

Multiple Random Walk and Its Application in Content-Based Image Retrieval*

Jingrui He¹, Hanghang Tong¹, Mingjing Li², Wei-Ying Ma², Changshui Zhang³

^{1,3}Department of Automation, Tsinghua University, Beijing 100084, China

²Microsoft Research Asia, 49 Zhichun Road, Beijing 100080, China

¹{hejingrui98, walkstar98}@mails.tsinghua.edu.cn, ²{mjli, wyma}@microsoft.com

³zcs@tsinghua.edu.cn

ABSTRACT

In this paper, we propose a transductive learning method for content-based image retrieval: Multiple Random Walk (MRW). Its basic idea is to construct two generative models by means of Markov random walks, one for images relevant to the query concept and the other for the irrelevant ones. The goal is to obtain the likelihood functions of both classes. Firstly, MRW generates two random walks with virtual absorbing boundaries, and uses the absorbing probabilities as the initial estimation of the likelihood functions. Then it refines the two random walks through an EM-like iterative procedure in order to get more accurate estimation of the likelihood functions. Class priors are also obtained in this procedure. Finally, MRW ranks all the unlabeled images in the database according to their posterior probabilities of being relevant. By using both labeled and unlabeled data, MRW can be seen as a transductive learning method, which has been demonstrated to outperform inductive ones by previous research work. Systematic experiments on a general-purpose image database consisting of 5,000 Corel images demonstrate the superiority of MRW over state-of-the-art techniques.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications – *image databases*; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *search process, relevance feedback*.

General Terms

Algorithms, Experimentation, Theory.

Keywords

Markov random walk, generative model, likelihood function, class prior, image retrieval, relevance feedback.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MIR '05, November 10-11, 2005, Singapore.

Copyright 2005 ACM 1-59593-244-5/05/0011...\$5.00.

1. INTRODUCTION

With the prevalence of World Wide Web, the number of accessible digital images is growing at an exponential speed. To make the best use of these resources, people need an efficient and effective tool to manage them. In such context, Content-Based Image Retrieval (CBIR) was proposed in the early 1990's. It depends on low-level features to find images relevant to the query concept, which is represented by the query example provided by the user. Comparing with traditional image retrieval paradigms, where the images are first annotated manually by keywords, and then retrieved according to the annotations, CBIR has the following advantages: 1. feature extraction can be performed fully automatically without human intervention; 2. the images' own content is always consistent.

The initial research work in CBIR focuses on exploring an ideal descriptor for image content. Generally speaking, these low-level features can be categorized into three major groups: color [4, 12, 20], texture [8, 10, 11, 23], and shape [16, 17, 28]. However, their performance is far from satisfactory due to the well-known gap between low-level features and high-level semantic concepts. To be specific, 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 [2].

The first attempt towards bridging this gap is searching for appropriate metrics to measure perceptual similarity. These metrics can be classified into pair-wise and global ones. The first class includes the simplest Minkowski distances ($L1$ distance, $L2$ distance, etc), as well as the more complicated ones, such as Dynamic Partial distance Function (DPF) [7] and Earth Mover's Distance (EMD) [14]. A representative example of the second class is the transductive learning framework named Manifold-Ranking Based Image Retrieval (MRBIR) proposed by He et al [2], which uses the manifold-ranking algorithm to explore the relationship among all the data points in the feature space, and then measures relevance between the query and all the images in the database accordingly. Comparing these two classes of metrics, the former has the advantage of very fast processing speed, while the latter is more accurate since it takes into consideration global relationship. However, this attempt is subject to the limited information available to describe the query concept, that is, we only have one query example provided by the user. Therefore, the

* This work was performed at Microsoft Research Asia.

proposed similarity metrics cannot always measure perceptual similarity accurately, and the retrieval precision is low.

A more promising line of research to bridge the semantic gap is to incorporate relevance feedback [13] into CBIR. Relevance feedback is a powerful tool borrowed from the community of information retrieval [15]. In each round, the CBIR system returns a certain number of images to the user, and asks for their relevance degree. In this way the system is able to gradually learn the user’s query concept, and improve the retrieval result. To make the best use of this additional information, an effective learning method should be used to identify the relevant images from the irrelevant ones. Again, state-of-the-art learning methods can be classified as inductive and transductive ones, according to whether unlabeled data is utilized in the training stage or not [2]. Inductive methods include Support Vector Machine (SVM) [25], boosting [22], etc. These methods essentially construct discriminative models for the two classes of images (relevant vs. irrelevant), and rank the images in the database according to the classification results. However, they all suffer from the small-sample-size problem, which might bring great degradation to the accuracy of the discriminative models. On the other hand, transductive methods provide a way to solve this problem by using unlabeled data, such as Discriminant-EM [24], MRBIR [2], etc. D-EM proposed by Wu et al [24] makes use of unlabeled data to construct a generative model, which will be used to measure relevance between the query and database images. However, if the components of data distribution are mixed up, its performance will be compromised. In MRBIR, when both relevant and irrelevant images are fed back by the user, it discriminately spreads their ranking scores via the unlabeled data, and then combines the scores to get the improved retrieval result. One major problem with MRBIR is that no general guideline has been given to spread the scores of irrelevant images, that is, how to determine the coefficient that affects the contribution of negative ranking scores to the final score. This problem affects the generalization performance of MRBIR to different databases.

In this paper, we propose a novel transductive method for CBIR: Multiple Random Walk (MRW). It constructs two generative models for the two classes of images (relevant and irrelevant) by means of Markov random walks, using both labeled and unlabeled data. The goal is to estimate the likelihood functions that given the positive (negative) class, a certain data point is generated. The difference between D-EM and MRW is the method used to construct the generative models. The former utilizes Gaussian distribution to represent both classes, which might be too simple to depict the complex distribution of image features; while the latter utilizes the graphical model to achieve this goal, which is suitable to capture the relationship among data points with various underlying distributions. On the other hand, when comparing MRBIR with MRW, we can see that the former is a special example of the latter where class priors are fixed beforehand (they correspond to the coefficient that affects the contribution of negative ranking scores to the final score), and the two generative models do not fully utilize the label information. By allowing the class priors to be determined automatically, MRW can be easily applied to different databases; by making full use of the label information to characterize the two random walks, we can better understand the query concept, and achieve higher precision.

In MRW, when we design the two generative models, two virtual absorbing boundaries are added to each random walk: one positive boundary, and the other non-positive boundary. Only labeled data are connected with the positive boundary; while the others are connected with the non-positive boundary. The transition probabilities between vertices other than the absorbing boundaries are based on the affinity matrix. After sufficiently large transition steps, the random walks output the absorbing probabilities that correspond to each data point, which, after normalization, would be taken as the initial estimation of the likelihood functions. Then through an EM-like iterative procedure, MRW refines the random walks in terms of their transition matrices, in order to better estimate the likelihood functions. Class priors are also obtained in this procedure. Finally, by feeding these likelihood functions and the class priors into the Bayes’ formula, MRW outputs the posterior probability that each data point is relevant to the query concept, and ranks the data points accordingly.

In the initial retrieval stage, where only one query example is available, the generative model for images relevant to the query concept can be readily constructed by Markov random walk, while that for irrelevant images cannot due to the lack of negative examples. Instead, the likelihood function of the negative class can be approximated based on pair-wise distance between low-level features. In relevance feedback, both relevant and irrelevant images are provided by the user. They can be used to construct the two generative models, which will lead to more accurate estimation of the posterior probabilities, and better retrieval results accordingly.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed Multiple Random Walk. Then we elaborate on the details of its application in CBIR in Section 3, including the initial retrieval stage and relevance feedback. Experimental results are given in Section 4, followed by concluding remarks in Section 5.

2. MULTIPLE RANDOM WALK

From the perspective of probability theory, image retrieval can be seen as a procedure of ranking images in the database according to their posterior probabilities of being relevant to the query concept. To obtain the posterior probabilities, we need to know the class priors (the prior probabilities of the positive and negative classes) and the likelihood functions (the probability of generating a certain data point given the class label). In this section, we introduce MRW in a general setting. After giving some notations, we will introduce the way by which MRW calculates the likelihood functions via Markov random walks, and then present the EM-like iterative procedure in which both the priors and the likelihood functions are refined gradually. Finally, we will give the flowchart of MRW to help illustrate this transductive method.

2.1 Notations

Consider a binary classification problem. Suppose that we have a set of n data points:

$\mathcal{X} = \{x_1, \dots, x_{n_+}, x_{n_++1}, \dots, x_{n_++n_-}, x_{n_++n_-+1}, \dots, x_n\} \subset \square^m$. The first n_+ points are positive examples with $y_i = 1, i = 1, \dots, n_+$, the next n_- points are negative examples with $y_i = -1, i = n_++1, \dots, n_++n_-$, while the remaining $n - n_+ - n_-$ points are unlabeled examples.

Furthermore, we define an undirected graph $G=(V,E)$, where V is the vertex set including the n data points, and E corresponds to the edges connecting two vertices with weights $w_{ij}(i,j=1,\dots,n)$. Our goal is to predict the class labels of all the unlabeled points based on their posterior probabilities of being positive and negative.

2.2 Generative Model in MRW

To begin with, let us come to a brief introduction to Markov random walk. A random walk is a finite Markov chain that is time-reversible [9]. To be specific, imagine a particle walking along the vertices of a graph. From a starting point, it moves to one of its neighbors with a certain probability in one step; then from this point, it moves again in subsequent step, etc. It is worth noticing that Markov random walk satisfies the following Markov property: $P[X_{t+1}|X_0,X_1,\dots,X_t]=P[X_{t+1}|X_t]$, where X_t denotes the status of the particle at time t . That is, the next move does not depend on previous status of the particle. Therefore, each random walk can be uniquely characterized by a transition matrix. Markov random walk has been applied in many research fields [9]. For example, by making use of the fact that the distribution of the particle status tends to the stationary distribution as $t \rightarrow \infty$, it can be used to design sampling algorithms. In this paper, we apply Markov random walk to construct generative models for the two classes.

We first define the positive random walk (the random walk for the positive class). Let W denote the $n \times n$ affinity matrix of graph G , with its elements being w_{ij} . Let D denote the $n \times n$ diagonal matrix, with the (i,i) -element equal to the sum of the i th row of W . With these two matrices, we define $P=D^{-1}W$. To construct the positive random walk, we add two virtual absorbing boundaries to G to get an enlarged graph $G^+=\left(\{V,v_{G^+}^+,v_{G^+}^-\},E^+\right)$, where $v_{G^+}^+$ and $v_{G^+}^-$ correspond to the positive absorbing boundary and the non-positive absorbing boundary respectively, and E^+ includes all the edges in E as well as those connecting the absorbing boundaries and the original vertices. To understand the meaning of absorbing boundaries, again let us consider the particle walking along the vertices of G^+ . An absorbing boundary is such a vertex that once the particle moves to it, it will by no means move to other vertices thereafter.

Next we design the connection between absorbing boundaries and the vertices in V for the positive random walk. It is reasonable to connect the positive absorbing boundary $v_{G^+}^+$ and the n_+ positive examples. In other words, starting from vertices corresponding to positive labeled examples, the particle have certain probabilities of being absorbed by $v_{G^+}^+$ in one step; while starting from other vertices, the particle cannot reach $v_{G^+}^+$ in one step. Likewise, we can also connect the non-positive absorbing boundary $v_{G^+}^-$ and the remaining $n-n_+$ examples. In other words, the particle may be absorbed by $v_{G^+}^-$ in one step if and only if it starts from vertices corresponding to the remaining $n-n_+$ examples.

Based on the above discussion, the $(n+2) \times (n+2)$ transition matrix of the enlarged graph G^+ can be written as follows:

$$P_{G^+} = \begin{bmatrix} 1 & 0 & 0_n^T \\ 0 & 1 & 0_n^T \\ B_{G^+} y^+ & B_{G^+} (1_n - y^+) & A_{G^+} P \end{bmatrix} \quad (1)$$

$$B_{G^+} = \begin{bmatrix} \beta & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \beta \end{bmatrix}, \quad A_{G^+} = \begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \alpha \end{bmatrix} \quad (2)$$

where 1_n (0_n) is an n -dimensional vector consisting of 1 s (0 s). y^+ is an n -dimensional label vector: $y_i^+ = 1$, for $i = 1, \dots, n_+$; $y_i^+ = 0$, for $i = n_+ + 1, \dots, n$. B_{G^+} and A_{G^+} are $n \times n$ diagonal matrices, with positive parameters α and β satisfying $\alpha + \beta = 1$. The first (second) row denotes the probabilities of the particle moving to the $n+2$ vertices in one step given that it is currently at $v_{G^+}^+$ ($v_{G^+}^-$). The other rows denote the transition probabilities given that the particle is currently at one of the vertices in V . From P_{G^+} , we can see that starting from positive labeled examples, the particle has a constant probability β of being absorbed by $v_{G^+}^+$ in one step, while starting from all the other examples, it has a constant probability β of being absorbed by $v_{G^+}^-$ in one step. The value of β should be small enough so that the particle can visit many vertices in V before being absorbed by $v_{G^+}^+$ or $v_{G^+}^-$, thus global information is effectively utilized.

By calculating $(P_{G^+})^\infty$, we can obtain the absorbing probability (the probabilities that starting from each vertex in V , the particle is finally absorbed by $v_{G^+}^+$) in the positive random walk:

$$p^+ = (I - A_{G^+} P)^{-1} \cdot B_{G^+} y^+ \quad (3)$$

Furthermore, we normalize p^+ to get f^+ such that $1_n^T \cdot f^+ = 1$, which is used as the likelihood function of the positive class:

$$P(x_i | y = 1) = f_i^+ \quad (4)$$

Negative random walk (the random walk for the negative class) is constructed in a similar way. Again, two virtual absorbing boundaries are added to G to get the enlarged graph $G^- = \left(\{V,v_{G^-}^+,v_{G^-}^-\},E^-\right)$. The $(n+2) \times (n+2)$ transition matrix on this enlarged graph is as follows:

$$P_{G^-} = \begin{bmatrix} 1 & 0 & 0_n^T \\ 0 & 1 & 0_n^T \\ B_{G^-} y^- & B_{G^-} (1_n - y^-) & A_{G^-} P \end{bmatrix} \quad (5)$$

$$B_{G^-} = \begin{bmatrix} \beta & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \beta \end{bmatrix}, A_{G^-} = \begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \alpha \end{bmatrix} \quad (6)$$

where y^- is an n -dimensional label vector: $y_i^- = 1$, for $i = n_+ + 1, \dots, n_+ + n_-$; $y_i^- = 0$, for $i = 1, \dots, n_+, n_+ + n_- + 1, \dots, n$. By calculating $(P_{G^-})^\infty$, we can obtain the absorbing probability (the probabilities that starting from each vertex in V , the particle is finally absorbed by $v_{G^-}^+$) in the negative random walk:

$$p^- = (I - A_{G^-} P)^{-1} \cdot B_{G^-} y^- \quad (7)$$

After normalization, we get f^- , which is used as the likelihood function of the negative class:

$$P(x_i | y = -1) = f_i^- \quad (8)$$

2.3 EM-like Iterative Procedure in MRW

When we calculate the likelihood functions, one potential problem with Equation (2) and Equation (6) is the constant probability β of the particle being absorbed by the absorbing boundaries in one step. Intuitively, starting from different data points, the particle should have different probabilities of being absorbed. On the other hand, according to the Bayes' formula, to obtain the posterior probabilities, besides the likelihood functions, we also need to know the class priors ($P(y=1)$ and $P(y=-1)$).

In this section, we will present an EM-like iterative procedure used in MRW to refine the generative models as well as to obtain the class priors.

Firstly, f^+ and f^- are used to initialize the likelihood functions of the positive class and negative class respectively. And the class priors, $P^{(0)}(y=1)$ and $P^{(0)}(y=-1)$, are both set to 0.5.

In E-step, by feeding the current estimation of the likelihood functions and the class priors into the Bayes' formula, we can obtain the posterior probabilities of each unlabeled example being positive and negative:

$$\begin{aligned} P^{(s+1)}(y_i = 1 | x_i) &= \frac{P^{(s)}(y=1)P^{(s)}(x_i | y=1)}{P^{(s)}(y=-1)P^{(s)}(x_i | y=-1) + P^{(s)}(y=1)P^{(s)}(x_i | y=1)} \\ P^{(s+1)}(y_i = -1 | x_i) &= \frac{P^{(s)}(y=-1)P^{(s)}(x_i | y=-1)}{P^{(s)}(y=-1)P^{(s)}(x_i | y=-1) + P^{(s)}(y=1)P^{(s)}(x_i | y=1)} \end{aligned} \quad (9)$$

In M-step, based on the current estimation of the posterior probabilities, we can refine the class priors as well as the random walks. To be specific, the class priors are updated as follows:

$$P^{(s+1)}(y=1) = \frac{1}{n} \sum_{i=1}^n P^{(s+1)}(y_i = 1 | x_i)$$

$$P^{(s+1)}(y=-1) = \frac{1}{n} \sum_{i=1}^n P^{(s+1)}(y_i = -1 | x_i) \quad (10)$$

Here we assume $P(x_i) = 1/n$, for $i = 1, \dots, n$.

To update the positive (negative) random walk, we make the following assumption: in the positive (negative) random walk, starting from a positive (negative) labeled example, the probability that the particle is absorbed by $v_{G^+}^+$ ($v_{G^-}^+$) in one step is equal to the current estimation of the posterior probability that this example is a positive (negative) one multiplied by β . Note that in this case, by introducing the posterior probabilities, MRW can be less affected by noise in the labeled data. Similarly, starting from all the other examples, the probability that the particle is absorbed by $v_{G^+}^-$ ($v_{G^-}^-$) in one step is equal to the current estimation of the posterior probability that this example is NOT a positive (negative) one multiplied by β . To be specific, in the positive random walk, if the i th example is a positive labeled one,

$$B_{G^+}^{(s+1)}(i, i) = \beta \cdot P^{(s+1)}(y_i = 1 | x_i) \quad (11)$$

$$A_{G^+}^{(s+1)}(i, i) = 1 - \beta P^{(s+1)}(y_i = 1 | x_i) \quad (12)$$

Otherwise,

$$B_{G^+}^{(s+1)}(i, i) = \beta (1 - P^{(s+1)}(y_i = 1 | x_i)) \quad (13)$$

$$A_{G^+}^{(s+1)}(i, i) = 1 - \beta (1 - P^{(s+1)}(y_i = 1 | x_i)) \quad (14)$$

where $B_{G^+}^{(s+1)}(i, i)$ and $A_{G^+}^{(s+1)}(i, i)$ are the (i, i) -element of $B_{G^+}^{(s+1)}$ and $A_{G^+}^{(s+1)}$ respectively. In the negative random walk, if the j th example is a negative labeled one,

$$B_{G^-}^{(s+1)}(j, j) = \beta \cdot P^{(s+1)}(y_j = -1 | x_j) \quad (15)$$

$$A_{G^-}^{(s+1)}(j, j) = 1 - \beta P^{(s+1)}(y_j = -1 | x_j) \quad (16)$$

Otherwise,

$$B_{G^-}^{(s+1)}(j, j) = \beta (1 - P^{(s+1)}(y_j = -1 | x_j)) \quad (17)$$

$$A_{G^-}^{(s+1)}(j, j) = 1 - \beta (1 - P^{(s+1)}(y_j = -1 | x_j)) \quad (18)$$

With these matrices, the updated likelihood functions can be obtained with equations (3), (4) and (7), (8). In this way, the label information is effectively incorporated into the two random walks, which is expected to improve the estimation of the likelihood functions.

The E-step and M-step are executed alternatively and iteratively until a stable state is reached. The estimation of the posterior probabilities in the last iteration are output and used in image retrieval. Since the proposed EM-like iterative procedure essentially works in a hill-climbing manner [1], it is not expected to reach a global optimal solution in all cases. However, according to the experimental results in Section 4, it still improves

the performance of state-of-the-art techniques in CBIR considerably.

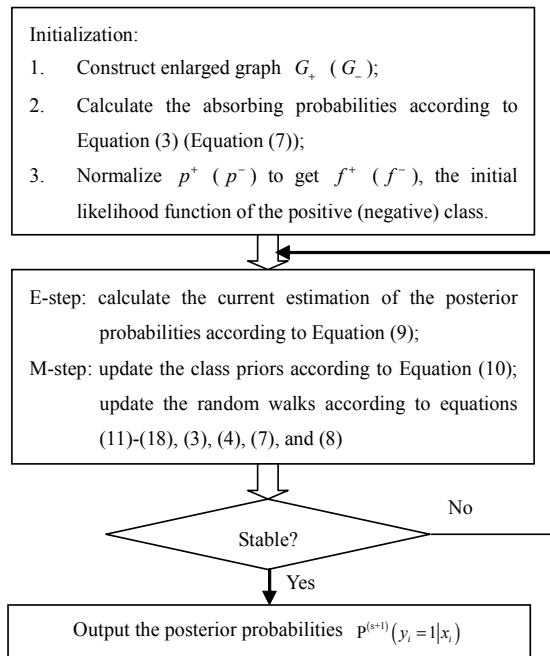


Figure 1. Flowchart of MRW.

2.4 Flowchart of MRW

According to the above discussion, the flowchart of MRW is summarized in Figure 1.

3. MULTIPLE RANDOM WALK IN CBIR

When applying MRW to CBIR, several aspects have to be well considered: 1. the initial retrieval stage where only one query image can be used as the labeled example; 2. relevance feedback where both relevant and irrelevant images are provided by the user.

3.1 The Initial Retrieval Stage

There are two typical query scenarios in CBIR: query by keyword (QBK) and query by example (QBE). While the former need additional information about the annotation of some database images, the latter can be performed solely based on low-level features, and it is the focus of this paper.

In the initial retrieval stage, there is only one image provided by the user to express his or her query concept. With this query example, the initial generative model of the positive class can be easily constructed according to Subsection 2.2. However, due to the lack of irrelevant images, the initial generative model of the negative class cannot be constructed since the particle will never be absorbed by the positive absorbing boundary in the negative random walk, no matter where it starts. As an alternative, we design the following scheme to define the likelihood function of the negative class. Let d_{\max} and d_{\min} denote the maximum and minimum distances between the low-level features of the query example and database images. Then the initial likelihood function of the negative class is defined as follows:

$$f_i^- = \frac{d(q,i) - d_{\min}}{\lambda + d_{\max} - d(q,i)} \quad (19)$$

where $d(q,i)$ is the distance between the low-level features of the query example and the i th image in the database; λ is a small positive parameter to avoid zero denominator. In this scheme, the larger $d(q,i)$ is, the more probable this image is generated by the negative class. Based on this underlying principle, the likelihood function of the negative class can be approximated using more complex functions. In this paper, Equation (19) is adopted for the sake of simplicity.

In the EM-like iterative procedure of MRW, the positive random walk will not be revised in each iteration. This is to avoid magnifying the estimation error since the number of labeled example is very limited (only one positive example). On the other hand, the likelihood function of the negative class is fixed at Equation (19) and will not be revised either. Therefore, only the class priors and the posterior probabilities are updated in each iteration.

Finally all the unlabeled images in the database are ranked in descending order of their posterior probabilities of being relevant to the query concept, which is obtained in the last iteration of the EM-like iterative procedure.

3.2 Relevance Feedback

In relevance feedback, the user will evaluate the system's current judgment on image relevance, and provide the labels for some unlabeled images, both relevant and irrelevant. With these labeled examples, we can construct the initial generative models for the two classes of images following the equations in Subsection 2.2. Next we can refine these models and the class priors via the EM-like iterative procedure, as in Subsection 2.3. Finally, we can obtain the posterior probability of each unlabeled image being relevant to the query concept, and rank the images accordingly.

4. EXPERIMENTAL RESULTS

We have evaluated the performance of MRW on a general-purpose image database consisting of 5,000 Corel images. These 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 database as a query, and average the results over the 5,000 queries. The precision vs. scope curve is used to evaluate the performance of various methods.

As is well known, feature selection is a large open problem and might have a great impact on the retrieval results [2]. Besides the traditional global features [4, 8, 11, 12, 17], recently a lot of research work has investigated regional features to depict image content [5]. However, regional features depend on image segmentation, the performance of which is far from satisfactory. Therefore, in our current implementation, we only adopt global features, which are listed in Table 1. Note that we have normalized each dimension of the features to [0,1] to eliminate the effect of different scales.

Table 1. Features used in the experiments

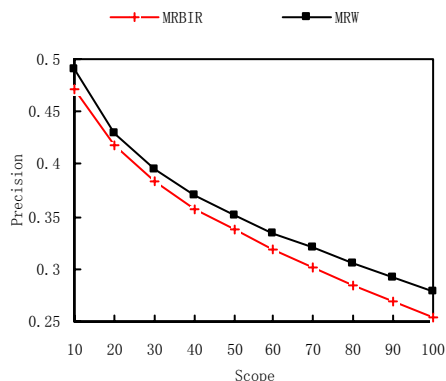
Name	Description & Dimension
Color Correlogram [4]	HSV space, 144 dimensions
Color Histogram [20]	HSV space, 36 dimensions
Tamura Feature [8]	20 dimensions
Pyramid Wavelet Texture Feature [10]	24 dimensions

As we adopt the method in [2] to construct the affinity matrix W , two parameters need to be set in MRW. The first one is the number of neighbors K used to construct the affinity matrix. In our experiments, it is set to 100 although the performance is not very sensitive to its specific value. The second parameter is β , which is closely related to the probability of the particle being absorbed by v_{G^+} (v_{G^-}) in one step, if it starts from a positive (negative) labeled example. As discussed in Subsection 2.2, it should be small enough so that the particle can visit many vertices in V before being absorbed by v_{G^+} (v_{G^-}). In our experiments, we set $\beta = 0.01$ ($\alpha = 0.99$), which is consistent with [2, 26, 27].

To demonstrate the effectiveness of MRW, we compare MRW with MRBIR both in the initial retrieval stage and in relevance feedback. The parameters of MRBIR are the same as in [2] in order to make a fair comparison. Note that since He et al has demonstrated the superiority of MRBIR over traditional methods in [2], we will not include their results in our figures.

4.1 The Initial Retrieval Stage

In this subsection, we compare MRW and MRBIR in the initial retrieval stage where only one query example is provided by the user. The comparison results are illustrated in Figure 2. From the figure, we can see that MRW consistently improves MRBIR no matter what the scope is. Furthermore, as the scope increases, the advantage of MRW over MRBIR becomes more obvious. Take P100 (the precision among the top 100 retrieved images) as an example, it is 0.254 using MRBIR, and is 0.280 using MRW, which improves MRBIR by 10.2%. Therefore, we can conclude that, by making use of the likelihood function of the negative class (Equation (19)) and by adjusting the class priors in the EM-like iterative procedure, MRW is able to learn the query concept more accurately.

**Figure 2. Comparison in the initial retrieval stage.**

4.2 Relevance Feedback

In relevance feedback, both positive and negative examples are provided by the user, so the number of labeled examples is much bigger than that in the initial retrieval stage, and the retrieval performance is expected to be considerably improved. Note that the most relevant images judged by the current system are presented to and labeled by the user in each round of relevance feedback, which is the so-called most relevant strategy in [3].

Next we compare MRW and MRBIR with respect to their ability to improve the retrieval performance in relevance feedback. As in [2], to provide a systematic evaluation, we fix the total number of images that are marked by the user to 20, but vary the times of feedback and the number of images fed back each time accordingly. The combinations include: 1 feedback with 20 images each time; 2 feedbacks with 10 images each time; and 4 feedbacks with 5 images each time. The experimental results are presented in Figure 3, Figure 4, and Figure 5 respectively.

From the figures, we can see that MRW outperforms MRBIR no matter which combination is used. Furthermore, as in the initial retrieval stage, when the scope increases, the advantage of MRW over MRBIR becomes more obvious. In contrast, the advantage of MRBIR over the traditional SVM-based method becomes less obvious with larger scope values [2]. Another interesting observation is that: as fewer images are fed back in each round of relevance feedback, the improvement of MRW over MRBIR becomes less remarkable when equal number of labeled images has been accumulated. For example, in Figure 3 where 20 images are fed back by the user in each round, P10 (the precision among the top 10 retrieved images) using MRBIR is 0.726, and it is 0.769 using MRW. The improvement is 5.92%. In Figure 4 where 10 images are fed back by the user in each round, P10 using MRBIR is 0.758, and it is 0.794 using MRW. The improvement is 4.75%. In Figure 5 where 5 images are fed back by the user in each round, P10 using MRBIR is 0.784, and it is 0.805 using MRW. The improvement is 2.69%. The reason may be explained as follows: when fewer images are fed back in each round, we tend to obtain more positive examples in the 20 images that are finally accumulated. On the one hand, more positive examples help improve the performance MRBIR considerably [2]; on the other hand, they may cause the sampling probability of labeled examples to deviate from the actual distribution, so MRW does not benefit much from the reduction of feed back images in each round.

4.3 Response Time

In MRW, the E-step and M-step are repeated until a stable state is reached. In most cases, the number of iteration steps is no more than 10. In the initial retrieval stage, we only have to calculate the absorbing probabilities once since the random walks are not refined in the EM-like iterative procedure, so the processing speed is fast (0.532 seconds, Pentium 4 1.80GHz, 512M RAM). In relevance feedback, although we have to calculate the absorbing probabilities in each iteration, we have designed an algorithm similar to that in [26, 27], which helps to reduce the response time greatly (5.125 seconds). Therefore, MRW meets the need of real-time retrieval tasks.

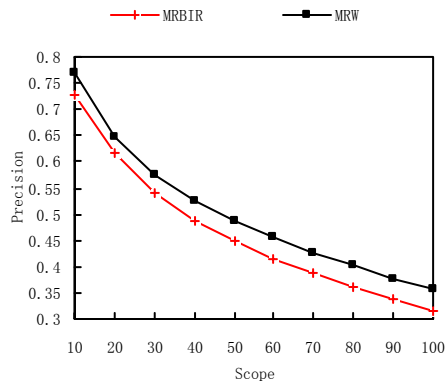


Figure 3. Comparison after the first round of relevance feedback with 20 images labeled by the user in each round.

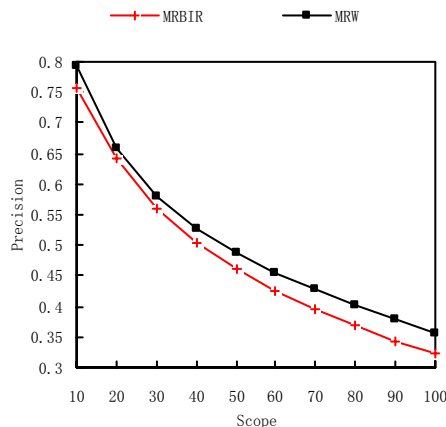


Figure 4. Comparison after the second round of relevance feedback with 10 images labeled by the user in each round.

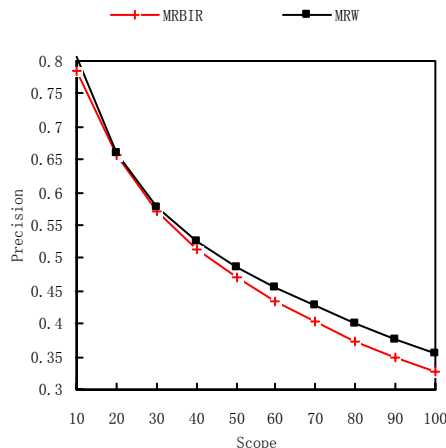


Figure 5. Comparison after the fourth round of relevance feedback with 5 images labeled by the user in each round.

5. CONCLUSION

In this paper, we have proposed a novel transductive method for CBIR named Multiple Random Walk. To construct the generative

models for images that are relevant and irrelevant to the query concept, we design two Markov random walks with positive and non-positive virtual absorbing boundaries. After normalization, the absorbing probabilities can be used as the initial estimation of the likelihood functions. Furthermore, an EM-like iterative procedure is designed to refine the random walks as well as to obtain the class priors. Finally, the estimation of the posterior probabilities are output by MRW and used in image retrieval.

In the initial retrieval stage where only one query image is provided by the user, the generative model for relevant images can be constructed by means of Markov random walk, while that for irrelevant images is constructed based on pair-wise distance between low-level features due to the lack of irrelevant images. In the EM-like iterative procedure, only the class priors are updated. Finally, all the unlabeled images in the database are ranked in descending order of their posterior probabilities of being relevant. In relevance feedback where both relevant and irrelevant images are provided by the user, we can construct both the positive and negative random walks using these labeled examples, thus obtain a more accurate estimation of the posterior probabilities. Systematic experimental results on a general-purpose image database consisting of 5,000 Corel images demonstrate the superiority of MRW over MRBIR.

6. REFERENCES

- [1] Bilmes, J.A. *A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*. Technical Report UMIACS-TR-97-021, International Computer Science Institute and U.C. Berkeley, 1998.
- [2] He, J., Li, M., Zhang, H.J., Tong, H., and Zhang, C. Manifold-ranking based image retrieval. *Proc. 12th ACM Int. Conf. on Multimedia*, pp. 9-16, 2004.
- [3] He, J., Li, M., Zhang, H.J., Tong, H., and Zhang, C. Mean version space: a new active learning method for content-based image retrieval. *Proc. 6th ACM SIGMM Int. Workshop on Multimedia Information Retrieval*, pp. 15-22, 2004.
- [4] Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R. Image indexing using color correlograms. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 762-768, 1997.
- [5] Jing, F., Li, M., Zhang, H.J., Zhang, B. An effective region-based image retrieval framework. *Proc. 10th ACM Int. Conf. on Multimedia*, pp. 456-465, 2002.
- [6] Kokare, M., Chatterji, B.N., and Biswas, P.K. Comparison of similarity metrics for texture image retrieval. *IEEE Conf. on Convergent Technologies for Asia-Pacific Region*, vol. 2, pp. 571-575, 2003.
- [7] Li, B., Chang, E., and Wu, C.T. DPF-a perceptual distance function for image retrieval. *Proc. IEEE Int. Conf. on Image Processing*, vol. 2, pp. 597-600, 2002.
- [8] Liu, F., and Picard, R.W. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 722-733, 1996.
- [9] Lovasz, L. Random walks on graphs: a survey. *Combinatorics Paul Erdos is Eighty*, vol. 2, pp. 1-46, 1993.

- [10] Mallat, S.G. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674-693, 1989.
- [11] Manjunath, B.S., and Ma, W.Y. Texture features for browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 837-842, 1996.
- [12] Pass, G. Comparing images using color coherence vectors. *Proc. 4th ACM Int. Conf. on Multimedia*, pp. 65-73, 1997.
- [13] Rocchio, J.J. Relevance feedback in information retrieval. *The SMART Retrieval System*, pp. 313-323, Prentice-Hall, Englewood Cliffs, NJ, 1971.
- [14] Rubner, Y., Tomasi, C., and Guibas, L. A metric for distributions with applications to image databases. *Proc. 6th IEEE Int. Conf. on Computer Vision*, pp. 59-66, 1998.
- [15] Rui, Y., Huang, T.S., Ortega, M., and Mehrotra, S. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Trans. Circuits and Systems for Video Technology*, vol. 8, pp. 644-655, 1998.
- [16] Schmid, C. A structured probabilistic model for recognition. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 485-490, 1999.
- [17] Schmid, C., and Mohr, R. Local grayvalue invariants for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 530-535, 1997.
- [18] Shi, J., and Malik, J. Normalized cuts and image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888-905, 2000.
- [19] Stricker, M., and Orengo, M. Similarity of color images. *Storage and Retrieval for Image and Video Databases, Proc. SPIE 2420*, pp 381-392, 1995.
- [20] Swain, M., and Ballard, D. Color indexing. *Int. Journal of Computer Vision*, 7(1): 11-32, 1991.
- [21] Szummer, M., and Jaakkola, T. Partially labeled classification with Markov random walks. *Neural Information Processing Systems*, 2002.
- [22] Tieu, K., and Viola, P. Boosting image retrieval. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 228-235, 2000.
- [23] Wang, J.Z., Wiederhold, G., Firschein, O., and Sha, X.W. Content-based image indexing and searching using Daubechies' wavelets. *Int. Journal of Digital Libraries*, vol. 1, no. 4, pp. 311-328, 1998.
- [24] Wu, Y., Tian, Q., and Huang, T. Discriminant-EM algorithm with application to image retrieval. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 155-162, 2000.
- [25] Zhang, L., Lin, F., and Zhang, B. Support vector machine learning for image retrieval. *Proc. IEEE Int. Conf. on Image Processing*, vol. 2, pp. 721-724, 2001.
- [26] Zhou, D., Bousquet, O., Lal, T.N., Weston, J., and Schölkopf, B. Learning with local and global consistency. *Neural Information Processing Systems*, 2003.
- [27] Zhou, D., Weston, J., Gretton, A., Bousquet, O., and Schölkopf, B. Ranking on data manifolds. *Neural Information Processing Systems*, 2003.
- [28] Zhou, X.S., Rui, Y., and Huang, T. Water-Filling: a novel way for image structural feature extraction. *Proc. IEEE Int. Conf. on Image Processing*, vol. 2, pp. 570-574, 1999.
- [29] Zhu, X., Ghahramani, Z., and Lafferty, J. Semi-supervised learning using Gaussian Fields and harmonic functions. *Proc 12th Int. Conf. on Machine Learning*, 2003.