Multimedia Systems (2006) 12(1): 34–44
DOI 10.1007/s00530-006-0033-1

REGULAR PAPER

Luo Si · Rong Jin · Steven C.H. Hoi · Michael R. Lyu

Collaborative image retrieval via regularized metric learning

Published online: 11 April 2006

Abstract In content-based image retrieval (CBIR), relevant images are identified based on their similarities to query images. Most CBIR algorithms are hindered by the semantic gap between the low-level image features used for computing image similarity and the high-level semantic concepts conveyed in images. One way to reduce the semantic gap is to utilize the log data of users' feedback that has been collected by CBIR systems in history, which is also called "collaborative image retrieval." In this paper, we present a novel metric learning approach, named "regularized metric learning," for collaborative image retrieval, which learns a distance metric by exploring the correlation between low-level image features and the log data of users' relevance judgments. Compared to the previous research, a regularization mechanism is used in our algorithm to effectively prevent overfitting. Meanwhile, we formulate the proposed learning algorithm into a semidefinite programming problem, which can be solved very efficiently by existing software packages and is scalable to the size of log data. An extensive set of experiments has been conducted to show that the new algorithm can substantially improve the retrieval accuracy of a baseline CBIR system using Euclidean distance metric, even with a modest amount of log data. The experiment also indicates that the new algorithm is more effective and more efficient than two alternative algorithms, which exploit log data for image retrieval.

L. Si (✉)
Language Technology Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA
E-mail: lsi@cs.cmu.edu

R. Jin
Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA
E-mail: rongjin@cse.msu.edu

S. C. H. Hoi · M. R. Lyu
Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
E-mail: {chhoi, lyu}@cse.cuhk.edu.hk

Keywords Content-based image retrieval · Relevance feedback · Log-based relevance feedback · Relevance feedback log · Users · Semantic gap · Metric learning · Regularization · Semidefinite programming

1 Introduction

Content-based image retrieval (CBIR) has been an active research topic in the last decade [9, 24, 30]. Although substantial research has been conducted, CBIR is still an open research topic mainly due to the difficulty in bridging the gap between low-level feature representation and high-level semantic interpretation. Several approaches have been proposed to reduce the semantic gap and to improve the retrieval accuracy of CBIR systems. One promising approach is the online user feedback [1, 6, 10–15, 17, 18, 21, 26–28, 33]. It first solicits users' relevance judgments on the initial retrieval results for a given query image. It then refines the representation of the initial query with acquired user judgments, and re-runs the CBIR algorithm again with the refined representation. However, collecting feedback information in an online manner can be time-consuming and therefore inconvenient for users. Given the difficulty in learning users' information needs from their relevance feedback, usually multiple rounds of relevance feedback are required before satisfactory results are achieved, which can significantly limit its application to real-world problems.

An alternative approach to bypass the semantic gap is to index image databases with text descriptions and allow users to pose textual queries against image databases. To avoid the excessive amount of labor on manual annotation, automatic image annotation techniques, such as [4, 7, 20, 22], have been developed. However, text descriptions generated by automatic annotation techniques are often inaccurate and limited to a small vocabulary, and therefore is insufficient to accommodate the diverse information needs from users.

Recently, there have been several studies on exploring the log data of users' relevance feedback to improve image retrieval [12, 15, 16, 25, 36]. In these studies, the CBIR system collects relevance judgments from a number of users, which is also called "*log data*" in this paper. In addition to the low-level features, each image is also represented by the users' relevance judgments in log data. Most of these studies hypothesized that when two images are similar in their semantic content, they tend to be either favored or disliked simultaneously by many users. As a result, similar images tend to share similar representation in users' relevance judgments. In [25], several weighting schemes are proposed for the low-level image features that are based on log data. In [10, 11], a manifold learning algorithm is applied to learn a low-dimensional manifold from log data that better reflects the semantic relation among different images. In [15, 16], the log data of users' relevance judgments are used to improve relevance feedback techniques for image retrieval. We refer to the image retrieval approaches based on log data as "collaborative image retrieval."

In this work, we explore the log data of users' relevance judgments in a way that is different from the previous work. Unlike [25] where manually designed weighting schemes based on log data are used to measure similarity of images, in this work, we propose to automatically learn the distance metric for the low-level features from the users' relevance judgements in log data. We hypothesize that, in each user feedback session, when two images are judged as relevant, they tend to be more similar in content than the case when one image is judged as relevant and the other is judged as irrelevant. Thus, our goal is to search for an appropriate distance metric for the low-level features such that the distance in low-level features is consistent with the users' relevance judgments in log data. To this end, we propose the "Min/Max" principle, which tries to minimize the distance between similar images and meanwhile maximize the distance between the feature vectors of dissimilar images. Based on this principle, we propose a new algorithm for metric learning, named "regularized distance metric learning," in which a regularization mechanism is introduced to improve the robustness of the learning algorithm. The new algorithm can be formulated into a Semidefinite Programming (SDP) problem [34], and therefore can be solved efficiently by the existing package for SDP, such as SeDuMi [32], and is scalable to the size of log data.

Our work distinguishes from the previous work on exploiting log data for image retrieval in that it deals with the *real-world users* whereas much of the previous research used the synthesized log data in its study. In particular, we try to address the following challenging issues with the *real* log data:

- *Image retrieval with modest-sized log data.* Most previous studies assume that large amount of log data are available, and do not consider the scenarios when the size of log data is limited. Developing retrieval techniques for modest-sized log data is important, particularly when a CBIR system is in its early development and has not accumulated large numbers of relevance judgments from users. It is also important when the target images are not popular and are only equipped with a small number of users' relevance judgments.
- *Image retrieval with noisy log data.* Most previous studies assume that log data are clean and contain no noise. This is an unrealistic assumption given that users' relevance judgments are subjective and real-world users could make mistakes in their judgments. In our experiments with real-world users, we usually observed a number of erroneous relevance judgments, ranging from 5 to 15% of all judgments. As will be shown later in the empirical study, the noise in users' relevance judgments can significantly degrade the retrieval accuracy of a CBIR system.
- *Efficiency and scalability.* Most previous studies emphasize the effectiveness of their algorithms on improving CBIR. Few of them examine the efficiency and scalability of their algorithms. The issue of efficiency and scalability is extremely important for this technique to be practical, particularly when we have to deal with large-sized log data.

The rest of this paper is arranged as follows: the next section discusses the related research. Section 3 describes the proposed regularized metric learning algorithm. Section 4 explains our experimental methodology. Section 5 presents the experimental results. Section 6 discusses the limitation and future work. Section 7 concludes this work.

2 Related work

This work is related to previous studies on utilizing users' log data to enhance content-based image retrieval. It is also related to the research on distance metric learning. We will review the previous work on using log data first, followed by the review of metric learning algorithms.

Users' log data have been utilized in the previous work [15] to improve online relevance feedback for CBIR. In [15], the users' relevance judgments in log data is used to infer the similarities among images. For online retrieval, a set of relevant and irrelevant images are first obtained through the solicitation of users' relevance judgments. Then, based on the log data, images that are most similar to the judged ones are added to the pool of labeled examples, including both relevant and irrelevant images. A discriminative learning model, such as support vector machines (SVM) [5], is trained with the expanded pool of labeled images to improve the retrieval accuracy. This work differs from ours in that it requires online feedback from users, while our algorithm focuses on improving the accuracy of the initial around of image retrieval. Another recent research related to our work is to apply manifold learning to image retrieval [10, 11]. Their work has considered using log data for both CBIR with online feedback and CBIR without online feedback. Using the Laplacian Eigenmap [3], they constructed a low-dimensional semantic space for the low-level image features using log data. Given the complicated distributions of image features, constructing a robust manifold for image features usually requires a large number of training data. In fact, according to our experiments, their algorithm works well when large numbers of users' relevance judgments are available. Its advantage appears to fade away when the size of log data is small. Finally, there are studies on designing weighting schemes for low-level image features based on log data [25]. In [25], weighting schemes, similar to the TF.IDF methods in text retrieval [29], have been proposed and computed based on the log data of users' relevance judgments.

Another group of related work is the learning of distance metric [23, 35]. One of the well-known research on this subject is [35], which learns a distance metric under pairwise constraints. As it serves as the baseline in this study, we briefly describe it here.

Let $\mathcal{C} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ be the data collection where n is the number of data points in the collection. Each $\mathbf{x}_i \in \mathbb{R}^m$ is a feature vector where m is the number of features. Let $\mathcal{S}$ be the set that contains pairs of similar data points, and $\mathcal{D}$ be the set that contains pairs of dissimilar data points. More precisely, we have

$$\begin{aligned}\mathcal{S} = \{(\mathbf{x}_i, \mathbf{x}_j) \mid &\text{ data points } \mathbf{x}_i \text{ and } \mathbf{x}_j \\ &\text{are likely to belong to the same class}\}\\ \mathcal{D} = \{(\mathbf{x}_i, \mathbf{x}_j) \mid &\text{ data points } \mathbf{x}_i \text{ and } \mathbf{x}_j \\ &\text{are unlikely to be in the same class}\}\end{aligned} \tag{1}$$

Let $\mathbf{A} \in \mathbf{S}^{m\times m}$ be the distance metric to be learned, which is a symmetric matrix of size $m \times m$. Then, for any two vectors $\mathbf{x}, \ \mathbf{y} \in \mathbb{R}^m$, their distance is expressed as:

$$\begin{aligned}d_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\|_{\mathbf{A}} = \sqrt{(\mathbf{x} - \mathbf{y})^{\mathrm{T}}\mathbf{A}(\mathbf{x} - \mathbf{y})}\\ &= \mathrm{tr}(\mathbf{A} \cdot (\mathbf{x} - \mathbf{y})(\mathbf{x} - \mathbf{y})^{\mathrm{T}})\end{aligned} \tag{2}$$

where product "·" is a pointwise matrix multiplication, and "tr" stands for the trace operator that computes the sum of diagonal elements of a matrix.

A is a valid metric as long as the distance between any two data points is nonnegative and satisfies the triangle inequality. This requirement is formalized as the positive semidefinite constraint for matrix **A**, i.e., $\mathbf{A} \succeq 0$ [34]. Furthermore, matrix **A** should be symmetric, namely $\mathbf{A} = \mathbf{A}'$. Note when **A** is an identity matrix $\mathbf{I}_{m\times m}$, the distance in Eq. (2) becomes

$$d_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^{\mathrm{T}}\mathbf{I}(\mathbf{x} - \mathbf{y})} = \sqrt{(\mathbf{x} - \mathbf{y})^{\mathrm{T}}(\mathbf{x} - \mathbf{y})}$$

Thus, we go back to the Euclidean distance.

Given the pairwise constraints in (1), [35] formulated the problem of metric learning into the following convex programming problem:

$$\begin{aligned}&\min_{\mathbf{A}} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}}^2\\ &\text{s. t.} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}}^2 \geq 1\\ &\quad \mathbf{A} \succeq 0\end{aligned} \tag{3}$$

In the above, optimal metric **A** is found by minimizing the sum of squared distance between pairs of similar data points, and meanwhile satisfying the constraint that the sum of squared distance between dissimilar data points is larger than 1. In other words, this algorithm tries to minimize the distance between similar data points and maximize the distance between dissimilar data points at the same time. This is consistent with our Min/Max principle discussed in the Introduction section.

The algorithm in (3) has been shown to be successful on several machine learning testbeds [35]. But one potential problem with this method is that it does not address the issue of robustness, which is important when training data are noisy or the amount of training data is limited. Our algorithm is able to improve the robustness of metric learning by introducing a regularizer into the objective function, which is similar to the strategy used in large margin classifiers [5]. Furthermore, the optimization problem in (3) may not be solved efficiently since it does not fall into any special class of convex programming, such as quadratic programming [8] and semidefinite programming [34]. In contrast, the proposed algorithm belongs to the family of semidefinite programming, which can be solved much more efficiently.

3 Regularized metric learning and its application to CBIR

As is discussed in the Introduction section, the basic idea of this work is to learn a desired distance metric in the space of low-level image features that effectively bridges the semantic gap. It is learned from the log data of users' relevance feedback based on the Min/Max principle, i.e., minimize/maximize the distance between the feature vectors of similar/dissimilar images. Log data, in this study, consist of a number of log sessions and each session corresponds to a different user query. In each log session, a user submits a query image to the CBIR system. After the initial results are retrieved by the CBIR system, the user provides relevance judgments for the top ranked images (i.e., 20 images in our experiment). To exploit the metric learning algorithm in (3) for log data, we convert binary relevance judgments into pairwise constraints as in (1). In particular, within each log session, images judged as relevant are regarded as similar to each other, and each dissimilar pair will consist of one relevant image and one irrelevant image. Thus, for each user query q, we have a set $\mathcal{S}_q$ for pairs of similar images and a set $\mathcal{D}_q$ for pairs of dissimilar images. Based on this treatment, we can now apply the framework in (3) to learn a distance metric $\mathbf{A}$ for low-level image features, i.e.,

$$\begin{aligned}
&\min_{\mathbf{A}} \sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}_q} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}}^2 \\
&\text{s. t.} \sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}_q} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}}^2 \geq 1 \\
&\quad \mathbf{A} \succeq 0
\end{aligned} \tag{4}$$

where Q stands for the number of sessions in log data.

Remark One natural question regarding the above treatment is that, although two images are judged as relevant by a user, they may still differ in many aspects. There are images that are judged differently by multiple users due to their different information needs. For example, two images could be judged both to be relevant by one user, and but only one being relevant by another user. Hence, it is questionable to treat relevant images as a similar pair. To answer this question, we need to understand that similar pairs $\mathcal{S}_q$ and dissimilar pairs $\mathcal{D}_q$ play different roles in (4). The pairs in dissimilar set $\mathcal{D}_q$ are used to form the constraint and the pairs in the similar set $\mathcal{S}_q$ are used to form the objective. Thus, a solution $\mathbf{A}$ to (4) must satisfy the constraint first before it minimizes the objective function. As a result, (4) only ensures the image pairs in $\mathcal{D}_q$ to be well separated in the feature space, but it does not guarantee that all the image pairs in $\mathcal{S}_q$ are close to each other. In other words, what is implied under the formulism in (4) is:

- When two images are judged as relevant in the same log session, they *could* be similar to each other,
- When one image is judged as relevant and another is judged as irrelevant in the same log session, they *must* be dissimilar to each other.

Clearly, the above assumption is closer to reality than the original one.

One problem with the formulism in (4) is that its solution may not be robust when the amount of log data is modest or the relevance judgments in log data are noisy. To enhance the robustness of metric learning, we form a new objective function for distance metric learning that takes into account both the discriminative issue and the robustness issue, formally as:

$$\begin{aligned}
\min_{\mathbf{A}} \|\mathbf{A}\|_{\mathrm{F}} &+ c_{\mathrm{S}} \sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}_q} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}}^2 \\
&- c_{\mathrm{D}} \sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}_q} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}}^2 \\
&\text{s. t. } \mathbf{A} \succeq 0
\end{aligned} \tag{5}$$

where $\|\mathbf{A}\|_{\mathrm{F}}$ stands for the Frobenius norm. If $\mathbf{A} = [a_{i,j}]_{m\times m}$, its Frobenius norm is defined as:

$$\|\mathbf{A}\|_{\mathrm{F}} = \sqrt{\sum_{i,j=1}^{m} a_{i,j}^2} \tag{6}$$

There are three items in (5). This item $\|\mathbf{A}\|_{\mathrm{F}}$ serves as the regularization term for matrix $\mathbf{A}$, which prevents any elements within $\mathbf{A}$ from being too large. In particular, it prefers a sparse distance metric, in which many elements of $\mathbf{A}$ are zeros or close to zeros. A similar idea has been used in support vector machines [5], in which the L2 norm of hyperplane weights is used for regularization. The second and third items in (5) represent the sum of squared distance between similar images and dissimilar images in log data. A discriminative distance metric $\mathbf{A}$ is learned such that similar images are close to each other in the space of image features and meanwhile dissimilar images are separated far away. Parameters c_{S} and c_{D} balance the tradeoff between the goal of minimizing distance among similar images and the goal of maximizing distance among dissimilar images. By adjusting these two parameters, we are also able to make a balanced tradeoff between the robustness of the learned distance metric and the discriminative power of the metric. Note that, compared to (4), the new formulism in (5) moves the image pairs in the dissimilar set to the objective function. As a result, we relax the requirement on the image pairs in $\mathcal{D}_q$: instead of assuming that all image pairs in $\mathcal{D}_q$ *must* be dissimilar to each other, we only assume that they *could* be dissimilar to each other. Through this relaxation, we are able to improve the robustness of metric learning, particularly when there are a number of errors in the log data of users' relevance judgments.

Using the distance expression in (2), both the second and the third items of objective function in (5) can be expanded into the following forms:

$$\begin{aligned}
c_{\mathrm{S}} &\sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}_q} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}}^2 \\
&= c_{\mathrm{S}} \operatorname{tr}\left(\mathbf{A} \cdot \sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}_q} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}}\right) \\
&= c_S \sum_{i,j=1}^{m} a_{i,j} s_{i,j}
\end{aligned} \tag{7}$$

and

$$\begin{aligned}
c_{\mathrm{D}} &\sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}_q} \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}}^2 \\
&= c_{\mathrm{D}} \operatorname{tr}\left(\mathbf{A} \cdot \sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}_q} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}}\right) \\
&= c_{\mathrm{D}} \sum_{i,j=1}^{m} a_{i,j} d_{i,j}
\end{aligned} \tag{8}$$

where

$$\mathbf{S} = [s_{i,j}]_{m\times m} = \sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}_q} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}}$$

$$\mathbf{D} = [d_{i,j}]_{m\times m} = \sum_{q=1}^{Q} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}_q} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}}$$

As indicated in (7) and (8), both terms are linear in matrix **A**.

Putting Eqs. (6), (7), and (8) together, we have the final formulism for the regularized metric learning:

$$\begin{aligned} &\min_{\mathbf{A}} \left(\sum_{i,j=1}^{m} a_{i,j}^2\right)^{1/2} + c_S \sum_{i,j=1}^{m} a_{i,j} s_{i,j} - c_D \sum_{i,j=1}^{m} a_{i,j} d_{i,j} \\ &\text{s. t. } \mathbf{A} \succeq 0 \end{aligned} \tag{9}$$

To convert the above problem into the standard form, we introduce a slack variable t that upper bounds the Frobenius norm of matrix **A**, which leads to an equivalent form of (9), i.e.,

$$\min_{\mathbf{A},t} \ t + c_S \sum_{i,j=1}^{m} a_{i,j} s_{i,j} - c_D \sum_{i,j=1}^{m} a_{i,j} d_{i,j} \tag{10}$$

$$\begin{aligned} \text{s. t. } &\left(\sum_{i,j=1}^{m} a_{i,j}^2\right)^{1/2} \le t \\ &\mathbf{A} \succeq 0 \end{aligned} \tag{11}$$

In the above optimization problem, the objective function is linear in both t and **A**. It has two constraints: the first constraint is called a second-order cone constraint [34], and the second constraint is a positive semidefinite constraint. Both these two types of constraints are special forms of convex constraints. They have been well studied in the optimization theory [34], and there exist very efficient solutions that guarantee to solve this problem in a polynomial time (i.e., polynomial in m^2, the square of the number of low-level image features). Note that, in the formulism in 11, we allow matrix **A** to be in *any* form as long as it is symmetric and positive definitive. In this work, an interior-point optimization method implemented in the SeDuMi [32] optimization toolbox is used to solve the optimization problem in (11).

4 Experiment methodology

4.1 Testbed

The collection of COREL image CDs contains a large number of real world images with semantic annotations. It has been widely used in previous CBIR research. In this work, two testbeds with images from 20 categories and 50 categories were created. Each category contains 100 images and is associated with specific semantic meaning such as antique, cat, dog, and lizard, etc. Given a query image from the testbed, a retrieved image is considered to be relevant when it belongs to the same category of the query image. The average precision of top retrieved images is used to measure the quality of retrieved results. Despite that such a definition of relevance judgments may not accurately reflect the characteristics of relevance judgments by real-world users, it is able to avoid the subjectiveness in manual relevance judgments. Furthermore, it automates the process of evaluation and allows different approaches to be compared based on the same ground truth. In practice, this evaluation methodology has been adopted by many studies of image retrieval, such as [10–12, 14, 15, 17, 33].

4.2 Low-level image feature representation

Low-level image feature representation is one of the key components for CBIR systems. Three types of visual features were used in this work, including color, edge, and texture. The same set of image features have been used in the previous research on image retrieval [15].

- *Color* Three types of color moments were used: color mean, color variance, and color skewness in three different color channels (i.e., H, S, and V). Thus, totally nine different features were used to represent color information.
- *Edge* Edge features have been shown to be effective in CBIR since it provides information about shapes of different objects. The histogram for edge direction was first obtained by applying the Canny edge detector [19] to images. Then, the edge direction histogram was quantized into 18 bins of every 20 degrees, which resulted in totally 18 different edge features.
- *Texture* Texture is another type of popular feature used in CBIR . In this work, we used texture features based on wavelet transformation. The Discrete Wavelet Transformation (DWT) was first applied to images with a Daubechies-4 wavelet filter [31]. Three levels of wavelet decomposition were used to obtain ten subimages in different scales and orientations. One of the subimages is a subsampled average image of the original one and was discarded as it contains less useful information. The entropies of the other nine subimages were used to represent the texture information of images.

Therefore, altogether 36 features were used in this work to represent images.

4.3 Log data of users' relevance feedback

The log data of users' relevance feedback were collected from real-world users of a CBIR system that is developed in the Chinese University of Hong Kong. Ten researchers participated in this experiment. In our experiment, for each log session, a sample query image was randomly generated.

Given the query image, the CBIR system did retrieval by computing the Euclidean distance between the query image and images in database. The top 20 most similar images were returned to users. Users provided relevance judgement for each returned image by judging if it is relevant to the query image. Each user was asked to provide ten or 15 log sessions on both the 20-category and the 50-category testbeds, respectively. All the feedback data from different log sessions were collected to build the users' log data.

An important issue for log data in real-world CBIR systems is that potentially users can make mistakes in judging the relevance of retrieved images. Thus, in reality there will be some amount of noise inside the log data of users' relevance feedback. Erroneous judgements can be caused by a variety of reasons, such as users' inconsistent and subjective judgments, and users' action mistakes. In order to evaluate the robustness of our algorithm, we collect log data with different amount of noises. The noise of log data is measured by its percentage of incorrect relevance judgments, i.e.,

$$P_{\text{noise}} = \frac{\text{Total number of wrong judgements}}{N_l \times N_{\text{log}}} \times 100\%$$

where N_l and N_{log} stand for the number of labeled examples acquired for each log session and the number of log sessions, respectively. To acquire log data with different amount of noise, we conduct experiments under two different setups. In the first setup, users' relevance judgments are collected under normal behaviors of users, which leads to relatively small numbers of mistakes. In the second setup, users are requested to provide feedback within a very short period of time, which leads to relatively higher mistakes. The reason for such a study is twofold: first, through this study, we are able to estimate the amount of noise that will be engaged in normal behaviors of real-world users; Second, the noisy log data is valuable to evaluate the robustness of our algorithms. Table 1 shows the two sets of collected log data for both datasets with different amounts of noise from real-world users. In total, 100 log sessions are collected for the 20-Category and 150 log sessions for the 50-Category dataset. Based on these log data with different configurations, we will be able to evaluate the effectiveness, the robustness, and the scalability of our algorithm for metric learning.

We would like to emphasize that the log data used in this work is created by collecting judgments from real world users. This is different from the log data of simulated users in [11], which are generated by conducting automatic retrieval for sample query images and acquiring relevance judgments based on images' category information. The log data of simulated users in [11] did not consider the data noise problem, which makes it less representative for real world applications than the data used in this work.

5 Experimental results

An extensive set of experimental results are presented in this section to illustrate the effectiveness, robustness, and scalability of our new regularized metric learning algorithm. Particularly, empirical studies were conducted to address the following three questions:

1. *How effective is our new algorithm in boosting the retrieval accuracy of a CBIR system by using the log data*? Experiments were conducted to compare the effectiveness of the distance metric learned by our new algorithm to the default Euclidean distance metric. We also compare the proposed metric learning algorithm to the algorithm in [35] for image retrieval, and to the manifold learning algorithm for CBIR that also uses log data [11].
2. *How does our new algorithm behave when the amount of users' relevance feedback is modest*? Experiments were conducted to study the effectiveness of our new algorithm by varying the size of the log data.
3. *How does our new algorithm behave when large amount of noise is present in the log data*? Experiments were conducted to study the effectiveness of our new algorithm with respect to different amount of noise.

5.1 Experiment I: effectiveness

Four algorithms are compared in this section for their accuracy of image retrieval:

1. A baseline CBIR system that uses the Euclidean distance metric and does not utilize users' log data. We refer to this algorithm as "Euclidean."
2. A CBIR system that uses the semantic representation learned from the manifold learning algorithm in [11]. We refer to this algorithm as "IML."
3. A CBIR system that uses the distance metric learned by the algorithm in [35]. We refer to this algorithm as "DML."
4. A CBIR system that uses the distance metric learned by the proposed regularized metric learning algorithm. We refer to this algorithm as "RDML."

Table 1 The characteristics of log data collected from the real-world users

	Normal log data		Noisy log data	
Datasets	# Log sessions	Noise (P_{noise}) (%)	# Log sessions	Noise (P_{noise}) (%)
20-Category	100	7.8	100	16.2
50-Category	150	7.7	150	17.1

Table 2 Average precision (%) of top-ranked images on the 20-Category testbed over 2,000 queries[a]

Top images	20	40	60	80	100
Euclidean	39.91	32.72	28.83	26.47	24.47
IML (%)	42.66 (6.9)	34.32 (4.9)	30.00 (4.1)	26.47 (0.3)	23.80 (−2.7)
DML (%)	41.45 (3.9)	34.89 (6.6)	31.21 (8.2)	28.63 (8.5)	26.44 (8.0)
RDML (%)	44.55 (11.6)	37.39 (14.3)	33.11 (14.8)	30.13 (14.1)	27.82 (13.7)

[a]The relative improvement of algorithm IML, DML, and RDML over the baseline Euclidean is included in the parenthesis following the average accuracy

Table 3 Average precision (%) of top-ranked images on the 50-Category testbed over 5,000 queries[a]

Top images	20	40	60	80	100
Euclidean	36.39	28.96	24.96	22.21	20.18
IML (%)	35.64 (−2.1)	29.16 (0.7)	24.75 (−0.8)	21.68 (−2.4)	19.32 (−4.3)
DML (%)	33.52 (−7.9)	27.15 (−6.3)	23.77 (−4.8)	21.48 (−3.3)	19.74 (−2.2)
RDML (%)	40.36 (10.9)	32.62 (12.6)	28.24 (13.1)	25.17 (13.4)	22.86 (13.3)

[a]The relative improvement of algorithm IML, DML, and RDML over the baseline Euclidean is included in the parenthesis following the average accuracy

All the algorithms were implemented with MATLAB. Specifically, for the implementation of the manifold learning algorithm for image retrieval (i.e., IML), we followed the procedure described in [11]. All the parameters in the algorithm IML were carefully tuned to achieve good retrieval accuracy. For the algorithm based on metric learning in [35] (i.e., DML), we download the code from the web site of the author[1], and slightly modified the downloaded code to fit it in the CBIR task. Finally, the proposed algorithm based on regularized metric learning (i.e., RDML) was implemented within MATLAB using the SeDuMi optimization toolbox [32] to solve the optimization problem in (11). Parameter c_S in (11) was set to 0.15 and 0.1 for the 20-Catgory and the 50-Category testbeds, respectively. Another parameter c_D was set to be one third of c_S.

The experiment in this section was conducted for the log data with small noise, i.e., 7.8% noise for the 20-Category testbed, and 7.7% noise for the 50-Category testbed. All the users' log data were used in this experiment, i.e., 100 and 150 log sessions for 20-Category and 50-Category testbeds, respectively. Every image in the database was used as a query image. The results of mean average precision for the top-ranked images are reported in Tables 2 and 3. Several observations can be drawn from Tables 2 and 3:

- Compared to the baseline model, the manifold learning method (IML) gains a small improvement for the 20-Category testbed, but it fails to improve the retrieval accuracy of CBIR for the 50-Category testbed. One possible explanation is that the IML method does not explicitly explore the Min/Max principle when it is using the log data. In particular, it is only able to exploit the images that have been judged as relevant and is unable to utilize the images judged as irrelevant. Note that the empirical results for the IML algorithm reported in this work is not consistent with the results reported in [11], where the IML method achieves a significant improvement over the Euclidean distance metric. After consulting the authors for IML, we believe that the inconsistency could be attributed to different characteristics of log data used in these two studies. Not only was a much larger amount of users' log data used in [11] than in this work, but also their log data did not include any noise. To further confirm the correctness of this explanation, we followed the same procedure described in [11] and constructed similar log data of simulated users. We tested our implementation of the IML algorithm using the simulated log data and observed a similar amount of improvement as reported in [11]. Based on these results, we have confirmation that the IML algorithm works well when a large amount of log data is available. It may fail to improve the performance of CBIR when the size of log data is small.
- The distance metric learning (DML) algorithm does achieve certain amount of improvement over the baseline algorithm on the 20-Category testbed. But it performs consistently worse than the Euclidean distance on the 50-Category testbed. These results indicate that distance metric learned by the DML algorithm may not be robust and can suffer from the overfitting problem. This is because images from the 50-Category testbed are much more diverse than images from the 20-Category testbed. In contrast, the size of log data for the 50-Category testbed is only slightly larger than that for the 20-Category testbed. Thus, log data may not be sufficient for representing the diversity of the 50-Category testbed, which leads the DML algorithm to over-fit log data and therefore degrades the retrieval accuracy.
- Compared to the baseline method, the proposed algorithm for regularized distance metric learning (RDML) is able to consistently achieve more than 10% improvement in mean average precision for the top-ranked images. These results indicate that the RDML algorithm is more robust than the other two algorithms in boosting the retrieval accuracy of CBIR with log data. We attribute the success of the RDML algorithm to the combination of the discriminative training, which is based on

[1] http://www-2.cs.cmu.edu/~epxing/publication.html

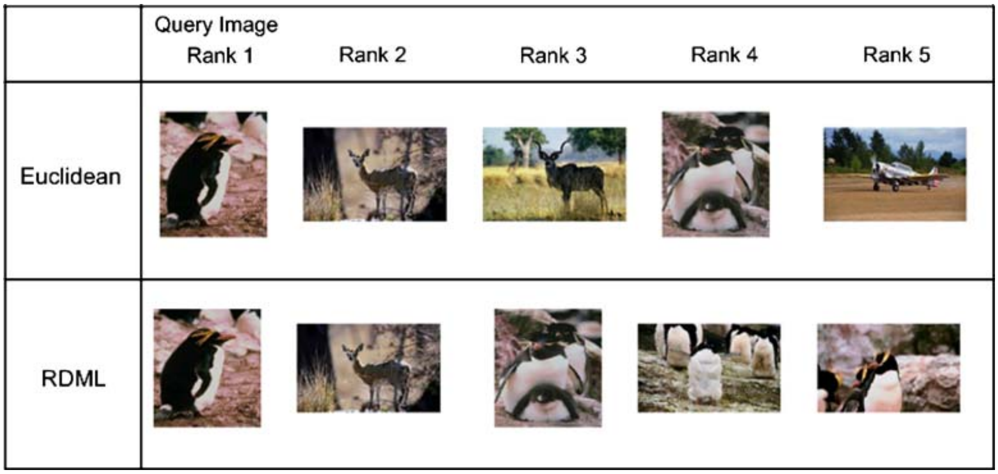

Fig. 1 The retrieval results of top-5 returned images of a sample query image (the first one in the next two rows) for CBIR systems with either the Euclidean distance metric (*first row*) or the distance metric learned by RDML (*second row*)

the Min/Max principle, and the regularization procedure, which results in more robust distance metric.

To further illustrate the behavior of the RDML algorithm, we list the retrieval results of a sample query image in Fig. 1. The first row of Fig. 1 shows the top-5 returned images from the CBIR system using Euclidean distance metric, while the second row represents the results by the CBIR system using the distance metric learned by the RDML algorithm. The first image of each row is the sample query image. It can be seen that the CBIR system using Euclidean metric only acquired two relevant images (including the query image) out of top five returned images, while the CBIR system using the RDML algorithm did a better work by retrieving two more relevant images (the fourth one and the fifth image on the second row).

5.2 Experiment II: efficiency and scalability

In addition to being more effective than the IML and the DML algorithm, the RDML algorithm can also be computed substantially more efficiently than the other two algorithms and is scalable to the size of log data. To manifest the efficiency and scalability of the proposed algorithm, we conducted a set of experiments to show the training time of these three algorithms. All the algorithms were run on a Windows XP operation system that is powered by a 2.0 GHz PC with 1 GB physical memory. The training times of these three algorithms are shown in Table 4. As indicated in Table 4, the RDML algorithm can be trained much more efficiently than the other two algorithms for both testbeds. Particularly, two observations can be drawn from Table 4:

Table 4 The training time cost (CPU seconds) of three algorithms on 20-Category (100 log sessions) and 50-Category (150 log sessions) testbeds

Algorithm	IML	DML	RDML
20-Category	82.5	3,227	19.2
50-Category	2,864	12,341	20.5

- The RDML algorithm is significantly more efficient than the DML algorithm. For both datasets, the training cost of the DML algorithm is at least two orders larger than that of the RDML algorithm. Note that both algorithms try to learn the distance metric **A** from the same log data and therefore have the same problem size. The RDML algorithm is more efficient than the DML algorithm because its related optimization problem can be solved efficiently by the semidefinite programming technique, while the DML algorithm has to solve a general convex programming problem that is usually much more time-consuming.
- The RDML algorithm is significantly more scalable to the size of log data than the IML algorithm. For the 20-Category testbed, both the IML algorithm and the RDML algorithm have similar training cost. However, for the 50-Category testbed, the training cost for the IML algorithm shoots up to about 3,000 s. Whereas the RDML algorithm is able to maintain its training cost almost unchanged between the 20-Category and the 50-Category. This is because the IML algorithm needs to solve a generalized eigenvalue decomposition problem [11], in which the problem size is not only dependent on the number of image features, but also dependent on the number of images in log data. Given the computational complexity of principal eigenvectors is of the order of n^3 where n is the number of variables, the IML algorithm cannot scale up to the size of log data. In contrast, the problem size for the RDML algorithm, only depends on the number of image features, thus is the same for both testbeds. As a result, regardless of the size of log data, the problem sizes of the RDML algorithm are the same, which leads to unchanged training cost.

Table 5 Average precision (%) of top-ranked images on the 20-Category testbed for IML, DML, and RMDL algorithm using small amounts of log data[a]

Top images	20	40	60	80	100
Euclidean	39.91	32.72	28.83	26.47	24.47
IML (#Log 67) (%)	39.01 (−2.3)	31.49 (−3.8)	27.64 (−4.1)	24.75 (−6.5)	22.43 (−8.3)
DML (#Log 67) (%)	41.03 (2.8)	34.73 (6.1)	31.26 (8.4)	28.67 (8.3)	26.47 (8.2)
RDML (#Log 67) (%)	43.80 (9.7)	36.15 (10.5)	32.00 (11.0)	29.20 (10.6)	26.89 (9.9)
IML (#Log 33) (%)	36.64 (−8.2)	29.72 (−9.2)	25.99 (−9.9)	23.41 (−11.6)	21.53 (−12.0)
DML (#Log 33) (%)	38.13 (−4.5)	31.99 (−2.2)	28.69 (−0.5)	26.34 (−0.5)	24.50 (−0.1)
RDML (#Log 33) (%)	42.56 (6.6)	35.12 (7.3)	31.01 (7.5)	28.17 (6.7)	26.11 (6.7)

[a]The relative improvement of algorithm over the baseline Euclidean is included in the parenthesis following the average accuracy

Table 6 Average precision (%) of top-ranked images on the 50-Category testbed for IML, DML, and RMDL using small amounts of log data[a]

Top images	20	40	60	80	100
Euclidean	36.39	28.96	24.96	22.21	20.18
IML (#Log 100) (%)	34.25 (−5.8)	27.65 (−4.5)	23.34 (−6.5)	20.69 (−6.8)	18.49 (−8.4)
DML (#Log 100) (%)	33.53 (−7.9)	26.84 (−7.3)	23.28 (−6.7)	20.93 (−5.8)	19.21 (−4.8)
RDML (#Log 100) (%)	39.10 (7.4)	31.62 (9.2)	27.28 (9.3)	24.30 (9.4)	22.02 (9.2)
IML (#Log 50) (%)	32.95 (−9.5)	26.87 (−7.2)	22.92 (−8.2)	20.35 (−8.4)	18.25 (−9.6)
DML (#Log 50) (%)	29.78 (−18.2)	23.26 (−19.7)	19.86 (−20.4)	17.70 (−20.3)	16.13 (−20.1)
RDML (#Log 50) (%)	38.96 (7.1)	31.44 (8.6)	27.08 (8.5)	24.09 (8.5)	21.76 (7.9)

[a]The relative improvement over the baseline Euclidean is included in the parenthesis following the average accuracy

5.3 Experiment III: different size of log data

In real-world CBIR applications, it may be difficult to acquire large amount of users' log data. This issue is especially important in the early stage of system development. It is also important when the target images are not popular and are only equipped with a few relevance judgments. In this case, the CBIR system has to provide retrieval service with limited amount of log data. A set of experiments was designed in this section to show the behavior of the RMDL algorithm together with the IML and the DML algorithm in response to different size of log data. Different from the experiments presented in the previous sections, where all users' log data are used, in this section, all the algorithms were trained with only part of users' log data. In particular, it was trained with one-third and two-third users' log data for both testbeds. The empirical results are shown in Tables 5 and 6.

It can be seen from these two tables that the advantage of the RMDL algorithm over the baseline algorithm using Euclidean distance metric decreases with less training data. However, even with very limited amount of training data, i.e., 33 log sessions for 20-Category and 50 log sessions for 50-Category, the RDML algorithm is still capable to gain notable improvement over the baseline model, which is about 7% for 20-Category and about 8% for 50-Category. Compared to the RDML algorithm, the IML algorithm and the DML algorithm suffer from substantially more degradation in the retrieval accuracy. In fact, for most cases when a small amount of log data is present, both the IML algorithm and the DML algorithm perform even worse than the straightforward Euclidean distance. In sum, this set of experiments demonstrates the robustness of the RDML algorithm in improving content-based image retrieval with the limited amount of users' log data, which can be important for real-world CBIR systems.

5.4 Experiment IV: noisy log data

Another practical problem with real-world CBIR applications is that the log data of user feedback are inevitable to contain certain amount of noise. The experimental results in previous sections have demonstrated that the RMDL algorithm is able to boost the retrieval results of a CBIR system when log data have only a small amount of noise. It is interesting to investigate the behavior of the RMDL algorithm when more noise is present in the log data of users' relevance feedback.

Experiments were conducted on both the 20-Category and the 50-Category testbeds using the log data that contain a large amount of noise. The details of users' log data with large noise have been described in Sect. 4.3. The experiment results for two testbeds using the RMDL algorithm are shown in Tables 7 and 8, respectively. It can be seen from the experiment results that the noise in users' log data does have a significant impact on the retrieval accuracy, which is consistent with our expectation. However, even when the noisy log data that contain over 15% incorrect relevance judgments, the RMDL algorithm still shows a consistent improvement over the baseline method using the Euclidean distance metric, although the improvement is small. In contrast, both the IML algorithm and the DML algorithm fail to improve the performance over the Euclidean distance when the log data is noisy. These results indicate the robustness of our new algorithm, which again is important for real-world CBIR applications.

Table 7 Average precision (%) of top-ranked images on the 20-Category testbed for IML, DML, and RMDL using noisy log data[a]

Top images	20	40	60	80	100
Euclidean	39.91	32.72	28.83	26.47	24.47
IML (large noise) (%)	37.94 (−4.9)	30.14 (−7.9)	25.93 (−10.1)	23.56 (−11.0)	21.97 (−10.2)
DML (large noise) (%)	38.62 (−3.2)	32.32 (−1.2)	28.95 (0.4)	26.61 (0.8)	24.62 (0.6)
RDML (large noise) (%)	41.19 (3.2)	34.15 (4.4)	30.40 (5.4)	27.92 (5.8)	25.89 (5.8)

[a]The relative improvement over the baseline Euclidean is included in the parenthesis following the average accuracy

Table 8 Average precision (%) of top-ranked images on the 50-Category testbed for IML, DML, and RMDL using noisy log data[a]

Top images	20	40	60	80	100
Euclidean	36.39	28.96	24.96	22.21	20.18
IML (large noise) (%)	33.80 (−7.1)	27.30 (−5.8)	23.56 (−5.0)	20.65 (−6.7)	18.36 (−8.1)
DML (large noise) (%)	32.85 (−9.7)	26.95 (−7.0)	23.55 (−5.7)	21.22 (−4.5)	19.49 (−3.4)
RDML (large noise) (%)	37.45 (2.9)	29.97 (3.5)	25.84 (3.5)	22.99 (3.5)	20.87 (3.4)

[a]The relative improvement over the baseline Euclidean is included in the parenthesis following the average accuracy

6 Limitation and future work

Based on the promising results achieved from the above extensive empirical evaluations, we conclude that the regularized metric learning algorithm is effective for improving the performance of CBIR systems by integrating the log data of users' relevance feedback. Through the regularization mechanism, the learned distance metric is more robust. By formulating the learning problem into a semidefinite programming problem, it can be solved efficiently and is scalable to the size of log data. However, it is necessary to address the limitation and the challenging issues with the proposed algorithm as well as feasible directions for solving these problems in our future work.

First, we realize that the selection of parameter c_S and c_D in the proposed algorithm is important to its retrieval performance. Although our empirical approach for choosing c_S and c_D has resulted in good performance, we plan to investigate other principled approaches for effectively tuning these two parameters. One potential approach is to automatically determine these two parameters using the cross-validation method. It divides the log data into 20%/80% partitions where 80% of the data is used for training and 20% for validation. The optimal values of c_S and c_D are found by maximizing the retrieval accuracy of the validation set.

Second, although our algorithm is robust to the noise present in the log data, the degradation in the retrieval accuracy caused by erroneous judgments is still quite significant. Hence, in the future, we plan to consider more sophisticated regularization approaches for metric learning, such as manifold regularization [2].

Third, in the proposed algorithm, a single distance metric is learned to describe the similarity between *any* two images. Given a heterogeneous collection that consists of multiple different types of images, a single distance metric may not be sufficient to account for diverse types of similarity functions. In the future, some interesting extensions can be naturally derived from our work. One possible way is to learn multiple query-dependent distance metrics with respect to different query types, which is similar to the idea of *query classification based retrieval* [8] in document information retrieval. Moreover, we may also learn multiple user-dependent distance metrics if users' preferences are available.

7 Conclusions

Content-based image retrieval (CBIR) has been an active research topic for many years. However, its retrieval accuracy is still not satisfactory due to the semantic gap between low-level image feature representation and high-level image semantic meaning. This paper proposes a novel algorithm for distance metric learning, which boosts the retrieval accuracy of CBIR by taking advantage of the log data of users' relevance judgments. A regularization mechanism is used in the proposed algorithm to improve the robustness of solutions, when the log data is small and noisy. Meanwhile, it is formulated as a positive semidefinite programming problem, which can be solved efficiently and therefore is scalable to the size of log data.

Experiment results have shown that the proposed algorithm for regularized distance metric learning substantially improves the retrieval accuracy of the baseline CBIR system that uses the Euclidean distance metric. It is also more effective and more efficient than two alternative algorithms that also utilize the log data to enhance image retrieval. More empirical studies indicate that the new algorithm gains notable improvement even with limited amount of users' log data. Furthermore, the new algorithm is rather robust to work in the environment where the log data is noisy and contains a number of erroneous judgments.

In sum, the new algorithm for regularized distance metric learning has a nice theoretical formalization and generates better empirical results than several other approaches. It can be computed efficiently with large-scale CBIR system and also works well in CBIR systems when users' log data are noisy and insufficient. All these advantages make the new algorithm proposed in this paper a good candidate for combining the log data and the low-level image features to improve the retrieval performance of CBIR systems.

Acknowledgements We would like to thank Eric Xing and Xiaofei He for the helpful discussion and clarification for their work as well as valuable comments and suggestions from anonymous reviewers. The last two authors were supported by two grants, one from the Shun Hing Institute of Advanced Engineering, and the other from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK4205/04E).

References

1. Ashwin, T.V., Navendu, J., Ghosal, S.: Improving image retrieval performance with negative relevance feedback. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP'01), UT (2001)
2. Belkin, M., Niyog, P., Sindhwani, V.: Manifold regularization: a geometric framework for learning from examples. Technical Report, Computer Science Technical Report TR-2004-06, University of Chicago (2004)
3. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Advances in Neural Information Processing Systems, vol. 14 (2002)
4. Blei, D., Jordan, M.I.: Modeling annotated data. In: Proceedings of the 26th International ACM SIGIR Conference (SIGIR'03), pp. 127–134 (2003)
5. Burges, C.J.C.: A tutorial on support vector machine for pattern recognition. Knowl. Discov. Data Min. **2**(2), 121–167 (1998)
6. Cox, I.J., Miller, M., Minka, T., Yianilos, P.: An optimized interaction strategy for Bayesian relevance feedback. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98), pp. 553–558. Santa Barbara, CA (1998)
7. Duygulu, P., Barnard, K., de Freitas, J., Forsyth, D.: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: Proceedings of the 7th European Conference on Computer Vision, pp. 97–112 (2002)
8. Gill, P.E., Murray, W., Wright, M.H.: Practical Optimization. Academic, London (1981)
9. Gong, Y., Chua Zhang, H.Z., Sakauchi, H.C.M.: An image database system with content capturing and fast image indexing abilities. In: IEEE International Conference on Multimedia Computing and Systems (1994)
10. He, J., Li, M., Zhang, H.J., Tong, H., Zhang, C.: Manifold ranking based image retrieval. In: Proceedings of ACM Multimedia 2004 (2004)
11. He, X., Ma, W.-Y., Zhang, H.-J.: Learning an image manifold for retrieval. In: Proceedings of ACM MM 2004 (2004)
12. He, X., King, O., Ma, W.-Y., Li, M., Zhang, H.J.: Learning a semantic space from user's relevance feedback for image retrieval. IEEE Trans. Circ. Syst. Video Technol. **13**(1), 39–48 (2003)
13. Heesch, D., Yavlinsky, A., Rüuger, S.: Performance comparison between different similarity models for cbir with relevance feedback. In: Proceedings of International Conference on Image and Video Retrieval (CIVR'03), LNCS 2728, pp. 456–466. Springer, Berlin Heidelberg New York (2003)
14. Hoi, C.H., Lyu, M.R.: Group-based relevance feeedback with support vector machine ensembles. In: Proceedings 17th International Conference on Pattern Recognition (ICPR'04). Cambridge, UK (2004)
15. Hoi, C.-H., Lyu, M.R.: A novel log-based relevance feedback technique in content-based image retrieval. In: Proceedings of ACM Multimedia 2004 (2004)
16. Hoi, C.H., Lyu, M.R., Jin, R.: Integrating user feedback log into relevance feedback by coupled svm for content-based image retrieval. In: Proceedings of the 1st IEEE International Workshop on Managing Data for Emerging Multimedia Applications (EMMA 2005) (2005)
17. Hoi, C.-H., Lyu, M.R.: Web image learning for searching semantic concepts in image databases. In: Poster Proceedings of the 13th International World Wide Web Conference (WWW'2004), New York (2004)
18. Huang, T.S., Zhou, X.S.: Image retrieval by relevance feedback: from heuristic weight adjustment to optimal learning methods. In: Proceedings of IEEE International Conference on Image Processing (ICIP'01), Thessaloniki, Greece (2001)
19. Jain, A.K., Vailaya, A.: Shape-based retrieval: a case study with trademark image database. Pattern Recognit. **9**, 1369–1390 (1998)
20. Jeon, J., Lavrenko, V., Manmatha, R.: Automatic image annotation and retrieval using cross-media relevance models. In: Proceedings of the 26th International ACM SIGIR Conference (SIGIR'03), pp. 119–126 (2003)
21. King, I., Zhong, J.: Integrated probability function and its application to content-based image retrieval by relevance feedback. Pattern Recognit. **36**(9), 2177–2186 (2003)
22. Lavrenko, V., Manmatha, R., Jeon, J.: A model for learning the semantics of pictures. In: Advances in Neural Information Processing Systems (NIPS'03) (2003)
23. Lebanon, G.: Learning Riemannian metrics. In: Proceedings of the 19th Conference on Uncertainty in Articial Intelligence (2003)
24. Lu, Y., Hu, C., Zhu, X., Zhang, H.J., Yang, Q.: A unified framework for semantics and feature based relevance feedback in image retrieval systems. In: MULTIMEDIA '00: Proceedings of the Eighth ACM International Conference on Multimedia, pp. 31–37. ACM, New York (2000)
25. Muller, H., Pun, T., Squire, D.: Learning from user behavior in image retrieval: Application of market basket analysis. Int. J. Comput. Vis. **56**(1–2), 65–77 (2004)
26. Porkaew, K., Chakrabarti, K., Mehrotra, S.: Query refinement for multimedia retrieval and its evaluation techniques in mars. In: Proceedings of ACM Multimedia (MM'99). Orlando, FL (1999)
27. Rui, Y., Huang, T.S.: A novel relevance feedback technique in image retrieval. In: Proceedings of the ACM Multimedia (MM'99) pp. 67–70. Orlando, FL (1999)
28. Rui, Y., Huang, T.S., Ortega, M., Mehrotra, S.: Relevance feedback: a power tool in interactive content-based image retrieval. IEEE Trans. Circ. Syst. Video Technol. **8**(5), 644–655 (1998)
29. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Proc. Manag.: Int. J. **24**(5), 513–523 (1988)
30. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
31. Smith, J., Chang, S.-F.: Automated image retrieval using color and texture. IEEE Trans. Pattern Anal. Mach. Intell. (1996)
32. Sturm, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optimization Methods and Software, pp. 11–12, 625–653. Special issue on Interior Point Methods (CD supplement with software) (1999)
33. Tong, S., Chang, E.: Support vector machine active learning for image retrieval. In: Proceedings of The Ninth ACM International Conference on Multimedia, pp. 107–118. ACM, New York (2001)
34. Vandenberghe, L., Boyd, S.: Semidefinite programming. SIAM Rev. **38**(1), 49–95 (1996)
35. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning with application to clustering with side-information. In: Thrun, S., Becker, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems, vol. 15, pp. 505–512. MIT Press, Cambridge, MA (2003)
36. Zhou, X.-D., Zhang, L., Liu, L., Zhang, Q., Shi,B.-L.: A relevance feedback method in image retrieval by analyzing feedback log file. In: Proceedings of the International Conference on Machine Learning and Cybernetics, vol. 3, pp. 1641–1646. Beijing (2002)