Multimed Tools Appl (2006) 30: 131–147
DOI 10.1007/s11042-006-0023-5

Probabilistic semantic network-based image retrieval using MMM and relevance feedback

Mei-Ling Shyu · Shu-Ching Chen · Min Chen · Chengcui Zhang · Chi-Min Shu

Published online: 9 August 2006

Abstract The performance of content-based image retrieval (CBIR) systems is largely limited by the gap between the low-level features and high-level semantic concepts. In this paper, a probabilistic semantic network-based image retrieval framework using relevance feedback is proposed to bridge this gap, which not only takes into consideration the low-level image content features, but also learns high-level concepts from a set of training data, such as access frequencies and access patterns of the images. One of the distinct properties of our framework is that it exploits the structured description of visual contents as well as the relative affinity measurements among the images. Consequently, it provides the capability to bridge the gap between the low-level features and high-level concepts. Moreover, such high-level concepts can be learned off-line, and can be utilized and refined based on the user's specific interest during the on-line retrieval process. Our experimental

M.-L. Shyu (✉)
Department of Electrical and Computer Engineering,
University of Miami, Coral Gables, FL 33124, USA
e-mail: shyu@miami.edu

S.-C. Chen • M. Chen
Distributed Multimedia Information System Laboratory, School of Computing and Information Sciences,
Florida International University, Miami, FL 33199, USA

S.-C. Chen
e-mail: chens@cs.fiu.edu

M. Chen
e-mail:mchen005@cs.fiu.edu

C. Zhang
Department of Computer and Information Sciences,
University of Alabama at Birmingham, Birmingham, AL 35294, USA
e-mail: zhang@cis.uab.edu

C.-M. Shu
Department of Environmental and Safety Engineering,
National Yunlin University of Science and Technology, Yunlin, Taiwan, Republic of China
e-mail: shucm@pine.yuntech.edu.tw

results demonstrate that the proposed framework can effectively assist in retrieving more accurate results for user queries.

Keywords Content-based image retrieval · Probabilistic semantic network · MMM mechanism · Relevance feedback

1 Introduction

With the advance in image processing, information retrieval, and database management, content-based image retrieval (CBIR) has been actively studied in recent years, which resulted in a number of systems and techniques, both in the academic and commercial domains. For example, IBM's QBIC system [4] and Virage's VIR engine [21] are two most notable commercial image retrieval systems, while VisualSEEK [19], Metaseek [1], PhotoBook [12] are well-known academic image retrieval systems. In addition, a prototype content-based image retrieval system, PicHunter, was presented in [3]. Most of the existing CBIR systems rely on the use of the low-level image information, such as color [20], shape [24], and texture [7] features, to search for similar images from the databases. Queries are issued through query-by-image example (QBE).

However, the performance of CBIR systems is largely limited by the gap between the low-level features and high-level semantic concepts. To overcome this problem, recent studies in CBIR have focused on the approaches based on relevance feedback (RF), which tries to establish the link between the high-level concepts and low-level feature representations by modeling the user's subjective perception from the user's feedback [15]. Most of the previous relevance feedback research can be classified into two approaches: re-weighting and query point movement [6]. For instance, the query-point movement approach has been applied to the image retrieval systems, such as MARS [14] and MindReader [6]. The MARS system also implements a re-weighting method called the standard deviation method. However, those RF-based systems have two major limitations as follows.

1. These approaches estimate the ideal query parameters only from the low-level image features. Due to the limited power of the low-level features in representing the high-level semantics, it may not be effective in modeling users' perceptions. As a result, in many cases, the desired query results could not be achieved even after a large number of user interactions. To address this issue, a framework that performs relevance feedback on both the images' low-level features and the semantic contents represented by keywords was proposed [9]. In their work, a semantic network was constructed as a set of keywords linked to the images in the database. Though the retrieval accuracy is improved by using this approach, extra effort is required to label the images manually with the keywords.
2. Though the feedback information provided in each interaction contains certain high-level concepts, it is solely used to improve the current query results for a specific user. In other words, no mechanism is included in these systems to memorize or to accumulate the relevance feedback information to improve both the current query accuracy and the future system performance.

To overcome such limitations, in this paper, we propose a probabilistic semantic network-based image retrieval framework which employs both relevance feedback and

Markov Model Mediator (MMM) mechanism for image retrieval. The MMM mechanism adopts the Markov model framework and the concept of the mediators. The Markov model is one of the most powerful tools available for scientists and engineers to analyze the complicated systems, whereas a mediator is defined as a program to collect and combine information from one or more sources, and finally yield the resulting information [22]. Markov models have been used in many applications. Some well-known examples are Markov Random Field Models [5], and Hidden Markov Models (HMMs) [13]. Some research works have been done to integrate the Markov model into the field of image retrieval. Lin et al. [8] used a Markov model to combine the spatial and color information. In their approach, each image in the database is represented by a pseudo two-dimensional HMM in order to adequately capture both the spatial and chromatic information about that image. HMMs were also used to parse video data in [23]. In [11], by using a probabilistic framework, the HMM was employed to model the time series of the feature vectors of events and objects for semantic level indexing and retrieval. In our earlier studies, the MMM mechanism has been applied to multimedia database management [17] and document management on the World Wide Web (WWW) [16]. In our recent work, we applied MMM to content-based image retrieval and the preliminary results were presented in [18], where the MMM mechanism functions as both the searching engine and the image similarity arbitrator for image retrieval, whereas the RF process provides the information of a specific user's concepts for MMM to improve the query performance.

In this paper, we further improve our work in [18] by constructing a probabilistic semantic network for content-based image retrieval using MMM and relevance feedback. Our experimental results demonstrate that the proposed semantic network can provide a better representation of the relative affinity relationships among images in the database. In addition, the proposed framework can support both accumulative learning which is conducted off- and on-line instant learning for individual users. The main contributions are summarized as follows:

1. Different from the common approaches which try to capture the semantic content of an individual image (it is more difficult and most likely incomplete), a probabilistic semantic network is constructed in our framework to represent the semantic relationships among images. Such a network is useful because image retrieval is actually a process to explore the relationships between the query image and the other images in the database.
2. In contrast to the work proposed in [9], which requires extra manual effort in labeling the images, our framework provides the capabilities to accumulate the previous feedback information and automatically mine the semantic relationships among the images to construct and update the probabilistic semantic network. Therefore, instead of starting each query with the low-level features, the users' perceptions are gradually embedded into our framework to improve the initial query results.
3. In addition to supporting accumulative learning, the proposed framework also supports the query refinement for individual users in real-time. In particular, a temporary semantic subnetwork is extracted from the original semantic network and updated based on the current user's interests. For the sake of system efficiency and avoiding the bias caused by a single user, such update is conducted on the temporary semantic subnetwork only, without affecting the original semantic network. In the meanwhile, the individual user's feedback is collected continuously, and the update of the whole semantic network is triggered only when the number of accumulated feedbacks reaches a threshold. Such update is conducted off-line to enable accumulative learning while maintaining efficiency.

The remainder of this paper is organized as follows. Section 2 discusses the architecture and the details of the proposed framework. Experimental results are presented in Section 3 to demonstrate the effectiveness of this framework. A brief conclusion is given in Section 4.

2 Architecture of the framework

The architecture of the proposed framework is shown in figure 1. As can be seen from this figure, a training process is used to construct the semantic network off-line. Then image retrieval is conducted by utilizing both the low-level features and the semantic network, which are captured by two matrices in the MMM mechanism, namely the feature matrix $\mathcal{B}$ and the relative affinity matrix $\mathcal{A}$, respectively. A feedback process refines the current retrieval results by updating the temporary semantic subnetwork. In the meanwhile, the user feedback information is collected continuously as the training data for subsequent updates of the whole semantic network. The discussions are detailed in the following three subsections.

2.1 Semantic network

Assume N is the total number of images in the image database and $I = \{i_1, i_2, \ldots, i_N\}$ is the image set. The semantic network is modeled by the relative affinity matrix $\mathcal{A}$, where $\mathcal{A} = \{a_{m,n}\}(1 \leq m, n \leq N)$ denotes the probabilities of the semantic relationships among the images based on users' preferences, and the relationships of the images in the semantic network are represented by the sequences of the states (images) connected by transitions. Figure 2 shows an example of the semantic network, where the lines with zero probabilities are omitted. Table 1 shows the corresponding $\mathcal{A}$ matrix.

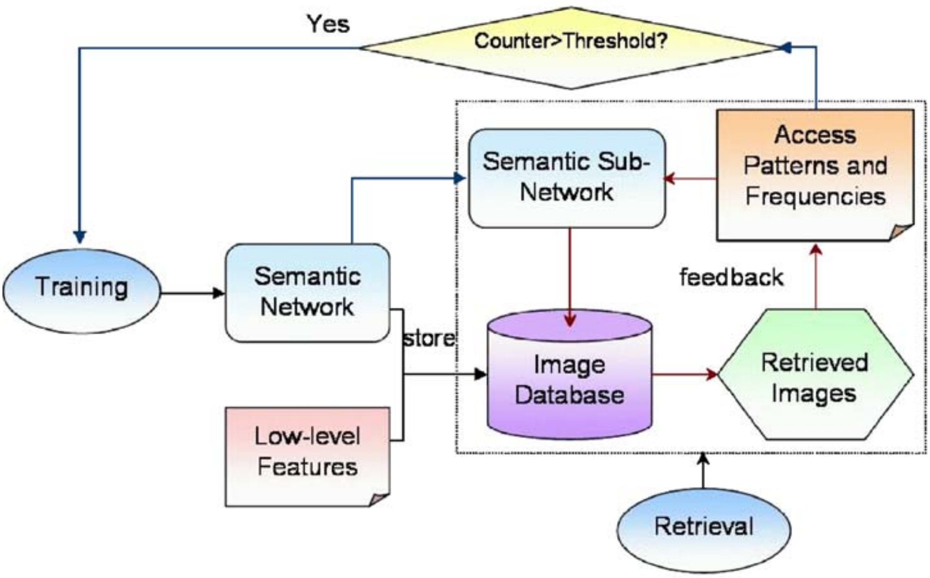

Fig. 1 Framework architecture

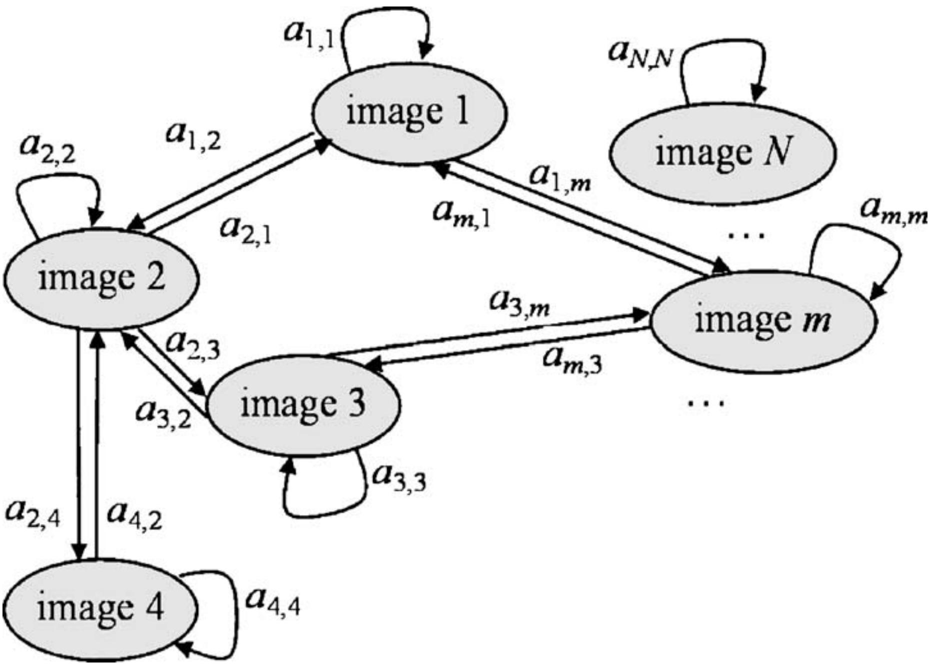

Fig. 2 Probabilistic semantic network

In the network, we define two different kinds of relationships between two images:

1. Directly related (R_D)
   $i_m\ R_D\ i_n \Leftrightarrow a_{m,n} \neq 0$ where $i_m, i_n \in I$, $a_{m,n} \in \mathcal{A}$
   For example: $i_1\ R_D\ i_2$, $i_2\ R_D\ i_3$, etc.
2. Indirectly related (R_I)
   $i_m\ R_I\ i_n \Leftrightarrow ((a_{m,n} = 0) \wedge (\ \exists i_x \in I \Rightarrow a_{m,x} \neq 0 \wedge a_{x,n} \neq 0))$ where $i_m, i_n, i_x \in I$, $a_{m,n}, a_{m,x}, a_{x,n} \in \mathcal{A}$, and $m \neq n$
   For example: $i_1\ R_I\ i_3$, $i_1\ R_I\ i_4$, etc.

In other words, R_D is the relationship between two directly linked images, while R_I exists between two images that are connected to a common image. For the purpose of constructing the semantic network, a set of training data is needed.

Table 1 The relative affinity matrix $\mathcal{A}$ of the example semantic network

	Img 1	Img 2	Img 3	Img 4	...	Img *m*	...	Img *N*
Img 1	$a_{1,1}$	$a_{1,2}$	0	0	...	$a_{1,m}$	...	0
Img 2	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{2,4}$	...	0	...	0
Img 3	0	$a_{3,2}$	$a_{3,3}$	0	...	$a_{3,m}$	...	0
Img 4	0	$a_{4,2}$	0	$a_{4,4}$	...	0	...	0
...	...	...	...	...	...	...	...	...
Img *m*	$a_{m,1}$	0	$a_{m,3}$	0	...	$a_{m,m}$	...	0
...	...	...	...	...	...	...	...	...
Img *N*	0	0	0	0	...	0	...	$a_{N,N}$

Fig. 3 The interface of the training system

2.1.1 Training data set

In order to construct the semantic network, a training data set is required to generate the probabilistic semantic relationships among the images. The source of the training data set is actually the log of user access patterns and access frequencies on the image data. Access patterns, in brief, denote the co-occurrence relationships among the images accessed by the user queries, while access frequencies denote how often each query was issued by the users.

A training system is implemented for this framework to collect the user access patterns and access frequencies. Figure 3 shows the system interface. During the training process, a group of users were asked to randomly issue queries and select positive and negative examples from the results for each query. The positive examples selected in each query are said to have the co-occurrence relationships with each other. Intuitively, they are semantically related. In addition, the more frequently two images are accessed together, the more closely they are related. The formal definition regarding the training data set is given as follows:

Definition 1 Assume N is the total number of images in the image database and a set of queries $Q = \{q_1, q_2, ..., q_{nq}\}$ were issued to the database in a period of time. The training data set consists of the following information:

– Let $use_{m,k}$ denote the usage pattern of image m with respect to query q_k per time period, where the value of $use_{m,k}$ is 1 when m is accessed by q_k and zero otherwise.
– The value of $access_k$ denotes the access frequency of query q_k per time period.

Table 2 gives some example queries issued to the image database with their corresponding access frequencies. The access patterns of the three sample images in figure 4

Table 2 The query access frequencies ($access_k$) and access patterns ($use_{k,m}$) of the sample images

Query	$access_k$	Img 1	Img 2	Img 3	...
q_1	4	1	1	0	...
q_2	1	0	1	1	...
q_3	$access_3$	0	0	0	...
...	...	...	...	...	...

versus the example queries are also shown in Table 2. In this table, the entry (k, m) = 1 indicates that the m^{th} image is accessed by query q_k. For example, suppose q_1 is a user-issued query related to retrieving images containing country house scenes. Img 1 and Img 2 are accessed together in q_1, with their corresponding entries in the access pattern matrix having the value 1. Let q_2 denote a query related to the concept of '*red flowers*,' then Img 2 and Img 3 will probably be accessed together in this query. However, since most of the users regard Img 2 as a country house scene more than a '*red flowers*' scene, the access frequency of q_1 is larger than that of q_2. Consequently, after the system training, Img 1 is more likely to be retrieved than Img 3, given that Img 2 is selected as the query image. Thus, the users' subjective concepts about the images are captured by the pair of user access patterns and user access frequencies.

2.1.2 Construct the semantic network

Based on the information in the training data set, we can capture the semantic relationships among the images in the database and construct the semantic network. In order to capture the semantic relationships among all the images, an assisting matrix *AFF* is defined, which is constructed by having the $aff_{m,n}$ to be the relative affinity relationship between two images m and n using the following definition.

Fig. 4 Three sample images (Img 1–Img 3)

Definition 2 The relative affinity measurement ($aff_{m,n}$) between two images m and $n(1 \leq m, n \leq N)$ indicates how frequently these two images are accessed together, where

$$aff_{m,n} = \sum_{k=1}^{nq} use_{m,k} \times use_{n,k} \times access_k \tag{1}$$

Let N be the total number of the images in the database. The matrix $\mathcal{A}$ is initialized by having $a_{m,n}$ be the element in the $(m,n)^{th}$ entry in $\mathcal{A}$, where

$$a_{m,n} = 1/N \tag{2}$$

Then $\mathcal{A}$ is constructed via the following equation.

$$a_{m,n} = \begin{cases} \dfrac{aff_{m,n}}{\sum\limits_{k \in d} aff_{m,k}} & \text{if } \sum\limits_{k \in d} aff_{m,k} \neq 0 \\ a_{m,n} & \text{otherwise.} \end{cases} \tag{3}$$

From the above equations, we can see that each row m in the $\mathcal{A}$ matrix represents the R_D relationship between the image i_m and the other images, while the whole $\mathcal{A}$ matrix contains the information of both the two kinds of relationships among the images in the database. The values in $\mathcal{A}$ are then used to construct the semantic network.

For the sake of efficiency, during a training period, the training system only collects all the user access patterns. Once the number of records reaches a threshold (e.g., 500), the update of $\mathcal{A}$ matrix is triggered automatically. All the computations are done off-line. Moreover, instead of using the whole $\mathcal{A}$ matrix, in the retrieval process we only utilize the R_D relationship between the query image and the other images together with the low-level features to generate the initial query results. In other words, for a specific query image $i_m(i_m \in I)$, only the m^{th} row in matrix $\mathcal{A}$ (denoted as A_m) is applied in order to reduce the computational load and I/O cost.

2.1.3 Refine the semantic subnetwork on-line

As mentioned above, the relative affinity matrix $\mathcal{A}$ is obtained based on the feedbacks provided by various users on different kinds of queries. Therefore, matrix $\mathcal{A}$ represents the general user concepts and can help to achieve better query results. However, in the retrieval process, different users may have different concepts about the query images. Therefore, in addition to supporting accumulative learning, the system also needs to support the instant learning which enables the query refinement for individual users on the fly. Moreover, it is most likely that the query images chosen by the users have no existing R_D or R_I relationships in the current semantic network. In this subsection, a refinement method for the semantic subnetwork is proposed to solve these problems based on users' feedback. Note that as discussed earlier, such refinement is conducted on the temporary subnetwork only.

For a query image $i_m(i_m \in I)$, the user can choose to accept the initial query results obtained by using the general user concepts, or to provide the feedback via indicating the positive and negative examples. The access patterns can be obtained based on these positive and negative examples, as mentioned in the previous subsection. Such access patterns are then used to update the $\mathcal{A}$ matrix to further improve the initial query results. More importantly, the user specified R_D relationship $\mathcal{A}_m^{'} = [a_k^{'}](1 \leq k \leq N)$ between i_m and other images can be obtained using Eqs. 1 and 3 with parameter m fixed and n varied from 1 to

Table 3 Capture R_D and R_I relationships between i_m and other images

1. Obtain non-zero items $[a'_{n_1}, a'_{n_2}, ..., a'_{n_T}]$ in A'_m, where T is the total number of images which have non-zero R_D relationships with $i_m(1 \le T \le N)$.
2. For each $a'_{n_i}(1 \le i \le T)$, get its corresponding $A_{n_i} = [a_{n_i,j}](1 \le j \le N)$ from matrix $\mathcal{A}$
3. Normalize each A_{n_i} as:
   if (max $(A_{n_i}) \neq \min(A_{n_i})$)
   $a_{n_i,j} = (a_{n_i,j} - \min(A_{n_i}))/(\max(A_{n_i}) - \min(A_{n_i}))$
   else
   $a_{n_i,j} = \begin{cases} 1 & \text{if } n_i = j \\ 0 & \text{otherwise} \end{cases}$
4. Define a matrix $P = \{p_{i,j}\}(1 \le i \le T, 1 \le j \le N)$
 For j = 1 to N
  For i = 1 to T
   $p_{i,j} = a'_{n_i} \times a_{n_i,j}$
  end
  v_j = the maximal value of the j^{th} column of P
 end
5. Normalize the vector $V = \{v_j\}(1 \le j \le N)$ as:
$$v_j = \frac{v_j}{\sum\limits_{1 \le i \le N} v_i}$$

N. Let vector $V_m = [v_j](1 \le j \le N)$ denote the information of both R_D and R_I relationships between i_m and other images. Table 3 shows the steps to calculate it. The idea is quite straightforward. As we know, each a'_{n_i} in $\mathcal{A}'_m$ represents the R_D probability from i_m to i_{n_i}, while each $a_{n_i,j}$ in $\mathcal{A}_{n_i}$ denotes the R_D probability from i_{n_i} to i_j. Therefore, the R_D probability from i_m to i_j connected by a common image i_{n_i} can be obtained via $a'_{n_i} \times a_{n_i,j}$. A pair of images can be indirectly related to each other via multiple paths in the semantic network. For example, Img 1 and Img 3 are indirectly related via two different ways — one through Img 2 and one through Img m (as shown in figure 2). In such a situation, the maximal probability is kept since this maximal probability indicates the actual degree of their semantic relationship. It is worth mentioning that normally the set of non-zero items in $\mathcal{A}'_m$ is quite small, so the algorithm is efficient in terms of space and time.

2.2 Low-level feature set

In this framework, currently we consider two kinds of low-level features: color information and texture information, for the images in the image database. Since the color feature is closely associated with image scenes and it is more robust to changes due to scaling, orientation, perspective and occlusion of images, it is the most widely used visual feature in image retrieval [10]. In our CBIR system, color information is obtained for each image from its HSV color space. The HSV color space is chosen for two reasons. First, it is perceptual, which makes HSV a proven color space particularly amenable to color image analysis [2]. Secondly, the benchmark results in [10] showed that the color histogram in the HSV color space performs the best. For the texture features, one-level wavelet transformation using Daubechies wavelets are used because it is proven to be suitable for image analysis. Since our focus is to evaluate the performance of the MMM retrieval mechanism rather than to explore the most appropriate features for image retrieval, in our study, each image has a feature vector of only 19 elements. Within the 19 features, 13 are for color descriptors and six are texture descriptors. The color features considered are 'black' (bl), 'gray' (gy), 'white' (w), 'red,' 'red-yellow' (ry), 'yellow' (y), 'yellow-green' (yg), 'green' (g), 'green-blue' (gb), 'blue' (b), 'blue-purple' (bp), 'purple' (p) and 'purple-red' (pr) according to the combinations of different ranges of the hue (H), saturation (S), and the intensity values (V). Colors with the number of pixels less than 5% of the total number of pixels are regarded as non-important and the corresponding positions in the feature vector have the value 0. Otherwise, we put the corresponding percentage of that color component to that position. For the texture features, one-level wavelet transformation using Daubechies wavelets are used to generate the horizontal detail sub-image, the vertical detail sub-image, and the diagonal detail sub-image. For the wavelet coefficients in each of the above three subbands, the mean and variance values are extracted. A temporary matrix ($\mathcal{BB}$) is created to contain all the image feature vectors. The rows of matrix $\mathcal{BB}$ are all the distinct images, while its columns contain all the distinct features. The value in the $(p,q)^{th}$ entry is greater than zero if feature q appears in image p, and zero otherwise. Then the feature matrix $\mathcal{B}$ can be obtained via normalizing $\mathcal{BB}$ per row. We consider that the color and texture information are of equal importance, such that the sum of the feature values of the color features should be equal to that of the texture features. In other words, the sum of the values that the features are observed from a given image should be 1, with 0.5 for the sum of all the color features and 0.5 for the sum of all the texture features. It is worth mentioning that our mechanism is flexible in the sense that the $\mathcal{B}$ matrix can be constructed by using any normalized vector-based image feature set.

2.3 Stochastic process for information retrieval

In this subsection, a stochastic retrieval process is defined to calculate the edge weights among the images utilizing both the low-level features and the semantic network. Assume N is the total number of images in the database, and the features of the query image q is denoted as $\{o_1, o_2, ..., o_T\}$, where T is the total number of non-zero features of the query image q. In our case, $1 \leq T \leq 19$ since there are 19 features in total.

Definition 3 $W_t(i)$ is defined as the edge weight from image i to q at the evaluation of the t^{th} feature (o_t) in the query, where $1 \leq i \leq N$ and $1 \leq t \leq T$.

Table 4 Image retrieval steps using the proposed framework

1. Given the query image q, obtain its feature vector $\{o_1, o_2, ..., o_T\}$, where T is the total number of non-zero features of the query image q.
2. Upon the first feature o_1, calculate $W_1(i)$ according to Eq. 4.
3. To generate the initial query results, set $a_{q,i}$ to be the value of $(q,i)^{th}$ entry in matrix $\mathcal{A}$. Otherwise, based on the user's feedback, calculate vector V_q by using the algorithm presented in Table 3 and let $a_{q,i}$ equal to v_i, the i^{th} entry in V_q.
4. Move on to calculate $W_2(i)$ according to Eq. 5.
5. Continue to calculate the next values for the W vector until all the features in the query have been taken care of.
6. Upon each non-zero feature in the query image, we can obtain a vector $W_t(i)(1 \leq t \leq T)$. We then sum up each value at the same position in the vectors $W_1(i)$, $W_2(i)$, ..., $W_T(i)$. Namely, $sumW_T(i) = \sum_T W_t(i)$ is calculated.
7. Find the candidate images by sorting their corresponding values in $sumW_T(i)$. The bigger the value is, the stronger the relationship exists between the candidate image and the query image.

Based on the definition, the retrieval algorithm is given as follows. At $t = 1$,

$$W_1(i) = (1 - |b_i(o_1) - b_q(o_1)|/b_q(o_1)) \quad (4)$$

The value of $W_{t+1}(i)$, where $1 \leq t \leq T - 1$, is calculated by using the value of $W_t(i)$.

$$W_{t+1}(i) = W_t(i)a_{q,i}(1 - |b_i(o_{t+1}) - b_q(o_{t+1})|/b_q(o_{t+1})) \quad (5)$$

As we mentioned before, $\mathcal{A}$ represents the relative affinity measures of the semantic relationships among the images in the probabilistic semantic network and $\mathcal{B}$ contains the low-level features. To generate the initial query results, the value of $a_{q,i}$ from matrix $\mathcal{A}$ is used. Once the user provides the feedback, a vector V_q is calculated using the algorithm presented in Table 3. Then $a_{q,i} = v_i(v_i \in V_q)$ is applied in Eq. 5. The stochastic process for image retrieval by using the dynamic programming algorithm is shown in Table 4.

3 Experiments

In this paper, we present a framework in which the semantic network and low-level features can be integrated seamlessly into the image retrieval process to improve the query results. In this section, the experimental results are presented to demonstrate the effectiveness of this framework.

3.1 Experimental image database system

In the experiments, 10,000 color images from the Corel image library with more than 70 categories, such as people, animal, etc., are used. In order to avoid the bias and to capture the general users' perceptions, the training process was performed by a group of ten university students, who are not involved in the design and development of our framework and have no knowledge of the image content in the database. Currently, we have collected 1,400 user access patterns through the training system, which covered less than half of the images in the database. The $\mathcal{A}$ matrix and the semantic network are constructed according to the algorithms presented in Section 2. For the low-level image features, the color and texture features of the images are considered and the $\mathcal{B}$ matrix is obtained using the procedures illustrated in Section 2. The constructions of these matrices can be performed off-line.

To test the retrieval performance and efficiency of the proposed mechanism, 80 randomly chosen images belonging to five distinct categories were used as the query images. Table 5 lists the descriptions for each category as well as the number of query images selected from each category.

For a given query image issued by a user, the proposed stochastic process is conducted to dynamically find the matching images for the user's query. The similarity scores of the

Table 5 The category distribution of the query image set

Category	Explanation	Number of query images
Landscape	Land, Sky, Mountain	16
Flower	Flower	16
Animal	Elephant, Panther, Tiger	16
Vehicle	Car, Bus, Plane, Ship	16
Human	Human	16

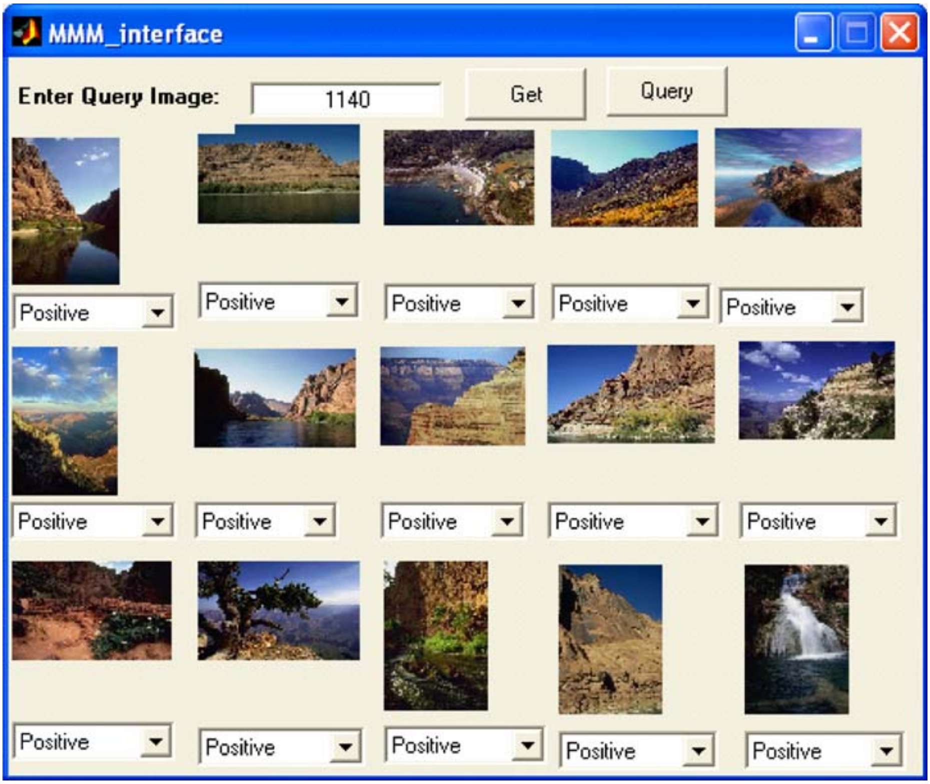

Fig. 5 The snapshot of a query-by-image example

images with respect to the certain query image are determined by the values in the resulting $sumW_T$ vectors according to the rules described in Table 4. Figure 5 gives a query-by-image example, in which the retrieved images are ranked and displayed in the descending order of their similarity scores from the top left to the bottom right, with the upper leftmost image being the query image. In this example, the query image belongs to the 'Landscape' category. As can be seen from this figure, the perceptions contained in these returned images are quite similar and the ranking is reasonably good.

In order to demonstrate the performance improvement and the flexibility of our model, we use the accuracy–scope curve to compare the performance of our proposed mechanism with a common relevance feedback method. In the accuracy–scope curve, the scope specifies the number of images returned to the users and the accuracy is defined as the percentage of the retrieved images that are semantically related to the query image.

3.2 Performance comparison

In our experiments, we compare the overall performance of our proposed MMM mechanism with the relevance feedback method (RF) proposed in [15] in the absence of the information of user access patterns and access frequencies. The RF method proposed in [15] conducts the query refinement based on re-weighting the low-level image features (matrix $\mathcal{B}$) alone. It is worth mentioning that in fact any normalized vector-based image

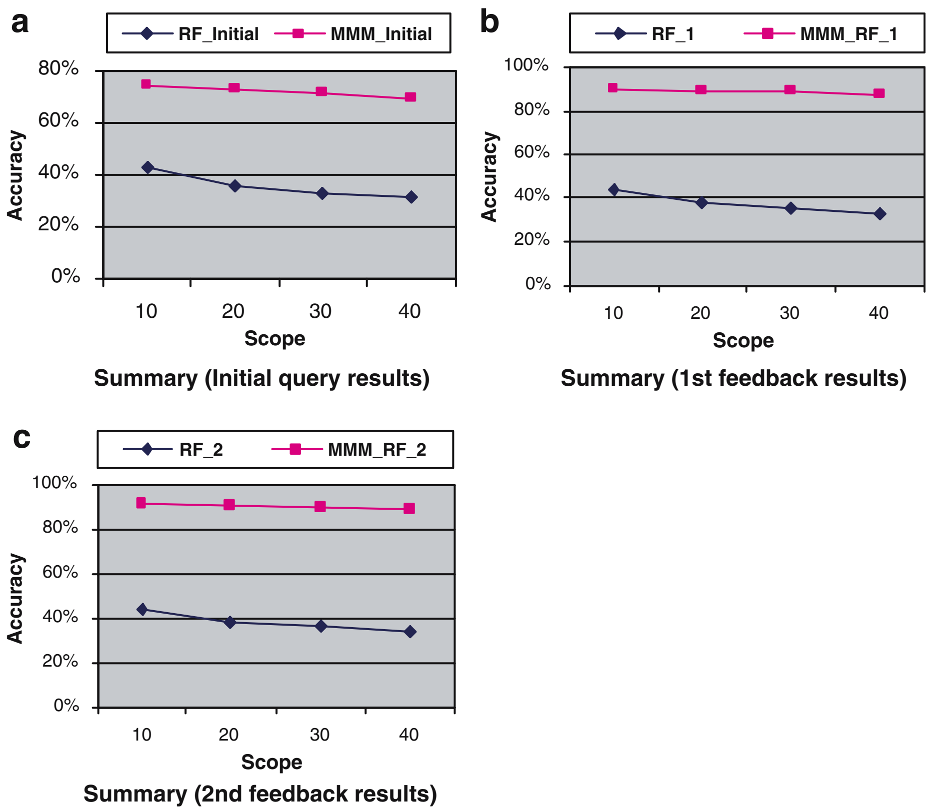

Fig. 6 Performance comparison

feature set can be plugged into the matrix $\mathcal{B}$. Figure 6 shows the curves for the average accuracy values of our CBIR system and the RF CBIR system, respectively. In figure 6a, 'MMM_Initial' and 'RF_initial' indicate the accuracy values of the MMM mechanism and the RF method at the initial retrieval time, respectively. The 'MMM_RF_1(2)' and the 'RF_1(2)' in figures 6b,c represent the accuracy values of the two methods after the first and the second rounds of user relevance feedback. The results in figure 6 are calculated using the averages of all the 80 query images. It can be easily observed that our proposed

Table 6 Accuracy and efficiency comparison between Relevance Feedback method and our proposed framework

Category	Relevance feedback			Proposed framework		
	Feedback number	Feedbacks per image	Accuracy (%)	Feedback number	Feedbacks per image	Accuracy (%)
Landscape	48	3	55.3	21	1.3	61.3
Flower	48	3	44.7	23	1.4	73.4
Animal	48	3	48.8	20	1.3	81.6
Vehicle	48	3	23.8	44	2.8	74.4
Human	48	3	26.9	33	2.1	75.0
Summary	240	3	39.9	141	1.8	73.1

method outperforms the RF method for the various numbers of images retrieved at each iteration. This proves that the use of the user access patterns and access frequencies obtained from the off-line training process can capture the subjective aspects of the user concepts. As another observation, the proposed method and the RF method share the same trend, which implies that the more iterations of user feedback, the higher accuracy they can achieve.

Table 6 lists the number of user feedback iterations observed in the RF method and the proposed method for each image category. For example, the number of query images in the 'Landscape' category is 16, and the numbers of user feedback iterations observed for those 16 images are 48 and 21 for the RF method and the proposed method, respectively. Thus, the number of feedback iterations per image is 48/16 = 3 for the RF method; while it is 1.3 for the proposed method. As can be seen from this table, the proposed method can achieve better retrieval performance even by using a smaller number of feedback iterations than that of the RF method in all five categories.

4 Conclusion

In this paper, a review of the recent efforts and techniques in CBIR is given, followed by the discussion of the current problems in the CBIR systems from the concern of lacking the mapping between the high-level concepts and the low-level features. Although Relevance Feedback (RF) has been proposed to bridge the gap, the performance is limited by the insufficient power of the low-level features in representing the high-level concepts. In addition, the users are required to take heavy responsibilities during the retrieval process to provide feedbacks in several iterations. The useful information contained in the user feedback is employed to improve the current query results only, without being further utilized to boost the system performance. In response to these issues, a probabilistic semantic network-based image retrieval framework using both relevance feedback and the Markov Model Mediator (MMM) mechanism is proposed. As a result, the semantic network and the low-level features are seamlessly utilized to achieve a higher retrieval accuracy. In addition, the access patterns and access frequencies from the previous feedbacks are accumulated to refine the probabilistic semantic network. Our experimental results demonstrate the effectiveness and efficiency of our framework for image retrieval.

Acknowledgments For Mei-Ling Shyu, this research was supported in part by NSF ITR (Medium) IIS-0325260. For Shu-Ching Chen, this research was supported in part by NSF HRD-0317692 and NSF EIA-0220562. For Chengcui Zhang, this research was supported in part by UAB Faculty Development Award, SBE-0245090 and the UAB ADVANCE program of the Office for the Advancement of Women in Science and Engineering.

References

1. Beigi M, Benitez A, Chang S-F (1998) Metaseek: a content-based meta search engine for images. In: Proceedings of IS&T/SPIE Storage and Retrieval for Image and Video Databases. San Jose, California, pp 118–128
2. Cheng HD, Sun Y (2001, December) A hierarchical approach to color image segmentation using homogeneity. IEEE Trans Image Process 9(12):2071–2082
3. Cox IJ, Miller ML, Minka TP, Papathornas TV, Yianilos PN (2000, January) The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. IEEE Trans Image Process 9(1):20–37

4. Flickner M et al (1995, September) Query by image and video content: the QBIC system. IEEE Computer 28(9):23–31
5. Frank O, Strauss D (1986) Markov graphs. J Am Stat Assoc 81:832–842
6. Ishikawa Y, Subramanya R, Faloutsos C (1998) Mindreader: query databases through multiple examples. In: Proceedings of the 24th VLDB Conference (VLDB'98). New York, USA, pp 218–227
7. Kaplan LM et al (1998) Fast texture database retrieval using extended fractal features. In: Proceedings of IS&T/SPIE Conference on Storage and Retrieval for Media Databases, pp 162–173
8. Lin HC et al (1997) Color image retrieval based on hidden Markov models. IEEE Trans Image Process 6(2):332–339
9. Lu Y, Zhang HJ, Liu W, Hu C (2003) Joint semantics and feature based image retrieval using relevance feedback. IEEE Trans Multimedia 5(3):339–347
10. Ma W-Y, Zhang HJ (1999) Content-based image indexing and retrieval. Handbook of multimedia computing, Chapter 13 CRC
11. Naphade MR, Huang TS (2001, March) A probabilistic framework for semantic video indexing, filtering and retrieval. IEEE Trans Multimedia 3(1):141–151
12. Pentland A, Picard RW, Sclaroff S (1994) Photobook: tools for content-based manipulation of image databases. In: Proceedings of IS&T/SPIE Storage and Retrieval for Image and Video Databases II, Vol. 2185, pp. 34–47
13. Rabiner LR, Huang BH (1986, January) An introduction to hidden Markov models. IEEE ASSP Mag 3(1):4–16
14. Rui Y, Huang TS, Mehrotra S (1997) Content-based image retrieval with relevance feedback in MARS. In: Proceedings of the 1997 International Conference on Image Processing (ICIP'97) (3-Volumn Set), pp 815–818
15. Rui Y, Huang TS, Mehrotra S (1998) Relevance feedback: a power tool for interactive content-based image retrieval. IEEE Trans on Circuit and Video Technology, Special Issue on Segmentation, Description, and Retrieval of Video Content, 18(5):644–655
16. Shyu M-L, Chen S-C, Shu C-M (2000, October) Affinity-based probabilistic reasoning and document clustering on the WWW. In: Proceedings of the 24^{th} IEEE Computer Society International Computer Software and Applications Conference (COMPSAC), Taipei, Taiwan, pp 149–154
17. Shyu M-L, Chen S-C, Kashyap RL (2000) A probabilistic-based mechanism for video database management systems. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME'00), New York, USA, pp 467–470
18. Shyu M-L, Chen S-C, Chen M, Zhang C, Sarinnapakorn K (2003) Image database retrieval utilizing affinity relationship. In: Proceedings of the first ACM International Workshop on Multimedia Databases (ACM MMDB'03), New Orleans, Louisiana, pp 78–85
19. Smith JR, Chang SF (1996, November) VisualSEEK: a fully automated content-based image query system. In: Proceedings of ACM International Conference on Multimedia, Boston, USA, pp 87–98
20. Stehling RO, Nascimento MA, Falcao AX (2000) On shapes of colors for content-based image retrieval. In: ACM International Workshop on Multimedia Information Retrieval (ACM MIR'00), Los Angeles, California, USA pp 171–174
21. Virage. http://www.virage.com
22. Wiederhold G (1992, March) Mediators in the architecture of future information systems. IEEE Computers 38 – 49
23. Wolf W (1997, April) Hidden Markov model parsing of video programs. In: Proceedings of the International Conference of Acoustics, Speech and Signal Processing. Munich, Germany, pp 2609–2611
24. Zhang DS, Lu G (2002, August) Generic Fourier descriptors for shape-based image retrieval. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME'02). Lausanne, Switzerland, pp 425–428

Dr. Mei-Ling Shyu received her Ph.D. degree from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN in 1999, and three Master's degrees from Computer Science, Electrical Engineering, and Restaurant, Hotel, Institutional, and Tourism Management from Purdue University. She has been an Associate Professor in the Department of Electrical and Computer Engineering (ECE) at the University of Miami (UM), Coral Gables, FL, since June 2005. Prior to that, she was an Assistant Professor in ECE at UM dating from January 2000. Her research interests include data mining, multimedia database systems, multimedia networking, and database systems. She has authored and co-authored more than 120 technical papers published in various prestigious journals, refereed conference/symposium/workshop proceedings, and book chapters. She is/was the guest editor of several journal special issues.

Dr. Shu-Ching Chen received his Ph.D. from the School of Electrical and Computer Engineering at Purdue University, West Lafayette, IN, USA in December, 1998. He also received Master's degrees in Computer Science, Electrical Engineering, and Civil Engineering from Purdue University. He has been an Associate Professor in the School of Computing and Information Sciences (SCIS), Florida International University (FIU) since August, 2004. Prior to that, he was an Assistant Professor in SCIS at FIU dating from August, 1999. His main research interests include distributed multimedia database systems and multimedia data mining. Dr. Chen has authored and co-authored more than 140 research papers in journals, refereed conference/symposium/workshop proceedings, and book chapters. In 2005, he was awarded the IEEE Systems, Man, and Cybernetics Society's Outstanding Contribution Award. He was also awarded a University Outstanding Faculty Research Award from FIU in 2004, Outstanding Faculty Service Award from SCIS in 2004 and Outstanding Faculty Research Award from SCIS in 2002.

Min Chen received her bachelor's degree in Electrical Engineering from Zhejiang University in China. She is currently a Ph.D. candidate in the School of Computing and Information Sciences at Florida International University (FIU). Her research interests include distributed multimedia database systems, image and video database retrieval, and multimedia data mining. She has authored and co-authored 17 technical papers published in various prestigious journals, refereed conference/workshop proceedings and book chapters. She is the recipient of several awards, including a Presidential Fellowship and the Best Graduate Student Research Award from FIU.

Dr. Chengcui Zhang is an Assistant Professor of Computer and Information Science at University of Alabama at Birmingham (UAB) since August, 2004. She received her Ph.D. from the School of Computer Science at Florida International University, Miami, FL, USA in August, 2004. She also received her bachelor and master degrees in Computer Science from Zhejiang University in China. Her research interests include multimedia databases, multimedia data mining, image and video database retrieval, and GIS data filtering. She is the recipient of several awards, including the UAB ADVANCE Junior Faculty Research Award from the National Science Foundation, UAB Faculty Development Award, and the Presidential Fellowship and the Best Graduate Student Research Award at FIU.

Dr. Chi-Min Shu received his B.S. degree in 1982 from Tunghai University. In 1987 and 1990, he obtained his M.S. and Ph.D. degrees from UMR in Chemical Engineering. Currently, he is a technical advisor for Unity Chemical Co., Ltd., a director of Center for Industry-Academia Collaboration, and a professor at NYUST. In 1990–1993, he worked at Bechtel Corporation as a process engineer. His research focuses on process safety, including thermal hazard analysis, runaway reaction, flammability studies, quantitative risk assessment, and planning of emergency response.