

Generalized Manifold-Ranking-Based Image Retrieval

Jingrui He, Mingjing Li, Hong-Jiang Zhang, Hanghang Tong, and Changshui Zhang

Abstract—In this paper, we propose a general transductive learning framework named generalized manifold-ranking-based image retrieval (gMRBIR) for image retrieval. Comparing with an existing transductive learning method named MRBIR [12], our method could work well whether or not the query image is in the database; thus, it is more applicable for real applications. Given a query image, gMRBIR first initializes a pseudo seed vector based on neighborhood relationship and then spread its scores via manifold ranking to all the unlabeled images in the database. Furthermore, in gMRBIR, we also make use of relevance feedback and active learning to refine the retrieval result so that it converges to the query concept as fast as possible. Systematic experiments on a general-purpose image database consisting of 5 000 Corel images demonstrate the superiority of gMRBIR over state-of-the-art techniques.

Index Terms—Image retrieval, manifold ranking, outside the database, relevance feedback.

I. INTRODUCTION

THE history of image retrieval can be traced back to the late 1970s, which aims to provide an effective and efficient tool for managing large image databases. Up until now, with the advent and popularity of World Wide Web, the number of digital images available for various purposes has grown tremendously, and many researchers have put their attention to the development of an image retrieval system, which works well in general or specific contexts [4], [14], [27], [30].

In the preliminary stage, image retrieval is based on keyword annotation, which is a natural extension of text retrieval. In this approach, images in the database are first annotated manually by keywords, and then retrieved according to their annotations. However, it suffers from several main difficulties, e.g., the large amount of labor required to annotate the whole database, and the inconsistency among different annotators in perceiving the same image. Therefore, this approach cannot be applied to real applications, especially when the size of the database is very large.

To overcome these difficulties, an alternative scheme, content-based image retrieval (CBIR) was proposed in the early 1990s, which makes use of low-level image features instead of keywords to represent images. Its advantage over keyword-

based image retrieval lies in the fact that low-level features can be extracted automatically without human intervention, and that the image's own content is always consistent. Present low-level features can be categorized into color [10], [13], [22], texture [3], [9], [11], [15], [23], shape [5], [26], etc. However, despite the great deal of research work dedicated to the exploration of an ideal descriptor for image content, the performance of these low-level features is far from satisfactory due to the well-known gap between visual features and semantic concepts, i.e., images of dissimilar semantic content may share some common low-level features, while images of similar semantic content may be scattered in the feature space.

To narrow or bridge the gap, a great deal of research work has been performed, which can be categorized into two major groups: one is to search for appropriate metrics to measure perceptual similarity; and the other is to incorporate relevance feedback (RF) into the retrieval process in order to learn better representation of images as well as the query concept.

In the initial retrieval stage, given the query image, several distance functions can be used to measure the similarity between the query and all the images in the database. Traditional distance functions are based on Minkowski metrics, such as L_1 , L_2 , L_∞ , etc. In [20], the authors compare the performance of different Minkowski metrics in texture image retrieval and draw a conclusion that Manhattan (L_1) distance performs better than Euclidean (L_2), Mahalanobis, and Chebychev (L_∞) distances. This conclusion is consistent with the experimental results of [10], [21], where L_1 distance outperforms other distances on color images. A more systematic study is presented in [25], where the authors propose a model of image retrieval systems and deduce a scheme for deriving the best similarity measure in a set of similarity measures, assuming a parametric model of the variability of feature vectors within the same class. More recently, to make up for the drawback of static feature weighting schemes combined with Minkowski metrics, Li *et al.* [2] propose a perceptual distance function (DPF), which is dynamically calculated in the subspace where the similarity between two images is maximized. Another example is the Earth Mover's Distance (EMD) [28], which has a rigorous probabilistic interpretation and has been successfully applied to image retrieval [8]. However, these metrics are based on pairwise distance calculation and oversimplify the relationship among all the images in the database. Therefore, their effectiveness is quite limited.

On the other hand, relevance feedback is an online learning technique used to improve the performance of information retrieval systems. With the additional information of the user's rating on the relevance of the retrieved images, the system dynamically learns the user's query concept, and gradually improves the retrieval result. Among others, a key issue in relevance feedback is the learning strategy. Traditional learning

Manuscript received January 9, 2005; revised February 12, 2006. This work was supported by National High Technology Research and Development Program of China (863 Program) under Contract 2001AA114190. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Benoit Macq.

J. He, H. Tong, and C. Zhang are with the Automation Department, Tsinghua University, Beijing 100084, China (e-mail: hejingrui98@mails.tsinghua.edu.cn; walkstar98@mails.tsinghua.edu.cn; zcs@tsinghua.edu.cn).

M. Li and H.-J. Zhang are with Microsoft Research Asia, Beijing 100080, China (e-mail: mjli@microsoft.com; hjzhang@microsoft.com).

Digital Object Identifier 10.1109/TIP.2006.877491

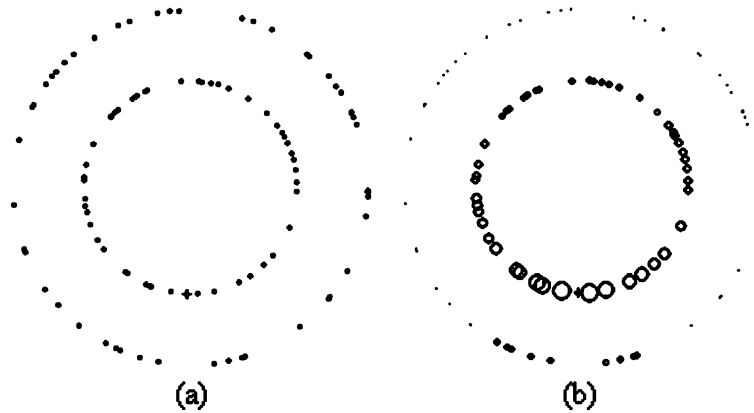


Fig. 1. Manifold ranking on a simple toy problem.

methods can be categorized into three major groups [16], [17], [24]: query reweighting [16], [27], [29], query point movement [17], [30], and query expansion [16], [17]. However, because these methods do not fully utilize the information embedded in feedback images, their performance can not reach a satisfactory level.

More recently, statistical learning methods have been incorporated into relevance feedback, and have been extensively demonstrated to outperform their traditional counterparts [1], [18], [19], [24], [31]. According to whether or not unlabeled data is utilized in the training stage, these methods can be classified into inductive and transductive ones.

The goal of an inductive method is to create a classifier which separates the relevant and irrelevant images and generalizes well on unseen examples. For example, the authors of [18] first compute a large number of highly selective features, and then use boosting to learn a classification function in this feature space; similarly, the relevance feedback method proposed in [19] trains a support vector machine (SVM) from labeled examples, hoping to obtain a small generalization error by maximizing the margin between relevant and irrelevant images. Moreover, to speed up the convergence to the target concept, active learning methods are also utilized to select the most informative images which will be presented to and marked by the user. For example, the support vector machine active learning algorithm (SVM_{active}) proposed by Tong *et al.* [24] selects the points near the SVM boundary so as to maximally shrink the size of the version space. Another active learning scheme, the maximizing expected generalization algorithm (MEGA) [1], judiciously selects samples in each round and uses positive examples to learn the target concept, while negative examples to bound the uncertain region. One major problem with inductive methods is the insufficiency of labeled examples, which might bring great degradation to the performance of the trained classifier.

Different from inductive methods, transductive methods aim to accurately predict the relevance of unlabeled images which are attainable during the training stage. For example, Discriminant-EM algorithm proposed by Wu *et al.* [31] makes use of unlabeled data to construct a generative model, which will be used to measure relevance between the query and database images. However, as pointed out in [31], if the components of data distribution are mixed up, which is often the case in CBIR, the performance of D-EM will be compromised.

In [12], we have proposed a novel transductive learning framework named manifold-ranking-based image retrieval (MRBIR). Different from traditional CBIR systems, in MRBIR, relevance between the query and database images is evaluated by exploring the relationship of all the data points in the feature space, which addresses the limitation of present similarity metrics based on pairwise distance. Comparing with D-EM, which uses unlabeled data to construct a generative model, MRBIR takes each unlabeled image as a vertex in a weighted graph that will propagate the ranking score of labeled examples. Furthermore, the proposed system can improve the retrieval result by means of relevance feedback, including feedback with only positive examples and with both positive and negative examples. More importantly, for the first time, MRBIR integrates transductive learning and active learning in a natural way, which results in a better performance than state-of-the-art techniques.

One major problem with MRBIR is that it can only deal with the problem where the query image is in the database, since the weighted graph takes the query point as a vertex. However, when the query image is not in the database, MRBIR cannot spread its ranking score to the images in the database. In most real applications, the query image is provided by the user, and it is unlikely to be found in the database. Thus, the usefulness of MRBIR is somewhat limited.

In this paper, we extend MRBIR to form a general framework named generalized manifold-ranking-based image retrieval (gMRBIR), which could work well no matter whether or not the query image is in the database. Its basic concern is how to effectively initialize the ranking scores of the vertices in the original graph. Given a query image, gMRBIR performs the following two-step procedure.

- 1) Initialization: Spread the ranking score of the query image to its K nearest neighbors in the database.
- 2) Propagation: Spread the ranking scores of the neighbors to all the unlabeled images via manifold ranking.

If the query image is in the database, MRBIR and gMRBIR lead to the same ranking result when $K = 1$; while, if the query image is not in the database, the first step of gMRBIR provides K seeds with various ranking scores, and the second step performs manifold ranking based on these seeds.

Taking advantage of the experimental results in [12], we incorporate the best schemes for relevance feedback and active learning used in MRBIR into gMRBIR in order to refine the retrieval results. Furthermore, some modifications are made to accommodate the speciality of gMRBIR.

The organization of the paper is as follows. In Section II, we briefly review related work. Section III presents the fundamental idea of gMRBIR. Learning methods for improving the retrieval result, including relevance feedback and active learning, are introduced in Section IV. In Section V, we provide experimental results to evaluate the performance of gMRBIR. Finally, we conclude in Section VI.

II. RELATED WORK

Different from traditional methods, which measure perceptual similarity based on pairwise distance, MRBIR makes use of a manifold ranking algorithm to measure relevance between the query and database images, which explores the relationship of all the data points in the feature space [12]. Fig. 1 presents a simple toy problem to illustrate the idea of manifold ranking. We are given a set of points sampled from two concentric circles, and a query in the inner circle [Fig. 1(a)]. The task is to rank all the data points according to their relevance to the query. Intuitively, the relevance of data points in the inner circle should decrease along the circle, and so do the data points in the outer circle; furthermore, all the points in the inner circle should have a higher relevance than those in the outer circle, as is shown in Fig. 1(b). However, if the points are ranked simply according to pairwise Euclidean distance to the query, we will by no means obtain such a result.

Given a set of points $\chi = \{x_1, \dots, x_q, x_{q+1}, \dots, x_n\} \subset \mathfrak{R}^m$, the first q points are the queries which form the query set, and the rest are to be ranked according to their relevance to the queries. Let $d: \chi \times \chi \rightarrow \mathfrak{R}$ denote a metric on χ which assigns to each pair of points x_i and x_j a distance $d(x_i, x_j)$, and $f: \chi \rightarrow \mathfrak{R}$ denote a ranking function which assigns each point x_i a ranking score f_i to form the vector \vec{f} . Finally, we define a vector $y = [y_1, \dots, y_n]^T$, in which $y_i = 1$ if x_i is a query, and $y_i = 0$, otherwise. Furthermore, we need to set two parameters: the number K_0 of nearest neighbors between which an edge should be put, and the width σ of the kernel for defining the edge weight.

The procedure of the manifold ranking algorithm used in MRBIR is presented below. An intuitive description of this algorithm is: a weighted graph is first formed which takes each data point as a vertex; assign a positive ranking score to each query while zero to the remaining points; all the data points then spread their scores to the nearby points via the weighted graph; the spread process is repeated until a global stable state is reached, and all the points except the query will have their own scores according to which they will be ranked. In the context of image retrieval, there is only one query in the query set. The resultant ranking score of an unlabeled image is in proportion to the probability that it is relevant to the query, with large ranking score indicating high probability.

Manifold Ranking Algorithm

- 1) Calculate the K_0 nearest neighbors for each point; connect two points with an edge if they are neighbors.
- 2) Form the affinity matrix W defined by $W_{ij} = \exp[-d^2(x_i, x_j)/2\sigma^2]$ if there is an edge linking x_i and x_j . Let $W_{ii} = 0$.
- 3) Symmetrically normalize W by $S = D^{-1/2}WD^{-1/2}$ in which D is the diagonal matrix with (i, i) -element equal to the sum of the i th row of W .
- 4) Let $\vec{f}(0)$ be a zero vector. Iterate $\vec{f}(t+1) = \alpha S \vec{f}(t) + (1-\alpha)y$ until convergence, where α is a parameter in $[0, 1)$.
- 5) Let f_i^* denote the i th component of the limit of the sequence $\{\vec{f}(t)\}$. Rank each point x_i according to its ranking scores f_i^* (largest ranked first).

By careful analysis, we have the following theorem [6], [7]: The sequence $\{\vec{f}(t)\}$ converges to¹

$$\vec{f}^* = (I - \alpha S)^{-1}y. \quad (1)$$

From (1), we can reach the following conclusions.

- 1) In the context of image retrieval, the matrix $A = (I - \alpha S)^{-1}$ can be calculated offline to facilitate fast online calculation of the ranking score \vec{f}^* since no iteration steps are needed.
- 2) The ranking scores f_i^* of all the unlabeled images are obtained from y . Therefore, we name y as the “seed vector” hereafter.

III. gMRBIR

A. Fundamental Idea

Recall that in MRBIR, the query image should be a vertex in the weighted graph; otherwise y would be a zero vector, and all the unlabeled images would end up with a zero ranking score. This condition is satisfied when the query image is in the database. However, in real applications, users tend to select a query image from outside the database. In these cases, MRBIR cannot be readily used to perform image retrieval. A natural solution is to form an enlarged graph which adds the query image as a vertex into the original graph. However, one major problem with this solution is that: each time the user provides a query image, the enlarged matrix $(I - \alpha S)^{-1}$ has to be calculated online, which will inevitably slower the processing speed of image retrieval systems.

Next, we will present our method used in gMRBIR for dealing with the problem of query images outside the database. The fundamental idea is how to properly initialize the ranking scores of the vertices in the original graph. Suppose that the affinity matrix of the enlarged graph with the query image as a vertex is denoted W_e , then we have

$$W_e = \begin{bmatrix} W & e \\ e^t & 0 \end{bmatrix} \quad (2)$$

¹We have omitted a constant coefficient which will not affect the ranking result.

where e is an n -dimensional vector. Its i th component $e(i) \neq 0$ if the i th image is among the K nearest neighbors of the query image, and its value is defined the same way as W_{ij} in step 2) of the manifold ranking algorithm. We will discuss about the value of K in Section III-B. After normalization [as in step 3) of the manifold ranking algorithm], we have

$$\begin{aligned} S_e &= D_e^{-1/2} W_e D_e^{-1/2} \\ &\approx \begin{bmatrix} D^{-1/2} & 0 \\ 0 & (e^t \cdot \vec{1})^{-1/2} \end{bmatrix} \\ &\quad \times \begin{bmatrix} W & e \\ e^t & o \end{bmatrix} \begin{bmatrix} D^{-1/2} & 0 \\ 0 & (e^t \cdot \vec{1})^{-1/2} \end{bmatrix} \\ &= \begin{bmatrix} S & s \\ s^t & 0 \end{bmatrix} \end{aligned} \quad (3)$$

where D_e is a diagonal matrix with (i, i) -element equal to the sum of the i th row of W_e , and s is the normalized vector of e , i.e.,

$$s(i) = \frac{e(i)}{(D(i, i) \cdot D_e(n+1, n+1))^{(1/2)}} \quad (4)$$

where $D(i, i)$ is the (i, i) -element of the diagonal matrix D , and $D_e(n+1, n+1) = \sum_{i=1}^n e(i)$. In (3), the approximation is because that $e(i)$ can be ignored compared to $D(i, i)$.

The i th component $s(i)$ of s can be seen as the similarity between the query and database images based on low-level features, if the database image is within the K nearest neighbors of the query. It is easily verified that $0 < s(i) < 1$ since $e(i) \leq D(i, i)$ and $e(i) \leq D_e(n+1, n+1)$. Thus, we make reasonable assumptions that $s^t \cdot s \approx 0$, then $s \cdot s^t \approx Z$, and the Frobenius norm $\|s \cdot s^t\|_F \ll \|S\|_F$.

Define $y_e = [y^t, 1]^t$ ($n+1$ -dimensional) as the seed vector of the enlarged graph. Recall that when the query image is not in the database, the seed vector y of the original graph is a zero vector. Let \vec{f}_e^* ($n+1$ -dimensional) denote the final ranking score of the enlarged graph. Then, we have (5)–(9), shown at the bottom of the page, where h_1 , h_2 , and h_3 are functions of a square matrix. The approximation in (9) is because that $s \cdot s^t$ can be ignored compared to S .

Recall that the first n dimensions of \vec{f}_e^* make up the ranking score \vec{f}^* of the original graph, i.e.,

$$\vec{f}^* = \alpha(I - \alpha S)^{-1} s. \quad (10)$$

Comparing (1) and (10), the difference is that the latter has an additional constant coefficient α (which will not affect the ranking result), and that the seed vector y is replaced by s . For the sake of discrimination, we will name s as the pseudo seed vector in contrast to the seed vector y . Ideally, s should be calculated according to (4). In our current implementation, to further simplify computation, we use the following equation to obtain this pseudo seed vector

$$s(i) = \frac{e(i)}{D_e(n+1, n+1)}. \quad (11)$$

In this way, we do not need to store the matrix D . The vector e together with the matrix S is enough to calculate S_e . Based on (10), if the query image is outside the database, gMRBIR will perform the following two-step procedure.

- 1) Initialization: Spread the ranking score of the query image to its K nearest neighbors in the database, and obtain the pseudo seed vector s according to (11).
- 2) Propagation: Spread the ranking scores of the pseudo seed vector s to all the unlabeled images via manifold ranking.

B. Discussion

Ideally, K_0 (which is used to obtain the affinity matrices W and S in step 1) of the manifold ranking algorithm) and K (which is used to obtain the vectors e and s in gMRBIR) should be the same, as they are integrated to form the enlarged affinity matrix W_e . From another point of view, K determines the number of nonzero elements in the pseudo seed vector s , which correspond to the images most similar to the query. Since such images and the associated similarity in terms of $s(i)$ are determined by low-level features, their precision can be very low, and a large K may introduce a lot of noise in the initialization step. Therefore, we deliberately set $K < K_0$. Experimental results in Section V support our supposition.

Although the two-step procedure in gMRBIR is designed to deal with the problem in which the query image is outside the

$$\vec{f}_e^* \stackrel{\text{Equation 1}}{=} (I - \alpha S_e)^{-1} y_e \quad (5)$$

$$\stackrel{\text{Taylor Expansion}}{\approx} (I + \alpha S_e + \alpha^2 S_e^2 + \alpha^3 S_e^3 + \dots) y_e \quad (6)$$

$$\stackrel{\text{Equation 3}}{\approx} \left(I + \alpha \begin{bmatrix} S & s \\ s^t & 0 \end{bmatrix} + \alpha^2 \begin{bmatrix} S & s \\ s^t & 0 \end{bmatrix}^2 + \alpha^3 \begin{bmatrix} S & s \\ s^t & 0 \end{bmatrix}^3 + \dots \right) \begin{bmatrix} y \\ 1 \end{bmatrix} \quad (7)$$

$$= \begin{pmatrix} (I - \alpha S)^{-1} + h_1(ss^t) & \alpha(I - \alpha S)^{-1} s + h_2(ss^t) \cdot s \\ \alpha s^t (I - \alpha S)^{-1} + s^t \cdot (h_2(ss^t))^t & 1 + \alpha^2 s^t (I - \alpha S)^{-1} s + s^t \cdot h_3(ss^t) \cdot s \end{pmatrix} \begin{bmatrix} y \\ 1 \end{bmatrix} \quad (8)$$

$$\approx \begin{bmatrix} \alpha(I - \alpha S)^{-1} s \\ 1 + \alpha^2 s^t (I - \alpha S)^{-1} s \end{bmatrix} \quad (9)$$

database, it is well suited to solve the problem in which the query image is within the database. To be specific, in such a situation, the query image would be the nearest neighbor of itself, and the corresponding element in the pseudo seed vector s would be sufficiently large as compared to the other elements. If we set $K = 1$, \vec{f}^* obtained from (1) and (10) would be the same except for a constant coefficient, i.e., MRBIR and gMRBIR would produce the same ranking result. From this point of view, MRBIR can be viewed as a speciality of gMRBIR.

IV. LEARNING METHODS IN GMRBIR

A. Relevance Feedback

As noticed by previous research work, in relevance feedback, relevant and irrelevant images should be processed differently due to the asymmetry between the two classes. According to the experimental results in [12], the simplest yet most effective scheme for incorporating feedback images in MRBIR is as follows.

- 1) Define two vectors y^+ and y^- . The element of the former one is set to 1 if the corresponding image is the query or a positive example; while the element of the latter one is set to -1 if the corresponding image is a negative example. All the other elements of the two vectors are set to 0.
- 2) Let \vec{f}^{+*} and \vec{f}^{-*} denote the ranking scores obtained from positive and negative examples respectively. $\vec{f}^{+*} = Ay^+$, $\vec{f}^{-*} = \gamma Ay^-$, where $A = (I - aS)^{-1}$, $\gamma \in (0, 1]$. The smaller γ is, the less impact negative examples will have on \vec{f}^* .
- 3) The final ranking score

$$\vec{f}^* = \vec{f}^{+*} + \vec{f}^{-*}. \quad (12)$$

When the query is within the database, it can always be marked as a relevant image in the first round of relevance feedback; thus, the vector y^+ has at least one nonzero element. However, when the query image is outside the database, y^+ may be a zero vector even after a few rounds of relevance feedback due to the lack of relevant images. In this case, \vec{f}^* will totally depend on irrelevant images to produce the final ranking score, which may be unreliable. Therefore, we revise \vec{f}^* as follows:

$$\vec{f}^* = \eta \vec{f}^0 + (1 - \eta) \vec{f}^{+*} + \vec{f}^{-*} \quad (0 < \eta < 1) \quad (13)$$

where \vec{f}^0 is calculated through (10) based on low-level features. η controls the contribution of \vec{f}^0 to \vec{f}^* , and it decreases with the total number of relevant images (n_r) feedback by the user, i.e.,

$$\eta = \exp(-n_r). \quad (14)$$

When $n_r = 0$, i.e., y^+ is a zero vector, $\eta = 1$, and \vec{f}^* is a weighted combination of the ranking scores obtained from low-level features and negative feedback images; as n_r increases, η becomes smaller. When sufficient relevant images are accumu-

TABLE I
IMAGE FEATURES USED IN gMRBIR

Name	Description & Dimension
Color Correlogram [13]	HSV space, 144 dimensions
Color Histogram [22]	HSV space, 36 dimensions
Tamura Feature [11]	Coarseness, Contrast, Directionality, 20 dimensions
Pyramid Wavelet Texture Feature [23]	24 dimensions

lated from relevance feedback, (13) becomes almost the same as (12).

B. Active Learning

Contrary to passive learning, active learning selects unlabeled images for labeling according to some principle in each round of relevance feedback, hoping to maximally improve the retrieval result. In [12], we adopt three active learning methods based on different principles. The first method selects the most relevant images; the second method selects the most inconsistent images which have the smallest $|f_i^*|$, while the third one selects the inconsistent images which are also quite similar to the query. In the context of image retrieval, where the goal is to identify the user's query concept using a limited number of labeled images, relevant images are more important than negative ones. Based on experimental results, the first method performs the best, which is consistent with theoretical analysis. Therefore, this method is incorporated into gMRBIR as the active learning method.

V. EXPERIMENTAL RESULTS

We have evaluated the performance of gMRBIR using a general-purpose image database consisting of 5 000 Corel images. The images are categorized into 50 groups, such as beach, bird, mountain, jewelry, sunset, etc. Each of the categories contains 100 images of essentially the same content, which serve as the groundtruth. We use each image in the whole database as a query. The remaining 4 999 images constitute the database, from which relevant images are to be found, i.e., the query is outside the database. The results are averaged over the 5 000 queries, and the precision versus scope curve is used to evaluate the performance of various methods.

Feature selection is a large open problem and might have a great impact on the results. Besides the traditional global features [9]–[11], [13], [22], [23], recently a lot of research work has investigated regional features to depict image content [8]. However, regional feature extraction depends on image segmentation, which is far from applicable in real applications. Therefore, in our current implementation, we only adopt global features, which are listed in Table I.² Note that we have normalized each dimension of the features to $[0, 1]$ to eliminate the effect of different scales.

A. gMRBIR Without Relevance Feedback

As validated by the experimental results in [12], $L1$ distance is more appropriate for defining edge weights in the affinity ma-

²We have experimented with many features and feature combinations. The comparison results are consistent with those provided in subsequent sections.

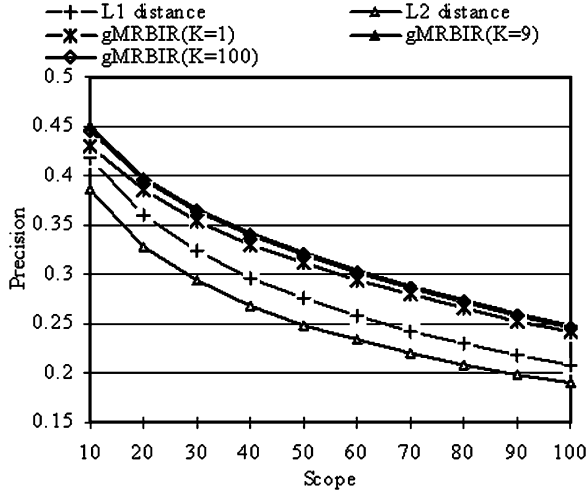


Fig. 2. Comparison without relevance feedback.

trix W than the commonly used $L2$ distance. Therefore, we define edge weights as follows:

$$W_{ij} = \exp \left[-\frac{d_{L1}(x_i, x_j)}{\sigma} \right] \quad (15)$$

where $d_{L1}(x_i, x_j)$ denotes the $L1$ distance between x_i and x_j . Therefore, we need to specify four parameters for gMRBIR to perform image retrieval: α , K_0 , K , and σ . Among them, α is set to 0.99, which is consistent with [6], [7], [12]. We resort to MRBIR to determine the values of K_0 and σ . To be specific, we include all the 4999 images in the database, take each image as a query, and use MRBIR to obtain the retrieval result. We compare different pairs of K_0 and σ and select the one leading to the highest average precision. Ideally, for each testing query, a different pair of parameters should be selected based on the remaining 4999 images in the database. However, in our experiments, we observe that the values do not vary a lot with different queries. So, we fix $K_0 = 100$ and $\sigma = 1$ for simplicity.

As we have discussed in Section III-B, K should be smaller than K_0 to achieve a relatively high precision in the initialization step. Like K_0 and σ , ideally, the value of K should be adapted to each query. One way for selecting proper K value at runtime is to find the K_0 nearest neighbors of the query in the database based on $L1$ or $L2$ distance, sort them in ascending order of distance to the query, and set K to be the number of neighbors before the largest discrepancy is observed. In this subsection, for the sake of simplicity, we only perform experiments with fixed K values for different queries, and compare their performance with that using $L1$ and $L2$ distances. The comparison results are provided in Fig. 2.

In our experiments, the best result is achieved at $K = 9$. When K is very small, e.g., $K = 1$, the nearest neighbors may not be semantically similar to the query image due to the imprecision of low-level features; thus, the performance of gMRBIR may not be very stable. When K is very big, on the other hand, e.g., $K = 100$, the performance is also inferior to $K = 9$, but the difference is not very prominent. The reason can be explained as follows: although the inclusion of many neighbors to initialize

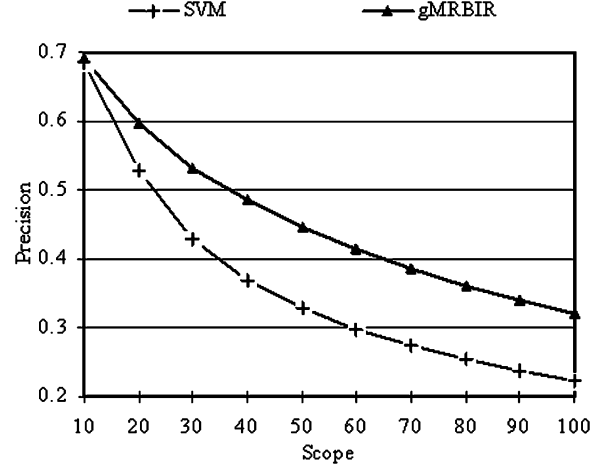


Fig. 3. One feedback with 20 images.

the pseudo seed vector s will introduce noise into the retrieval process, their effect is not very obvious, since the edge weight W_{ij} declines very fast with the increase of the distance between the query and database images ((15)).

Comparing the performance of gMRBIR with that of $L1$ and $L2$ distances, we can see that our method achieves significant improvement over traditional methods based on pairwise distances.

B. gMRBIR With Relevance Feedback

In this subsection, we perform experiments to test the performance of the learning methods (including relevance feedback and active learning) used in gMRBIR and compare with SVM [19]. Users' feedback processes are simulated as follows. In each round of relevance feedback, a certain number of images that are most relevant to the query based on the judgment of the current system are feedback and examined (the most relevant strategy). Images from the same (different) category as the query are used as new positive (negative) examples. For SVM, $L1$ distance is utilized in the initial retrieval stage and the adopted kernel is Gaussian kernel. When both positive and negative examples are available, an SVM classifier can be trained to predict the relevance of unlabeled images. Note that this method will not be affected no matter whether or not the query image is in the database. For gMRBIR, the parameter γ in relevance feedback is set to 0.25, which has been demonstrated to achieve the best result in [12]. To provide a systematic evaluation, we fix the total number of images that are labeled in relevance feedback to 20, but vary the times of feedback and the number of images feedback each time accordingly. The combinations used in this experiment include: one feedback with 20 images each time, two feedbacks with ten images each time, and four feedbacks with five images each time. In all these experiments, gMRBIR outperforms SVM by a large margin. In Figs. 3–5, we present their retrieval results after the first round of relevance feedback when 20 images, ten images, and five images are feedback, respectively.

Comparing the three figures with Fig. 2, we can see that when training data is not sufficient, SVM brings degradation to the performance when $L1$ distance is utilized to measure relevance

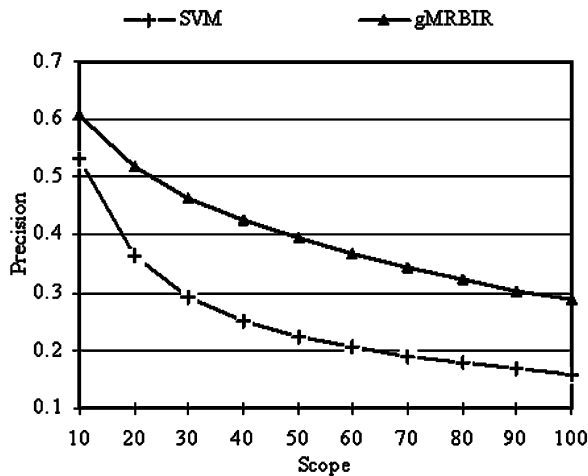


Fig. 4. One feedback with ten images.

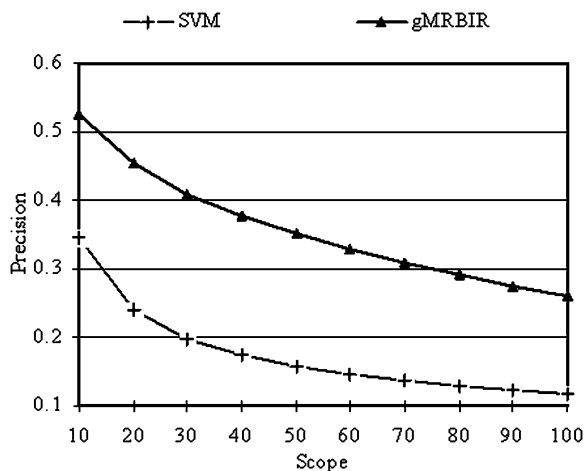


Fig. 5. One feedback with five images.

between the query and database images; while gMRBIR consistently improves the initial retrieval result no matter how many images are feedback by the user. Take P50 (precision within the top 50 retrieved images) as an example. Using $L1$ distance, we get P50 at 27.5%; when 20 images are feedback by the user, SVM improves P50 to 32.8%; when ten images are feedback by the user, P50 using SVM becomes 22.5%; when only five images are feedback by the user, P50 is further decreased to 15.8%. Using gMRBIR, we get initial P50 at 32.2%; when 20 images are feedback by the user, gMRBIR improves P50 to 44.7%; when ten images are feedback by the user, gMRBIR improves P50 to 39.4%; when only five images are feedback by the user, P50 is still improved to 35.1%.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a general transductive learning framework named generalized manifold-ranking-based image retrieval. If the query image is in the database, it could produce the same ranking result as MRBIR if the number of neighbors used to initialize the pseudo seed vector is small, while, if the query image is outside the database, it first spreads the ranking score of the query image to its neighbors, and

then further spreads the scores to all the unlabeled images via manifold ranking. Furthermore, the best learning methods used in MRBIR are incorporated into gMRBIR with some modifications to ensure that reasonable retrieval results can be obtained in all cases. Experimental results on a general-purpose image database show that, in the initial retrieval stage, gMRBIR is much better than traditional methods based on pairwise distances; in relevance feedback, gMRBIR consistently improves the retrieval result and outperforms SVM. Currently, we are trying to construct multiple graphs based on different features, and then incorporate information from these graphs to perform retrieval, which will hopefully make better use of these features and give even higher precision.

ACKNOWLEDGMENT

The authors would like to thank S. Yan, X. Zheng, L. Zhang, and X. Yi for their valuable discussions and enlightening comments, as well as the reviewers for their comments and suggestions.

REFERENCES

- [1] B. Li, E. Chang, and C. S. Li, "Learning image query concepts via intelligent sampling," in *Proc. IEEE Int. Conf. Multimedia & Expo*, 2001, pp. 961–964.
- [2] B. Li, E. Chang, and C. T. Wu, "DPF—a perceptual distance function for image retrieval," in *Proc. IEEE Int. Conf. Image Processing*, 2002, vol. 2, pp. 597–600.
- [3] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
- [4] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, "Efficient and effective querying by image content," *J. Intell. Inf. Syst.*, vol. 3, no. 3–4, pp. 231–262, 1994.
- [5] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 530–535, May 1997.
- [6] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Chölkopf, "Learning with local and global consistency," presented at the NIPS 2003.
- [7] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Chölkopf, "Ranking on data manifolds," presented at the NIPS 2003.
- [8] F. Jing, M. Li, H. J. Zhang, and B. Zhang, "An effective region-based image retrieval framework," presented at the 10th ACM Int. Conf. Multimedia 2002.
- [9] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 7, Jul. 1996.
- [10] G. Pass, "Comparing images using color coherence vectors," in *Proc. 4th ACM Int. Conf. Multimedia*, 1997, pp. 65–73.
- [11] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.*, vol. 8, no. 6, pp. 460–472, Jun. 1978.
- [12] J. He, M. Li, H. J. Zhang, H. Tong, and C. Zhang, "Manifold-ranking based image retrieval," in *Proc. 12th ACM Int. Conf. Multimedia*, 2004, pp. 9–16.
- [13] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1997, pp. 762–768.
- [14] J. R. Smith and S. F. Chang, "VisualSEEK: A fully automated content-based query system," in *Proc. 4th ACM Int. Conf. Multimedia*, 1996, pp. 87–98.
- [15] J. Z. Wang, G. Wiederhold, O. Firschein, and X. W. Sha, "Content-based image indexing and searching using Daubechies' wavelets," *Int. J. Digit. Libraries*, vol. 1, no. 4, pp. 311–328, 1998.
- [16] K. Porkaew, M. Ortega, and S. Mehrotra, "Query reformulation for content based multimedia retrieval in MARS," in *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, 1999, vol. 2, pp. 747–751.

- [17] K. Porkaew and K. Chakrabarti, "Query refinement for multimedia similarity retrieval in MARS," in *Proc. 7th ACM Int. Conf. Multimedia*, 1999, pp. 235–238.
- [18] K. Tieu and P. Viola, "Boosting image retrieval," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2000, vol. 1, pp. 228–235.
- [19] L. Zhang, F. Lin, and B. Zhang, "Support vector machine learning for image retrieval," in *Proc. IEEE Int. Conf. Image Processing*, 2001, vol. 2, pp. 721–724.
- [20] M. Kokare, B. N. Chatterji, and P. K. Biswas, "Comparison of similarity metrics for texture image retrieval," in *IEEE Conf. Convergent Technologies for Asia-Pacific Region*, 2003, vol. 2, pp. 571–575.
- [21] M. Stricker and M. Orengo, "Similarity of color images," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1995, pp. 381–392.
- [22] M. Swain and D. Ballard, "Color indexing," *Int. J. Comput. Vis.*, vol. 7, no. 1, pp. 11–32, 1991.
- [23] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [24] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," presented at the 9th ACM Int. Conf. Multimedia 2001.
- [25] J. P. Tarel and S. Boughorbel, "On the choice of similarity measures for image retrieval by example," in *Proc. 10th ACM Int. Conf. Multimedia*, 2002, pp. 446–455.
- [26] X. S. Zhou, Y. Rui, and T. Huang, "Water-Filling: A novel way for image structural feature extraction," in *Proc. IEEE Int. Conf. Image Processing*, 1999, vol. 2, pp. 570–574.
- [27] Y. Ishikawa, R. Subramanya, and C. Faloutsos, "MindReader: Querying databases through multiple examples," presented at the 24th Int. Conf. Very Large Data Bases 1998.
- [28] Y. Rubner, C. Tomasi, and L. Guibas, "A metric for distributions with applications to image databases," in *Proc. IEEE Int. Conf. Computer Vision*, 1998, pp. 59–66.
- [29] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 644–655, Sep. 1998.
- [30] Y. Rui, T. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int. Conf. Image Processing*, 1997, pp. 815–818.
- [31] Y. Wu, Q. Tian, and T. Huang, "Discriminant-EM algorithm with application to image retrieval," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2000, vol. 1, pp. 155–162.



Jingrui He received the B.S. and M.S. degrees from the Automation Department, Tsinghua University, China, in 2002 and 2005, respectively. She is currently pursuing the Ph.D. degree at the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Her research interests include statistical machine learning, information retrieval, and multimedia.



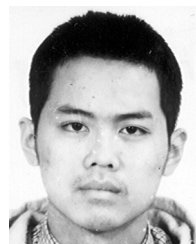
Mingjing Li received the B.S. degree in Electrical Engineering from the University of Science and Technology of China in 1989 and the Ph.D. degree in pattern recognition from the Institute of Automation, Chinese Academy of Sciences, in 1995.

He joined Microsoft Research Asia, Beijing, in July 1999. His research interests include pattern recognition and multimedia search.



Hong-Jiang Zhang received the B.S. degree from Zhengzhou University, China, and the Ph.D. degree from the Technical University of Denmark, both in electrical engineering, in 1982 and 1991, respectively.

From 1992 to 1995, he was with the Institute of Systems Science, National University of Singapore, where he led several projects in video and image content analysis and retrieval and computer vision. From 1995 to 1999, he was a Research Manager at Hewlett-Packard Labs, responsible for research and technology transfers in the areas of multimedia management, computer vision and intelligent image processing. In 1999, he joined Microsoft Research Asia, Beijing, where he is currently the Managing Director of Advanced Technology Center. He has authored three books, over 300 referred papers, eight special issues of international journals on image and video processing, content-based media retrieval, and computer vision, as well as over 50 patents or pending applications. He currently serves on the editorial boards of five IEEE/ACM journals and a dozen committees of international conferences.



Hanghang Tong received the B.S. and M.S. degrees from the Automation Department, Tsinghua University, China, in 2002 and 2005, respectively. He is currently pursuing the Ph.D. degree at the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

His research interests include data mining and statistical machine learning.



Changshui Zhang received the B.S. degree in Mathematics from Peking University, China, in 1986, and the Ph.D. degree from Tsinghua University, Beijing, China, in 1992.

He is currently a professor with Tsinghua University. His interests include pattern recognition, machine learning, computer vision, image processing, complex networks, etc.