0}_{k_2}]^T$; $g(i)$ is used to encode the geometrical information over this local geometrical patch.

3) Local Weakly Similar Patches for Labeled and Unlabeled Images: Recent research has shown that unlabeled samples may be helpful to improve the classification performance. During the last decade, various semi-supervised techniques have attracted an increasing amount of attention. In [44], Semi-supervised Discriminant Analysis (SDA) was proposed to find a projection which respects the discriminant structure inferred from the labeled samples, as well as the intrinsic geometrical structure inferred from both labeled and unlabeled samples. In [19], Hoi et al introduced a Laplician regularizer to a supervised metric learning approach and showed that the semi-supervised metric learning method can learn effective distance metrics by exploiting unlabeled samples when labeled samples are limited and noisy. Inspired by the recent advance in the semi-supervised research, in this part, we design a new regularizer term based on labeled and unlabeled images, and then introduce this term to our regularized subspace learning framework to find an effective semantic concept subspace.

Unlabeled images are attached to the labeled log images: $X = [x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+n_u}]$, where the first n images are judged by user in RF iterations, and the remaining n_u images have no label information. For each image $x_i \in X, i = 1, \ldots, n + n_u$, we first find its k_3 nearest neighborhood samples $x_{i_1}, \ldots, x_{i_{k_3}}$ in all images including both labeled and unlabeled images. And then the image x_i and its associated k_3 nearest images $X_{u(i)} = [x_i, x_{i_1}, \ldots, x_{i_{j_3}}]$ form a local weakly similar patch. The key to semi-supervised learning algorithm is the prior assumption of consistency. For subspace learning techniques, it can be interpreted as nearby data will have similar low-dimensional representations. The local weakly similar patch for one image is illustrated in Fig.4. Particularly, for the new representations of each patch, i.e., $Y_{u(i)} = [y_i, y_{i_1}, \ldots, y_{i_{k_3}}]$, we minimize the sum of the weighted squared distances between y_i and $y_{i_1}, \ldots, y_{i_{k_3}}$, and we have

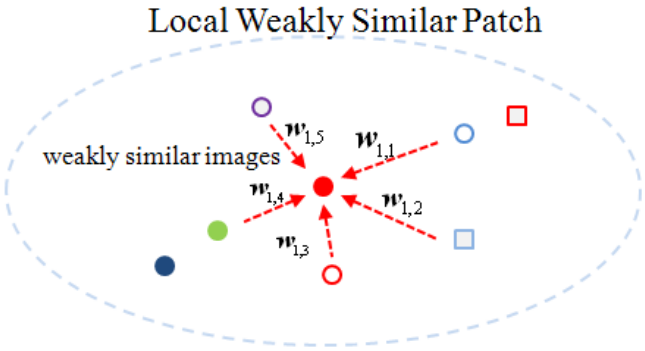


Fig. 4. The illustration of the Local Weakly Similar Patch for an image and its associated nearby labeled and unlabeled images. Solid dots and hollow dots denote labeled and unlabeled images, respectively. For each given image, the Local Weakly Similar Patch attempts to impose local consistency constraints on this image and its associated nearby images. Minimizing the objective function of the Local Weakly Similar Patch will incorporate the local weakly similarity information of unlabeled images in the reduced subspace.

$$r(y_i) = \min \sum_{j=1}^{k_3} ||y_i - y_{i_j}||^2 \frac{w_{i,j}}{k_3} \tag{10}$$

Similarly, to rewrite the local weakly similar patch into a compact form, we can rephrase Eq.(10) of the patch of y_i as follows,

$$\begin{aligned} r(y_i) &= \min \sum_{j=1}^{k_3} ||y_i - y_{i_j}||^2 \frac{\omega_{i,j}}{k_3} \\ &= \min tr \left(Y_{u(i)} \begin{bmatrix} \sum_{j=1}^{k_3} \bar{\omega}_{i,j} & -\bar{\omega}_i^T \\ -\bar{\omega}_i & diag(\bar{\omega}_i) \end{bmatrix} Y_{u(i)}^T \right) \\ &= \min tr(Y_{u(i)} L_{u(i)} Y_{u(i)}^T) \end{aligned} \tag{11}$$

where the weight $\omega_{i,j} = \exp(-||x_i - x_j||^2 / \delta^2)$ is the Laplacian heat kernel according to LE [31]; the patch $Y_{u(i)} = [y_i, y_{i_1}, \ldots, y_{i_{k_3}}]$; $L_{u(i)} = \begin{bmatrix} \sum_{j=1}^{k_3} \bar{w}_{i,j} & -\bar{w}_i^T \\ -\bar{w}_i & diag(\bar{w}_i) \end{bmatrix}$; the vector $\bar{w}_i = [\underbrace{\frac{w_{i,1}}{k_3}, \ldots, \frac{w_{i,j}}{k_3}}_{k_3}]$; $u(i)$ encodes the weakly similar information between labeled images and unlabeled images.

4) Conjunctive Patches Subspace Learning with Side Information: Each of the constructed patches has its own coordinate system. To get a consistent coordinate, we can first align each of these three different kinds of patches together to obtain a consistent coordinate according to an alignment trick [45], [46], respectively. For each image x_i, the associated patch $Y_i = [y_i, y_{i_1}, \ldots, y_{i_k}]$ can be rewritten as $Y_i = YS_i$, where $Y = [y_1, \ldots, y_N]$, $N = n + n_u$ is the number of labeled and unlabeled images and $S_i = R^{N \times (k+1)}$ is the selection matrix. And S_i is defined according to [45], [46] as follows,

$$(S_i)_{st} = \begin{cases} 1, \text{if } s = F_i(t) \\ 0, else \end{cases} \tag{12}$$

where $F_i = [i, i_1, \ldots, i_k]$ is the index vector for samples in Y_i .

And then, we can integrate all the three different kinds of patches defined in Eq.(5), Eq.(7) and Eq.(10) together through the regularized subspace learning framework in Eq.(3), i.e.,

$$\begin{aligned}
&\min f(W, X_l, S, D) + \beta_1 g(W, X_l, S) + \beta_2 r(W, X_l, X_u) \\
&= \sum_{i=1}^{n} \min tr(Y_{d(i)} L_{d(i)} Y_{d(i)}^T) + \beta_1 \sum_{i=1}^{n} \min tr(Y_{d(i)} L_{g(i)} Y_{d(i)}^T) \\
&+\beta_2 \sum_{i=1}^{n+n_u} \min tr(Y_{u(i)} L_{u(i)} Y_{u(i)}^T) \\
&= \min tr \left(\sum_{i=1}^{n} Y_{d(i)} L_{d(i)} Y_{d(i)}^T \right) + \beta_1 tr \left(\sum_{i=1}^{n} Y_{d(i)} L_{g(i)} Y_{d(i)}^T \right) \\
&+\beta_2 tr \left(\sum_{i=1}^{n+n_u} Y_{u(i)} L_{u(i)} Y_{u(i)}^T \right) \\
&= \min tr \left(Y \left(\sum_{i=1}^{n} S_{d(i)} L_{d(i)} S_{d(i)}^T \right) Y^T \right) \\
&+\beta_1 tr \left(Y \left(\sum_{i=1}^{n} S_{d(i)} L_{g(i)} S_{d(i)}^T \right) Y^T \right) \\
&+\beta_2 tr \left(Y \left(\sum_{i=1}^{n+n_u} S_{u(i)} L_{u(i)} S_{u(i)}^T \right) Y^T \right) \\
&= \min tr \left(W^T X \left(\sum_{i=1}^{n} S_{d(i)} L_{d(i)} S_{d(i)}^T \right) X^T W \right) \\
&+\beta_1 tr \left(W^T X \left(\sum_{i=1}^{n} S_{d(i)} L_{g(i)} S_{d(i)}^T \right) X^T W \right) \\
&+\beta_2 tr \left(W^T X \left(\sum_{i=1}^{n+n_u} S_{u(i)} L_{u(i)} S_{u(i)}^T \right) X^T W \right) \\
&= \min tr \left(W^T X \left(D + \beta_1 G + \beta_2 U \right) X^T W \right)
\end{aligned} \tag{13}$$

where D encodes the discriminative information and $D = \sum_{i=1}^{n} \left(S_{d(i)} L_{d(i)} S_{d(i)}^T \right)$; G encodes the geometrical information and $G = \sum_{i=1}^{n} \left(S_{d(i)} L_{g(i)} S_{d(i)}^T \right)$; U encodes the weakly similar information of unlabeled images and $U = \sum_{i=1}^{n+n_u} \left(S_{u(i)} L_{u(i)} S_{u(i)}^T \right)$; $\beta_1, \beta_2 > 0$ are tuning parameters, which are used to trade off the contributions of the three different terms.

The above regularized subspace learning framework can be further improved. Because, in the extreme case, when the two trade off parameters $\beta_1 \to 0$ and $\beta_2 \to 0$, the above optimization problem will result in trivial solutions by shrinking the entire space, i.e., obtaining the optimal solution of $W^* = 0$. Therefore, we should impose some constraints on the mapping matrix W on Eq.(13) and then the problem can be converted to a constrained optimization problem of the mapping matrix W.

Remark I: To avoid trivial solutions and find the mapping matrix W, various different constraints may be used to impose on this optimization problem, which will lead to different constrained optimization problem. A simple constraint $tr(W^T W) = 1$ can be imposed on this optimization function. This problem will result in a standard Eigenvalue decomposition problem and the W is the eigenvector corresponding to the smallest non zero eigenvalue. This method always produces rank one solutions. In other words, the original input space will be projected onto a line by this transformation. However, in many cases it is desirable to obtain a compact low dimensional feature representation of the original input space.

Remark II: Various distance metric learning approaches with side information have been designed to learn such a distance metric M. However, some of these methods are actually based on the second-order statistical properties of the training data as the discriminative loss function in CPSL, and thus involve to solve a semidefinite programing problem [39], [19]. For example, in [39], Ghodsi et al defined a loss function, which attempts to minimize the squared induced distance between similar samples while maximizing the squared induced distance between dissimilar samples. Additionally, two constraints are also imposed on this loss function to avoid trivial solutions, i.e.,

$$\begin{aligned}
&\min_M \frac{1}{|S|} \sum_{(x_i, x_j) \in S} ||x_i - x_j||_M^2 - \frac{1}{|D|} \sum_{(x_i, x_j) \in D} ||x_i - x_j||_M^2 \\
&s.t. M \succeq 0, \ tr(M) = 1
\end{aligned} \tag{14}$$

where the first constraint ensures a valid metric, and the second constraint excludes the trivial solutions where all distances are zeros. This loss function is then converted into a linear objective and solved by semidefinite programming for finding a proper distance metric M. However, the computational burden of this method is too high, and this significantly limits its potential applications to high dimensional data.

Although various different constraints can be imposed on Eq.(13) to avoid trivial solutions, they are actually arbitrary. Considering this, we impose $W^T W = I$ on the Eq.(13), to avoid the trivial solutions, which can be solved by conducting the standard Eigenvalue decomposition and the mapping matrix W is formed by the L eigenvectors associated with the first L smallest eigenvalues. This constrained optimization problem can also lead to closed-form solutions as in [39], [19] but without the runtime inefficiency. Additionally, we can easily obtain the distance metric M by resorting to the mapping matrix W.

IV. The Collaborative Image Retrieval System

A. Overview of our CIR Framework

In this subsection, we firstly give an overview of our CIR system, which can systematically integrate the user relevance judgements with a regular RF scheme for image retrieval. The CIR system assumes that the user expects the best possible retrieval results for each query image, i.e., the system is usually required to return the most semantically relevant images based on the previous RF information. Meanwhile, the user will never label a large number of images at each RF iteration and only do a few rounds of RF iterations. To deal with this type of scenario, the following CIR framework is proposed.

As shown in Fig.5, when a query image is provided, the low-level visual features are firstly extracted. Then, all images in the image database are sorted based on a predefined similarity metric. If the user is satisfied with the results, the image retrieval process can end. However, most of the time, the RF

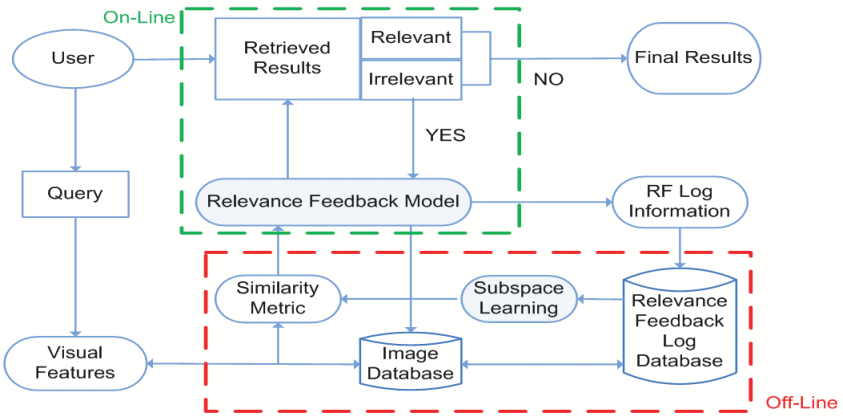


Fig. 5. The framework of our CIR system

is actually needed because of the poor retrieval performance of the system. The CIR system requires the user to label some top similar and dissimilar images as positive and negative feedbacks, respectively. Based on the user on-line feedback information, a RF model can be trained based certain machine learning techniques. The similarity metric can be updated together with the RF model. Then, all the images are resorted based on the recalculated similarity metric. If the user is satisfied with the refined results, the RF is no longer required (i.e., we denote "No" in Fig.5) and the system gives the final retrieved results. On the contrary, the RF will be performed iteratively (i.e., we denote "Yes" in Fig.5).

From Fig.5, it can be noticed that our CIR system is different from regular on-line RF schemes based CBIR systems. The CIR system integrates regular on-line RF schemes with an off-line feedback log data exploiting scheme. In Fig.5, we can see that the CIR system first collects the user on-line RF information, which can be stored in a RF log database. If the user feedback log data is unavailable, the CIR system performs exactly like traditional RF based CBIR systems. When the user RF information is available, the algorithm can effectively exploit the user feedback log data. Thus, the CIR system can accomplish a retrieval task in less iterations than regular RF schemes based system with the help of the user historical feedback log data.

B. Corel Image Database and Image Representation

Fig. 6. Some example images in the log database groups

To perform empirical evaluation of our proposed method, firstly we should provide a reliable image database with semantic groups. Corel Photo Gallery is a professionally catalogued image database and is widely used to evaluate the performance of a CBIR system in the past few years [10], [19], [47], [48]. To validate the effectiveness of the proposed algorithm, we group the images into a number of classes based on the ground truth. 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 while many images with the same concept are in different categories. Therefore, existing categories of the original Corel Photo Gallery are ignored and reorganized into 80 conceptual classes based on the ground truth, such as, lion, castle, bus, aviation, dinosaur, horse, etc. Note that each class of the Corel Photo Gallery has a clearly distinct concept and the quality of the images can be considered very high. Finally, the Corel Image Database comprises totally 10,763 real-world images. This way of using the images with semantic categories is able to help to evaluate the retrieval performance automatically, which significantly reduces subjective errors compared to manual evaluations.

Collecting the user historical feedback log data is an important step for a collaborative image retrieval task. However, to our best knowledge, there is no public data set for the application of exploiting user historical feedback log data for image retrieval. Moreover, for an RF procedure, different users are likely to have different opinions on judging similar and dissimilar images with the query image. In our experiments, to conduct objective evaluation and effectively investigate the performance of weakly supervised learning approaches, we have to provide a reliable log database to run these weakly supervised algorithms. It is not difficult to build a log data database based on an existing real-world database, e.g., Corel Image Database. Here, we first randomly select 10 classes according to the ground truth of the images from the Corel Image Database and form a log data database, which contains 1385 real-world images. And then, to distinguish between the supervised learning task and the weakly supervised learning task, we divide each class of the database into two groups with equal size. Therefore, the log data database comprises 20 groups with 10 different concepts. We randomly select 10 and 30 images uniformly from each group, and therefore we can gather two labeled log data sets. The similar constraints are imposed on the images within the same group, while the dissimilar constraints are imposed on the images with

different concepts. Finally, we can obtain two log databases with different number of log data, i.e., 200 log images, 600 log images. Some example images in the log database are shown in Fig.6.

To represent images, we use three different sets of low-level visual features in our experiments, i.e., color [49], local descriptors [50] and shape [51]. For color, a 9-dimensional color moment feature in Luv color space is first employed. Then, we select three measures (i.e., hue, saturation, and value) and use them to form a histogram. Hue and saturation are both quantized into eight bins and value into four bins. The local dense features, i.e., the Webber Local Descriptors (WLD) [50], are extracted to describe the local visual features of images, which result in 240-dimensional values. Moreover, we employ the edge directional histogram 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 edges. Each of these features has its own capability to characterize the content of images. The system combines the three different kinds of low-level visual features into a vector with 510 values. Then all feature components are normalized to normal distributions with zero mean and one standard deviation to represents the images.

V. Experimental results

In this section, we evaluate the performance of the proposed method in exploiting the user historical feedback log data for CIR. We design the experiments for performance evaluation in four aspects. First of all, we use six synthetic data sets to illustrate the effectiveness of the discriminative loss function in seeking the discriminative directions in RF. Secondly, we investigate the performance of the proposed method by exploiting historical feedback log database for an image retrieval task without a RF scheme. Then, we report the performance of our CIR system by exploiting the user on-line feedback log data and compare it with a regular RF scheme (i.e., SVM RF) based image retrieval system. Finally, we study the sensitivity of important parameters of the proposed CPSL method. In our experiments, all methods are implemented with MATLAB 7.6.0 and all experiments are performed on a desktop computer with 3.0 GHz Intel Duo Core CPU, 3 GB memory and Windows XP system.

A. Experiments with synthetic data sets

In order to visualize the effectiveness of the discriminative loss function (i.e., Eq.(5)) of CPSL in seeking the most discriminative directions in RF, the first experiment is executed on six synthetic datasets. In each round of RF, the user judges a set of similar and dissimilar images with the query image, which are positive and negative feedbacks, respectively. The positive and negative feedbacks are generated with various strikingly different distributions since the distributions of feedback data are usually complicated in real world. Regarding the set of positive feedbacks and the set of negative feedbacks as two different classes, LDA treats the two different sets of feedback samples equally. Based on the assumption that

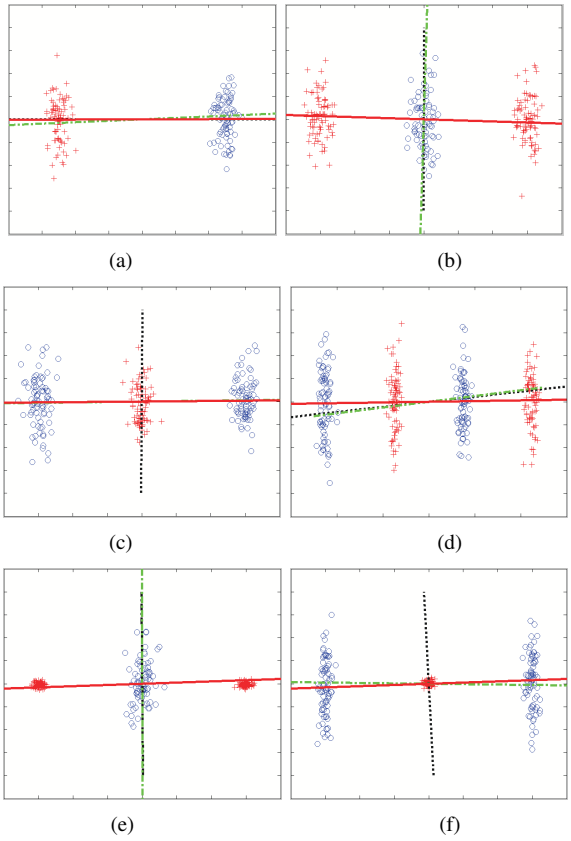


Fig. 7. The performance comparisons of three different subspace learning methods (i.e., LDA, BDA and CPSL) for two sets of samples (i.e., similar samples and dissimilar samples) in RF. In experiments, red "+" and blue "o" denote similar and dissimilar samples, respectively. Black dotted lines, green dot dash lines and red full lines indicate the LDA, the BDA and the CPSL, respectively. (a)-(f) show the experimental results of three subspace learning methods when handling samples with various different distributions, respectively.

"all positive examples are alike, and each negative example is negative in its own way", the BDA [9] was proposed to formulate the RF as a (1+x) class subspace learning problem. However, it is still not very reasonable to conclude that all positive feedbacks come from one class with a Gaussian distribution. Actually, each positive feedback is similar with each of the remaining positive feedbacks, and each negative feedback is dissimilar with each of the positive feedbacks. Consequently, different from traditional supervised learning problems (e.g., LDA and BDA), RF is intrinsically a weakly supervised learning problem and can involve only the similar and dissimilar pairwise constraints for feedback samples. Any unreasonable assumption for the class labels of feedback samples will result in performance degradation.

From Fig.7, we clearly see that LDA can find the best discriminative direction only when the set of positive feedbacks and the set of the negative feedbacks are distributed as Gaussian with similar covariance matrices, as shown in Fig.7(a), but may be confused when the distribution of the feedbacks is more complicated, as given in Fig.7(b), (c), (d), (e) and (f). Regarding RF as a (1+x) class problem, BDA can only find the direction that positive feedbacks are well separated with the negative feedbacks when the positive feedbacks

have Gaussian distribution, e.g., Fig.7(c) and (f). However, the BDA may also be confused when the distribution of positive feedbacks is more complicated, as shown in Fig.7(b), (d) and (e). The discriminative loss function in the CPSL method only involves the local similar and dissimilar pairwise constraints of feedback samples and does not impose any label constraints on the feedback samples, which is more appropriate for RF in image retrieval. Consequently, the discriminative loss function in CPSL can effectively find the discriminative subspace comparing with classical supervised subspace learning methods with explicit label information in RF.

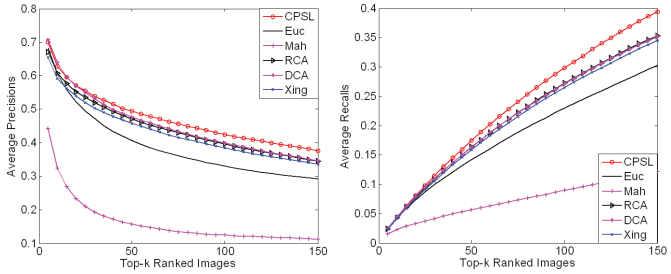
B. Experiments on the CIR system with historical feedback log data

In this subsection, we will evaluate the effectiveness of the proposed CPSL method based on two experiments: firstly, we investigate the CPSL method by exploiting the historical feedback log data for an image retrieval task without a RF scheme. And then, we show the performance of our CIR system by exploiting the user on-line historical feedback log data and compare it with a regular RF scheme (i.e., SVM RF) based image retrieval system on a large real-world Corel Image Database.

1) Performance evaluation by exploiting the feedback log database for image retrieval: In this part, we intend to examine if the proposed algorithm is comparable or better than the previous representative weakly supervised metric learning techniques in the distance metric learning community. We compare the CPSL method with two major distance metrics (i.e., the Euclidean metric and the Mahalanobis metric), three representative weakly supervised metric learning approaches (i.e., RCA [26], DCA [27] and Xing [25]). In experiments, we do not compare the proposed method with supervised learning techniques since they often require explicit class labels, which are not suitable for CIR. Moreover, in this subsection, the CPSL method does not involve any unlabeled samples for fair comparison with RCA, DCA and Xing. Parameters in each method were determined empirically to achieve its best performance in this paper. The parameter sensitivity of the CPSL method will be carefully analyzed in the next subsection.

All of the compared algorithms are implemented on two log databases as described in Section IV.B, i.e., a log database with 200 log images and a log database with 600 log images. In experiments, 500 queries are first randomly selected from the database and then the image retrieval is automatically done by a computer. We use Average Precision (AP) and Average Recall (AR) to evaluate the performance of compared algorithms. The AP refers to the percentage of relevant images in top ranked images presented to the user and is calculated as the averaged values of all the queries. The AR shows the fraction of the related images that are successfully retrieved and is defined as the percentage of the retrieved images among all relevant images in the database. In experiments, we calculated the APs and the ARs over the 500 queries at different positions from top 10 to top 150 to obtain the AP and AR curves.

Fig.8 shows the experimental results of the compared algorithms on the database with 200 log images. The detailed results are given in Table I and Table II. From the results, we can draw several observations. Firstly, we notice that directly using the Euclidean distance metric in a high dimensional visual feature space is not proper due to the semantic gap. Moreover, a simple Mahalanobis distance metric does not outperform the Euclidean distance metric. In fact, when the number of the log data (i.e., 200 log images) is much less than the dimension of the image features (i.e., 510 dimension), the covariance of the log data is singular, which significantly degrades the performance of the Mahalanobis distance metric for image retrieval. To avoid the singular problem, the regularization item $(\sigma^2 I, \sigma^2 = 0.01)$ is added to the covariance matrix in experiments, which is widely used to enhance the generalization property of the algorithm. And then, all of the metric learning methods (i.e., RCA, DCA, Xing and CPSL) can perform better than the Euclidean distance metric by exploiting the log data. In experiments, the optimal metric learned by RCA is computed as the inverse of the average covariance matrix of the chunklets. Similar to the Mahalanobis distance metric, the RCA will also encounter the singular covariance matrix when dealing with high-dimensional images. In experiments, the RCA is preceded by constraints based LDA which reduces the dimension to that of the CPSL method as described in [26]. By doing this, we notice that the RCA can show much better performance than the Euclidean distance metric by exploiting the similar pairwise constraints. The DCA incorporated the dissimilar constraints into the RCA and was formulated into a trace ratio problem. In [27], the authors proposed to attack this problem by using a direct method as in the fisher's LDA [52]. However, much discriminative information in the null space of the dissimilar scatter has been discarded in solving this problem [53]. Although the DCA incorporates the dissimilar pairwise constraints into the RCA, the performance of the DCA has been significantly degraded due to the problem of numerical computation in handling this trace ratio problem. Actually, the DCA cannot show better performance than the RCA for some results, as shown in top 70 to top 100 results in Table II. Xing et al formulated the weakly supervised metric learning into a convex optimization problem, which can be solved by an iterative projection algorithm. However, this method will involve a high computational burden when dealing with high dimensional images (i.e, 510 dimension in this paper), which is always the case in CBIR. The CPSL



(a) (b)

Fig. 8. Average Precision curves and Average Recall curves of the six compared methods for the 200 log data(i.e.,(a) Average Precision, (b) Average Recall))

TABLE I
AVERAGE PRECISIONS IN TOP N RESULTS OF THE SIX COMPARED METHODS (I.E., EUCLIDEAN DISTANCE METRIC, MAHALANOBIS DISTANCE METRIC, RCA, DCA, XING AND CPSL) FOR THE LOG DATABASE WITH 200 LOG IMAGES

Top	10	20	30	40	50	60	70	80	90	100
Euc	55.71	49.46	45.10	41.97	39.46	37.62	36.13	34.78	33.62	32.52
Mah	26.93	21.02	18.10	16.40	15.21	14.28	13.49	12.96	12.63	12.33
RCA	57.68	53.54	50.57	48.20	46.06	44.46	43.02	41.54	40.20	38.97
DCA	59.59	55.08	51.37	48.73	46.68	44.98	43.33	41.79	40.51	39.32
Xing	56.20	52.02	48.95	46.76	44.95	43.20	41.61	40.35	39.12	37.94
CPSL	59.52	55.45	52.72	50.23	48.63	47.11	45.62	44.29	43.09	41.89

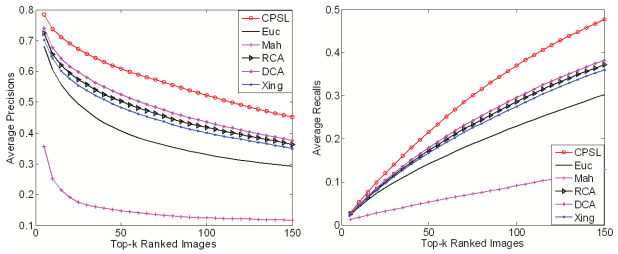
TABLE II
AVERAGE RECALLS IN TOP N RESULTS OF THE SIX COMPARED METHODS (I.E., EUCLIDEAN DISTANCE METRIC, MAHALANOBIS DISTANCE METRIC, RCA, DCA, XING AND CPSL) FOR THE LOG DATABASE WITH 200 LOG IMAGES

Top	10	20	30	40	50	60	70	80	90	100
Euc	5.84	8.63	11.01	13.16	15.09	17.03	18.87	20.54	22.21	23.71
Mah	2.87	3.75	4.55	5.30	6.02	6.68	7.30	7.96	8.64	9.29
RCA	6.06	9.34	12.30	15.05	17.52	19.95	22.23	24.29	26.26	28.10
DCA	6.18	9.50	12.32	15.00	17.56	19.98	22.15	24.18	26.16	28.02
Xing	5.91	9.09	11.93	14.61	17.13	19.41	21.54	23.64	25.58	27.39
CPSL	6.31	9.79	13.03	15.95	18.85	21.56	24.07	26.46	28.75	30.85

can learn a distance metric M by resorting to the mapping matrix W and solve the formulated constrained function with a standard Eigen value decomposition method, which is much effective and efficient when handling high dimensional images and never meets the problem of numerical computation.

From the results, we can see that the proposed CPSL can significantly outperform the two major distance metrics and three compared metric learning approaches for overall evaluation. Moreover, we also conduct the same comparisons on the database with 600 log images and the results are shown in Fig. 9, Table III and Table IV. Similar to the experimental results on the database with 200 log images, the proposed CPSL method can also show much better performance than the compared weakly supervised metric learning algorithms when dealing with 600 log images. Additionally, the performance of each of the weakly supervised learning algorithms on the 600 log data is much better than the corresponding results on the 200 log data since more training samples are involved to train a reliable distance metric for image retrieval. It should be noted that the results of the Euclidean distance metric on 600 log data is the same as the corresponding results on 200 log data since no training procedure is involved. Comparing with the results on 200 log data, the Mahalanobis distance metric cannot show better performance on 600 log data since the similar and dissimilar constraints are actually not utilized to calculate the metric. Moreover, it is difficult to obtain a reliable and stable Mahalanobis distance metric when the number of log data is small and the dimension of the data is high. Therefore, it is not proper to directly use the Mahalanobis distance metric for image retrieval when exploiting the user historical log data.

2) Performance evaluation on our CIR system: In this part, we show the performance of our CIR system by exploiting the user on-line feedback log data on a large database with 10,763 Corel images and compare it with a regular RF scheme based image retrieval system. The SVM based RF scheme is



(a) (b)

Fig. 9. Average Precision curves and Average Recall curves of the six compared methods for the 600 log data(i.e.,(a) Average Precision, (b) Average Recall))

one of the most popular techniques for image retrieval, which considers the RF as a strict two-class on-line classification problem. But it totally ignores the distinct properties of the two groups of training feedbacks, that is, all positive feedbacks share a common concept while each negative feedback differs in various concepts. Moreover, it does not take into account the unlabeled samples although they are very helpful in constructing a good classifier. With the assumption that different semantic concepts live in different subspaces and each image can live in many subspaces, it is the goal of RF schemes to figure out "which one". However, it will be a burden for the SVM RF schemes to tune the internal parameters to adapt to the changes of the subspaces. In this subsection, we show that our CIR system can effectively address the two drawbacks by off-line exploiting the user on-line historical feedback log data.

The experiments are simulated by a computer automatically. First, 400 queries are randomly selected from the database and the RF is automatically done by a computer. At each round of RF, the first 3 relevant images are marked as positive feedbacks and all the other irrelevant images in top 20 results

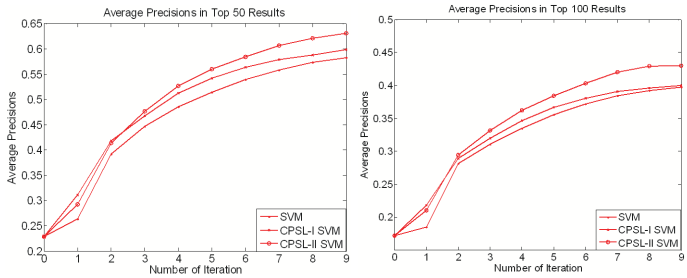
TABLE III
AVERAGE PRECISIONS IN TOP N RESULTS OF THE SIX COMPARED METHODS (I.E., EUCLIDEAN DISTANCE METRIC, MAHALANOBIS DISTANCE METRIC, RCA, DCA, XING AND CPSL) FOR THE LOG DATABASE WITH 600 LOG IMAGES

Top	10	20	30	40	50	60	70	80	90	100
Euc	55.71	49.46	45.10	41.97	39.46	37.62	36.13	34.78	33.62	32.52
Mah	21.45	17.60	16.09	15.18	14.51	13.87	13.36	12.86	12.62	12.44
RCA	62.04	57.44	54.03	51.44	49.22	47.25	45.44	43.75	42.40	41.19
DCA	64.25	59.86	56.32	53.68	51.36	49.39	47.51	45.76	44.29	42.87
Xing	60.09	55.48	52.26	49.49	47.25	45.33	43.71	42.02	40.71	39.50
CPSL	71.11	67.33	64.35	61.91	59.93	58.10	56.53	54.81	53.08	51.47

TABLE IV
AVERAGE RECALLS IN TOP N RESULTS OF THE SIX COMPARED METHODS (I.E., EUCLIDEAN DISTANCE METRIC, MAHALANOBIS DISTANCE METRIC, RCA, DCA, XING AND CPSL) FOR THE LOG DATABASE WITH 600 LOG IMAGES

Top	10	20	30	40	50	60	70	80	90	100
Euc	5.84	8.63	11.01	13.16	15.09	17.03	18.87	20.54	22.21	23.71
Mah	2.32	3.16	4.05	4.92	5.75	6.50	7.24	7.92	8.70	9.51
RCA	6.44	9.94	13.06	15.95	18.61	21.09	23.38	25.47	27.57	29.61
DCA	6.68	10.34	13.57	16.55	19.30	21.90	24.27	26.44	28.59	30.56
Xing	6.28	9.64	12.68	15.43	17.97	20.34	22.58	24.56	26.57	28.47
CPSL	7.63	11.99	16.01	19.76	23.38	26.79	30.07	33.03	35.72	38.28

are marked as negative feedbacks. The procedure is close to real-world circumstances since the irrelevant images usually largely outnumber the relevant ones in a real-world image retrieval system.



(a) (b)

Fig. 10. The APs in top 50 and top 100 results of the three different RF schemes for image retrieval(i.e., SVM RF, CPSL-I SVM RF and CPSL-II SVM RF))

We compare the regular SVM RF with two new SVM RF schemes, i.e., CPSL-I SVM RF and CPSL-II SVM RF. The regular SVM implements the RF task in the original high dimensional low-level visual feature space. The CPSL-I SVM RF first exploits the user on-line historical feedback log data by finding a semantic subspace, in which all positive feedbacks are clustered and all negative feedbacks are separated with all positive feedbacks as much as possible. And then the SVM implements the RF in this reduced semantic subspace. The CPSL-II SVM RF incorporates the information of unlabeled samples into the CPSL-I SVM RF through a regularized learning framework. In experiments, the CPSL-I method and the CPSL-II method are implemented by setting the trade-off parameter $\beta_2 = 0$ and $\beta_2 = 1/(n^u)$, respectively. For patch building parameters, we set $k_1, k_2, k_3 = 4$ according to manifold learning approaches [30], [29], [31], [32], [33]. Considering the computable efficiency, we randomly select $n_u = 400$ unlabeled images in each round of RF iteration. The optimal dimensionality of the reduced subspace for the CPSL-I method and the CPSL-II method is empirically set in experiments to preserve more geometry information of images. For all SVM-based algorithms, we empirically set the kernel parameters to achieve the best performance in experiments. Fig.10 (a) and (b) show the APs in top 50 and top 100 retrieved results, respectively.

The CPSL-I method can effectively exploit the user on-line feedback log data and find a semantic concept subspace, in which all positive feedbacks are clustered and all negative feedbacks are separated with positive feedbacks as much as possible. And then the SVM RF is implemented in this reduced semantic subspace for an image retrieval task. From the results, we notice that the CPSL-I SVM RF can outperform the regular SVM RF by exploiting the user on-line feedback log data. However, the performance difference between CPSL-I SVM RF and the regular SVM RF gets smaller after a few rounds of RF because of the overfitting problem. The CPSL-II SVM RF method can effectively integrate the information of unlabeled samples through a regularized learning framework into the construction of the classifier and alleviate the overfitting problem encountered by the CPSL-I SVM RF. As shown in Fig.10, when considering more RF iterations, the CPSL-II SVM RF is more effective than both of the CPSL-I SVM RF and the regular SVM RF.

In experiments, the mapping matrix W can be obtained by using the Eigen value decomposition. The time cost to calculate W is $O\left((n+n_u)^3\right)$. Afterwards, we project all images to this semantic subspace and then apply the new similarity metric with respect to the query to sort all images in the database. The time cost for calculating the Euclidean distance

in the semantic subspace L between the query and all images in the database is $O(NL)$, wherein N is the cardinality of the database. Therefore, for a query image, the time cost for CPSL based CBIR system is $O\left((n+n_u)^3\right)+O(NL)$. And the time cost for a conventional CBIR system in the high dimensional visual feature space H is $O(NH)$. Usually, for a CBIR system, the cardinality of the database N is very large and $H >> L$; therefore, the proposed method is very effective for an image retrieval task.

C. Parameter sensitivity

In this subsection, we study the parameter sensitivity of the CPSL method for an image retrieval task. The analyses are performed based on the experiments conducted on two log databases (i.e., 200 log data and 600 log data). In experiments, we analyze some factors: k_1 and k_2 in Eq.(5) for patch building, the trade off parameter β_1 in Eq.(13) and the dimension of the projected features for the CPSL method. Firstly, 500 query images are randomly selected from the database, and then the image retrieval process is automatically done by a computer. The APs in top 50 results is utilized for overall performance evaluation.

1) Evaluation on the number of neighboring samples: The two parameters k_1 and k_2 in Eq.(5) play an important role in building the local discriminative patch, which is the most critical aspect in CPSL. Generally, for a local discriminative patch, k_1 is the number of similar images which are involved to describe the compactness of the patch, and k_2 is the number of dissimilar images which are used to characterize the dispersiveness of the patch. Both of the two parameters (i.e., k_1 and k_2) reveal the data information from different aspects. In experiments, the trade-off parameter β_1 is set as 0 for alleviating the effect of the geometrical information and the reduced dimension for the two sets of log images is empirically fixed at 11 and 17, respectively. By varying k_1 and k_2, Figs.11(a) and (b) show the AP surface of CPSL subject to different k_1 and k_2 for the two log databases, respectively. From Fig.11, we can notice that the two parameters k_1 and k_2 can significantly affect the performance of the CPSL method in learning a subspace for an image retrieval task. As given in Fig.11(a), when k_1 and k_2 are larger than 4 and 10, respectively, the system can show much stable performance for 200 log images. Similarly, in Fig.11(b), when k_1 and k_2 are larger than 8 and 10, respectively, the CPSL method can achieve more satisfying results for 600 log images. Generally, smaller values of k_1 and k_2 mean that fewer similar and dissimilar images are involved to construct the local discriminative patch, and therefore insufficient training data lead to the degenerated performance of the system.

2) Evaluation on the trade-off parameter β_1: Empirically, the geometry information is useful for finding the semantic subspace. In this part, we turn to investigate the influence of the trade-off parameter β_1 in Eq.(13) for CPSL when building the local discriminative patch and the local geometrical patch for labeled log images. A small β_1 reflects the importance of separating dissimilar samples from similar ones, i.e., the CPSL focuses on the local discriminative information but ignores the local geometrical information. Fig.12 shows the performance of CPSL with different β_1, from which we can have the following observations.

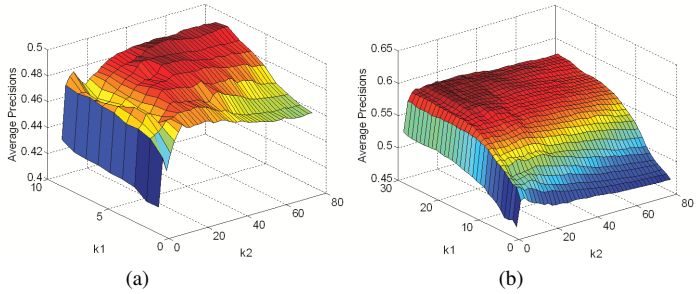


Fig. 11. The AP surface of the CPSL algorithm subject to different k1 and k2 for two log database (i.e.,(a) 200 log data and (b) 600 log data)

When β_1 is small, e.g., $\beta_1 = 0$, the performance is unsatisfactory. This is because that in this situation the local discriminative information is mainly preserved while important local geometrical information within labeled images with similar pairwise constraints is less considered. The performance of the CPSL increases when β_1 is growing and reaches the optimal value at $\beta_1 = 5$. And then, the APs decrease when β_1 is larger than this best setup, in which case the local geometrical information dominates the local patch and the local discriminative information is ignored.

Therefore, both the discriminative information and the geometrical information can reflect the important information contained in local patches from different aspects for complimentary. A suitable combination of them is essential to achieve good performance for the CPSL method.

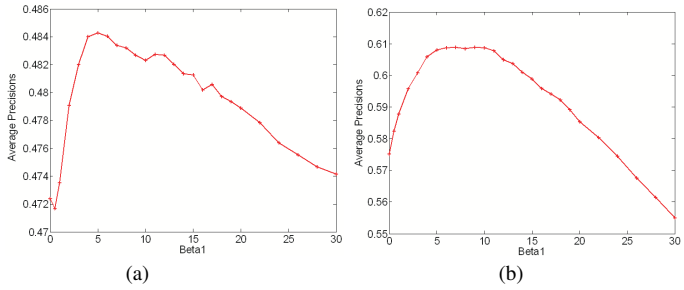
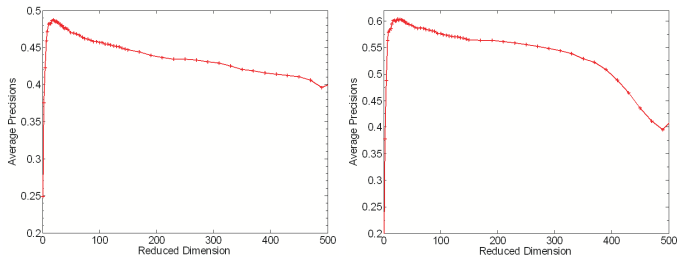


Fig. 12. Performance of CPSL with different β_1 for the two log database (i.e., (a) 200 log data and (b) 600 log data)

3) Evaluation on the projected subspace: Different from the weakly supervised distance metric learning methods [26], [27], [25], the proposed CPSL method aims to learn a mapping matrix, which can find a low dimensional subspace from the original high dimensional space. To find out an appropriate dimension of the projected semantic subspace, we have investigated the influence of the dimension in the following experiments. Fig.13 shows the performance of CPSL with features projected onto the subspaces with different dimensions. From Fig.13, we can notice that when the projected dimension is too low, (e.g., less than 11 and 17, respectively), the reduced subspace is insufficient to encode the semantic concepts of images, which makes the retrieval performance poor. When

the dimension equals or closes to that of the original high dimensional space (i.e., 510 in this paper), no or less benefit can be obtained from this subspace learning method. From the experimental results, we can notice that the CPSL method can achieve its best performance with the dimension of 11 and 17 for the two log databases, respectively. Moreover, lower dimensional data can lead to a less computational cost than higher dimensional data for an image retrieval task.



(a) (b)

Fig. 13. Performance of CPSL with features projected onto the subspaces with different dimensions for two log databases (i.e.,(a) 200 log data and (b) 600 log data)

D. Discussions and future work

In the proposed image retrieval system, several aspects can be improved. For instance, a much larger image database will be utilized in the current platform. Recently, CBIR based on a large scale social web database (e.g., 1 million Flickr images) has attracted much attention [3], [54]. In these systems, large scale social web images are first selected from social web sites (e.g., Flickr) and then manually grouped into semantic classes according to the associated textual information. However, different users have different opinions on a same web image (e.g., a Flickr image), and thus will categorize the same image into different semantic groups. The CBIR results from such image databases created by different people will be subjective and are difficult to objectively evaluate or compare. Moreover, the images from the social web image database (e.g., Flickr images) are often down-sampled and compressed by the web server, and thus illustrate considerable appearance differences from another database (e.g., Corel Image Database). Consequently, the images from different sources (e.g., Flickr and Corel Photo Gallery) may appear significantly different visual feature distributions in terms of statistical properties (i.e., mean, intra-class and inter-class variance) although they actually share a same semantic concept. Therefore, it will be interesting but still an open question to fairly evaluate the performance of a CBIR system based on large scale images from multiple sources (e.g., Flickr, Corel Photo Gallery and Personal Photo Database).

Basically, devising a reasonable similarity metric, which can reflect the semantic relation between a pair of images, plays a key role for an image retrieval task. The similarity metric learned from the training data can be well generalized to the testing data in the same database (e.g., Corel Image Database in this paper). However, due to the tremendously distribution differences between images from one data source (e.g., Corel Photo Gallery) and images from other data sources (e.g., INRIA Holidays Dataset [55], Flickr) in terms of various statistical properties, the similarity metric learned from one data source (e.g., Corel Photo Gallery) cannot be directly applied to the image data from other data sources (e.g., INRIA Holidays Dataset [55], Flickr). Recently, cross domain learning (a.k.a., domain adaptation or transfer learning) methods [56], [57] have been identified to be an effective scheme to address this problem, i.e., performing the learning task in the source domain and applying the learned model to the target domain which is usually governed by a different distribution from the source domain. Therefore, it is very promising to combine the proposed method with cross domain learning schemes for future studies.

To enhance the retrieval performance, the indexing of database is very important for a CBIR system. Generally, there are two types of image indexing methods [1], [3]. Classification based indexing technique aims to improve the retrieval precision of the system [58]. In this method, each image in the database is assigned one or more distinct labels. Then, based on these labels, the indexing of database can be constructed through their associated semantic labels. Therefore, the search results will be more satisfactory and cater to most of the users. The other indexing method is the low level visual feature based indexing technique [59], which can be used to speed up the retrieval procedure. There are many low level visual feature based indexing techniques, e.g., various tree-based indexing structures for high dimensional data. The two indexing methods have their respective advantages from different aspects. As a consequence, it is promising to combine the classification and visual feature information in the indexing structures to improve both of the retrieval precision and speed.

VI. Conclusion

In this paper, we have studied the problem of subspace learning with side information and presented a novel subspace learning technique, termed as Conjunctive Patches Subspace Learning (CPSL) with side information, to exploit the user historical feedback log data for a Collaborative Image Retrieval (CIR) task. The proposed scheme can effectively integrate the discriminative information of labeled log images, the geometrical information of labeled log images and the weakly similar information of unlabeled images together through a regularized learning framework. We have formally formulated this subspace learning problem into a constrained optimization task and then present an effective algorithm to solve this problem with closed-form solutions. Extensive experiments on both synthetic data sets and a real-world Corel image database have shown the effectiveness of the proposed scheme in exploiting the user historical feedback log data for CIR.

Acknowledgment

The authors would like to thank the handling associate editor and the anonymous reviewers for their constructive comments on this manuscript in the three rounds of review. The authors would also like to acknowledge the Ph.D. grant

from the Institute for Media Innovation, Nanyang Technological University, Singapore. This work was partially supported by the SINGAPORE MINISTRY OF EDUCATION Academic Research Fund (AcRF) Tier 2, Grant Number: T208B1218.

REFERENCES

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

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

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

[4] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644 –655, Sept. 1998.

[5] X.S. Zhou and T.S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Systems*, vol. 8, no. 6, pp. 536–544, Apr. 2003.

[6] Y. Rui and T.S. Huang, "Optimizing learning in image retrieval," in *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition*, 2000, pp. 236 –243.

[7] Y. Chen, X. S. Zhou, and T.S. Huang, "One-class svm for learning in image retrieval," in *Proceedings IEEE International Conference on Image Processing*, 2001, pp. 34 –37.

[8] P. Hong, Q. Tian, and T.S. Huang, "Incorporate support vector machines to content-based image retrieval with relevance feedback," in *Proceedings IEEE International Conference on Image Processing*, 2000, pp. 750 –753.

[9] X. S. Zhou and T.S. Huang, "Small sample learning during multimedia retrieval using biasmap," in *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition*, 2001, pp. 11 –17.

[10] D. Tao, X. Tang, X. Li, and Y. Rui, "Direct kernel biased discriminant analysis: a new content-based image retrieval relevance feedback algorithm," *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 716 –727, 2006.

[11] L. Wang, K. L. Chan, and P. Xue, "A criterion for optimizing kernel parameters in kbda for image retrieval," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 3, pp. 556 –562, 2005.

[12] D. Xu, S. Yan, D. Tao, S. Lin, and H. Zhang, "Marginal fisher analysis and its variants for human gait recognition and content- based image retrieval," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2811 –2821, 2007.

[13] X. He, W. Ma, and H. Zhang, "Learning an image manifold for retrieval," in *Proceedings of the 12th Annual ACM International Conference on Multimedia*, New York, NY, USA, 2004, pp. 17–23.

[14] J. He, M. Li, H. Zhang, H. Tong, and C. Zhang, "Manifold-ranking based image retrieval," in *Proceedings of the 12th Annual ACM International Conference on Multimedia*, New York, NY, USA, 2004, pp. 9–16.

[15] H. Muller, T. Pun, and D. Squire, "Learning from user behavior in image retrieval: Application of market basket analysis," *International Journal of Computer Vision*, vol. 56, pp. 65–77, 2004.

[16] C.H. Hoi and M. R. Lyu, "A novel log-based relevance feedback technique in content-based image retrieval," in *Proceedings of the 12th Annual ACM International Conference on Multimedia*, New York, NY, USA, 2004, pp. 24–31.

[17] C.H. Hoi, M.R. Lyu, and R. Jin, "A unified log-based relevance feedback scheme for image retrieval," *IEEE Transactions. on Knowledge and Data Engineering*, vol. 18, no. 4, pp. 509 – 524, 2006.

[18] L. Si, R. Jin, C.H. Hoi, and M.R. Lyu, "Collaborative image retrieval via regularized metric learning," *Multimedia Systems*, vol. 12, pp. 34–44, 2006.

[19] C.H. Hoi, W. Liu, and S.F. Chang, "Semi-supervised distance metric learning for collaborative image retrieval and clustering," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 6, no. 3, pp. 1–26, 2010.

[20] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of educational psychology*, vol. 24, no. 6, pp. 417, 1933.

[21] X. He, M. Ji, and H. Bao, "Graph embedding with constraints," in *Proceedings of the 21st International Jont Conference on Artifical Intelligence*. Morgan Kaufmann Publishers Inc., 2009, pp. 1065–1070.

[22] S. Si, D. Tao, and K.P. Chan, "Evolutionary cross-domain discriminative hessian eigenmaps," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 1075–1086, 2010.

[23] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328 –340, 2005.

[24] D. Tao, X. Li, X. Wu, and S.J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700 –1715, 2007.

[25] E.P. Xing, A.Y. Ng, M.I. Jordan, and S. Russell, "Distance metric learning with application to clustering with side-information," *Advances in Neural Information Processing Systems*, pp. 521–528, 2003.

[26] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning a mahalanobis metric from equivalence constraints," *Journal of Machine Learning Research*, vol. 6, no. 1, pp. 937–965, 2006.

[27] C.H. Hoi, W. Liu, M.R. Lyu, and W. Ma, "Learning Distance Metrics with Contextual Constraints for Image Retrieval," in *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2072–2078.

[28] M.A.A. Cox and T.F. Cox, "Multidimensional scaling," *Handbook of data visualization*, pp. 315–347, 2008.

[29] S.T. Roweis and L.K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323, 2000.

[30] J.B. Tenenbaum, V. Silva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319, 2000.

[31] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems*, Cambridge, MA, 2002, MIT Press.

[32] X. He and P. Niyogi, "Locality Preserving Projections," in *Advances in Neural Information Processing Systems*, Cambridge, MA, 2004, MIT Press.

[33] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40 –51, 2007.

[34] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," in *Advances in Neural Information Processing Systems*, Cambridge, MA, 2004, MIT Press.

[35] K.Q. Weinberger and L.K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[36] A. Globerson and S. Roweis, "Metric learning by collapsing classes," in *Advances in Neural Information Processing Systems*, Cambridge, MA, 2006, MIT Press.

[37] J.V. Davis and I.S. Dhillon, "Structured metric learning for high dimensional problems," in *Proceeding of the 14th ACM International Conference on Knowledge Discovery and Data mining*. ACM, 2008, pp. 195–203.

[38] L. Wu, R. Jin, S.C.H. Hoi, J. Zhu, and N. Yu, "Learning bregman distance functions and its application for semi-supervised clustering," in *Advances in Neural Information Processing Systems*, Cambridge, MA, 2009, MIT Press.

[39] A. Ghodsi, D. Wilkinson, and F. Southey, "Improving embeddings by flexible exploitation of side information," in *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, San Francisco, CA, USA, 2007, pp. 810–816, Morgan Kaufmann Publishers Inc.

[40] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural computation*, vol. 7, no. 2, pp. 219–269, 1995.

[41] V.N. Vapnik, *The nature of statistical learning theory*, Springer Verlag, 2000.

[42] L.Wang, *Support Vector Machines: Theory and Applications*, Springer Berlin, 2005.

[43] L.Wang and X.Fu, *Data Mining with Computational Intelligence*, Springer Berlin, 2005.

[44] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Proceedings IEEE International Conference on Computer Vision*, 2007.

[45] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," *SIAM Journal of Scientific Computing*, vol. 26, pp. 313–338, 2002.

[46] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1299–1313, 2009.
[47] Xiaofei He, Deng Cai, and Jiawei Han, "Learning a maximum margin subspace for image retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 2, pp. 189 –201, 2008.
[48] Wei Bian and Dacheng Tao, "Biased discriminant euclidean embedding for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 545 –554, 2010.
[49] M.J. Swain and D.H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.
[50] J. Chen, S. Shan, C. He, G. Zhao, P. Matti, X. Chen, and W. Gao, "Wld: A robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705 –1720, 2010.
[51] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, no. 8, pp. 1233 – 1244, 1996.
[52] H. Yu and J. Yang, "A direct lda algorithm for high-dimensional data – with application to face recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067 – 2070, 2001.
[53] H. Gao and J. W. Davis, "Why direct lda is not equivalent to lda," *Pattern Recognition*, vol. 39, no. 5, pp. 1002 – 1006, 2006.
[54] Jia Deng, A.C. Berg, and Li Fei-Fei, "Hierarchical semantic indexing for large scale image retrieval," in *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition*, 2011, pp. 785 –792.
[55] Herve Jegou, Matthijs Douze, and Cordelia Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *Processding 10th European Conference on Computer Vision*, 2008, pp. 304–317.
[56] Bo Geng, Dacheng Tao, and Chao Xu, "Daml: Domain adaptation metric learning," *IEEE Transactions on Image Processing,*, vol. 20, no. 10, pp. 2980 –2989, 2011.
[57] Lixin Duan, Ivor W. Tsang, and Dong Xu, "Domain transfer multiple kernel learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 465 –479, march 2012.
[58] N. Vasconcelos, "Image indexing with mixture hierarchies," in *Proceedings IEEE International Conference on Computer Vision and Pattern Recognition*, 2001, pp. 3–10.
[59] A. Natsev, Rajeev Rastogi, and K. Shim, "Walrus: a similarity retrieval algorithm for image databases," *IEEE Transactions on Knowledge and Data Engineering,*, vol. 16, no. 3, pp. 301 – 316, mar 2004.

Lining Zhang (S'11) received the B.Eng. 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 and 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, and the Ph.D. degree in computer vision from King's College, London University, London, U.K. 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 over ten 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 has published over 190 refereed papers in international journals and conferences.

Dr. Lin is a Chartered Engineer (U.K.), a fellow of Institution of Engineering Technology, and an Honorary Fellow, Singapore Institute of Engineering Technologists. He organized special sessions in IEEE International Conference on Multimedia and Expo (ICME 2006, 2012), 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), and MobiMedia 2011. 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 cochairs 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, and elected as a Distinguished Lecturer of APSIPA (2012).