

Contact author:

Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33124-0640
USA
Email: shyu@miami.edu
Telephone: (O) 305-284-5566
(Fax) 305-284-4044

Topic area: Content-Based Image Retrieval

Title: A Stochastic Content-Based Image Retrieval Mechanism

Authors: Mei-Ling Shyu, Shu-Ching Chen, and Chengcui Zhang

Topics relevant to the chapter:

Special Purpose Database Technologies – Multimedia Databases: images, streams, etc.
Information Retrieval with Database Systems.

A Stochastic Content-Based Image Retrieval Mechanism

Mei-Ling Shyu^{1*}, Shu-Ching Chen², Chengcui Zhang²

¹Department of Electrical and Computer Engineering, University of Miami
Coral Gables, FL, USA
shyu@miami.edu

²Distributed Multimedia Information System Laboratory
School of Computer Science, Florida International University
Miami, FL, USA
{chens, czhang02}@cs.fiu.edu

Abstract

Multimedia information, typically image information, is growing rapidly across the Internet and elsewhere. To keep pace with the increasing volumes of image information, new techniques need to be investigated to retrieve images intelligently and efficiently. Content-based image retrieval (CBIR) is always a challenging task. In this chapter, a stochastic mechanism, called Markov Model Mediator (MMM), is used to facilitate the searching and retrieval process for content-based image retrieval, which serves as the retrieval engine of the CBIR systems and uses stochastic-based similarity measures. Different from the common methods, our stochastic mechanism carries out the searching and similarity computing process dynamically, taking into consideration not only the image content features but also other characteristics of images such as their access frequencies and access patterns. Our experimental results demonstrate that the MMM mechanism together with the stochastic process can assist in retrieving more accurate results for user queries.

Keywords: Content-based image retrieval, Markov model, stochastic process, similarity measures.

* To whom correspondence should be directed (shyu@miami.edu).

1. Introduction

The availability of today's digital devices and techniques offers people more opportunities than ever to create their own digital images. Moreover, Internet has become the biggest platform to get, distribute and exchange digital image data. The rapid increase in the amount of image data and the inefficiency of traditional text-based image retrieval have created great demands for new approaches in image retrieval. As a consequence of such fast growth of digital image databases, the development of efficient search mechanisms has become more and more important. Currently, Content-Based Image Retrieval (CBIR) emerges and dedicates to tackling such difficulties. CBIR is an active research area where the image retrieval queries are based on the content of multimedia data.

In contrast to the text-based approach, CBIR operates on a totally different principle, i.e., to retrieve the stored images from a collection of images by comparing the features that were automatically extracted from the images themselves. Content-based image retrieval involves a matching process between a query image and the images stored in the database. The first step of the process involves extracting a feature vector for the unique characteristics of each image. The features used for retrieval can be either primitive or semantic, but the extraction process must be automatic. A quantified similarity value between two images is obtained by comparing their feature vectors. The commonly used image features include color, shape and texture. Queries are issued through query by image example (QBE), which can either be provided or constructed by users, or randomly selected from the image database. A lot of research work has been done, which resulted in a number of systems and techniques, both in the academic and commercial domains. For example, IBM's QBIC system [Faloutsos94, Flickner95] and Virage's VIR engine [Virage] are two most notable commercial image retrieval systems, while VisualSEEk [Smith96] and PhotoBook [Pentland94] are well-known academic image retrieval systems.

For large image collections, traditional retrieval methods such as sequential searching do not work well since it is time expensive, and tends to ignore the relationships among all images by only considering the relationship between the query image and a single image in the database. Various kinds of data structures, approaches and techniques have been proposed to manage image databases and hasten the retrieval process.

The first aim of this chapter is to take an overview of the currently available content-based image retrieval (CBIR) systems. Then, with the focus on the searching process, we present a conceptually coherent framework that adopts a stochastic mechanism called Markov Model Mediator (MMM) for the content-based image retrieval problem. With an explicit model of image query patterns, given the target image, the proposed framework carries out the searching and similarity computing process dynamically by taking into consideration not only the image content features but also their access frequencies and access patterns.

The remainder of this paper is organized as follows. Section 2 gives the literature review as well as the motivations of the proposed mechanism. Section 3 gives the key components of the MMM mechanism and introduces the stochastic process for information retrieval. Section 4 presents our experiments in applying the MMM mechanism to content-based image retrieval. The experimental results are also provided in this section. In Section 5, the future trends are discussed. A brief conclusion is given in Section 6.

2. Background

The objective of a CBIR system is to offer the user an efficient way in finding and retrieving those images that are qualified for the matching criteria of the users' queries from the database. Most of the existing CBIR systems retrieve images in the following manner. First, they build the indexes based on the low-level features such as color, texture and shape for the images in the database. The

corresponding indexes of a query image are also generated upon the time the query is issued. Second, they search through the whole database and measure the similarity of each image to the query image. Finally, the results are returned in a sorted order of the similarity matching level.

Lots of approaches for retrieving images on the basis of color similarity have been described in the literature, but most of them are actually variations of the same basic idea. The most commonly used matching technique is histogram intersection [Li00]. Variants of this technique are now used in a big proportion of the current CBIR systems. Methods of improving the original technique include the use of cumulative color histograms, combining histogram intersection with some element of spatial matching [Stricker96], and the use of region-based color querying [Carson97]. As for texture similarity, the useful measures include the degree of *contrast*, *coarseness*, *directionality*, *regularity*, *periodicity*, *randomness*, etc. Alternative methods of texture analysis for retrieval include the use of Gabor filters [Ma98] and fractals [Kaplan98]. Unlike texture, shape is a fairly well defined concept. Two main types of shape features are commonly used, namely (1) the *global* features such as aspect ratio, circularity and moment invariants, and (2) the *local* features such as sets of consecutive boundary segments. Alternative methods proposed for shape matching include elastic deformation of templates [Zhong00].

An impediment to research on CBIR is the lack of mapping between the high-level concepts and the low-level features. In order to overcome this problem and to better capture the subjectivity of human perception of the visual content, the concept of relevance feedback (RF) associated with CBIR was proposed in [Rui97]. Relevance feedback is an interactive process in which the user judges the quality of the retrieval results performed by the system by marking those images that the user perceives as truly relevant among the images retrieved by the system. This information is then used to refine the original query. However, even if the user provides a good initial query, it is still a problem of how to navigate through the image database.

No matter what information and what techniques are used for the construction of the image indexes, and no matter what similarity measurement strategies are employed, as far as the searching process is concerned, simple approaches such as sequential searching, are commonly put into operations to find the group of similar images for the queries. Such kinds of approaches may be adequate for small databases. However, as the scales and volumes of the databases increase considerably, they are deficient. Moreover, these approaches focus on only the relationship between the query image and the target image, neglecting the relationships among all the images within the database, which may result in inflexible and incomplete searching results.

There have been quite a few techniques being proposed and employed to alleviate the time consumption problem and to speed up the retrieval process, such as efficient indexing structures, compact representations, and pre-filtering techniques [Hafner95]. The QBIC system [Faloutsos94, Flickner95], for an instance, uses the pre-filtering technique and the efficient indexing structure like R-trees to accelerate its searching performance. However, little has been done in considering the complicated relationships of the image objects to each other.

In this chapter, we present a content-based retrieval system that employs the Markov model mediator (MMM) mechanism to retrieve images, which functions as both the searching engine and image similarity arbitrator. In our previous studies, the MMM mechanism has been applied to multimedia database management [Shyu00a, Shyu00c] and document management on the World Wide Web (WWW) [Shyu00b, Shyu01a, Shyu01b]. 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 to scientists and engineers for analyzing complicated systems, whereas a mediator is defined as a program that collects and combines information from one or more sources, and finally yields the resulting information [Wiederhold92]. A Markov model consists of a number of states connected by transitions. The Markov property states that given the current state of the system, the

future evolution is independent of its history. In other words, the states represent the alternatives of the stochastic process and the transitions contain probabilistic and other data used to determine which state should be selected next. All the transitions $S_i \rightarrow S_j$ such that $\Pr(S_j | S_i) > 0$ are said to be allowed, the rest are prohibited. Markov models have been used in many applications. Some well-known examples are Markov Random Field Models in [Frank86], and Hidden Markov Models (HMMs) [Rabiner86]. Some research works have been done to integrate the Markov model into the field of image retrieval. Lin et al. [Lin97] 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 hidden Markov model (HMM) in order to adequately capture both the spatial and chromatic information about that image. [Wolf97] used the hidden Markov model (HMM) to parse video data. In [Naphade01], the hidden Markov model was employed to model the time series of the feature vector for the cases of events and objects in their probabilistic framework for semantic level indexing and retrieval.

Our proposed CBIR system employs the MMM mechanism as well as the stochastic-based similarity measures for dynamical content-based image retrieval, which retrieves images with respect to the user queries. Our method also builds an index vector for each image within the database, but unlike the common methods mentioned above, our method considers not only the relationship between the query image and the target image, but also the relationships among all images in the database. A stochastic process that takes into account the image content features and other characteristics (such as the access frequencies and access patterns) of the images is proposed. Several experiments are conducted and the experimental results demonstrate that the MMM mechanism together with the stochastic process can assist in retrieving more accurate results for user queries.

3. The Stochastic Model

3.1 Markov Model Mediator (MMM) Mechanism

Markov model mediator (for short, MMM) is a probabilistic-based mechanism that adopts the Markov model framework and the mediator concept [Shyu00a, Shyu00b, Shyu00c, Shyu01a, Shyu01b]. In our CBIR system, each image database is modeled by a MMM. The MMM mechanism is defined as follows.

Definition 1: A MMM is represented by a 5-tuple $\lambda = (\mathcal{S}, \mathcal{F}, \mathcal{A}, \mathcal{B}, \Pi)$, where \mathcal{S} is a set of images called states; \mathcal{F} is a set of distinct features of the images; \mathcal{A} denotes the state transition probability distribution, where each entry (i, j) actually indicates the relationship between images i and j ; \mathcal{B} is the observation symbol probability distribution; and Π is the initial state probability distribution.

Each image database is modeled by a MMM, where \mathcal{S} consists of all the images in the image database and \mathcal{F} includes all the distinct features for the images in \mathcal{S} . The elements in \mathcal{S} and \mathcal{F} determine the dimensions of \mathcal{A} and \mathcal{B} . If there are totally n images in \mathcal{S} , and the number of distinct features in \mathcal{F} is m , then the dimensions of \mathcal{A} is $n \times n$, while \mathcal{B} has the size of $n \times m$. The relationships of the images are modeled by the sequences of the MMM states connected by transitions. A training data set consisting of the access patterns and access frequencies of the queries issued to the database is used to train the model parameters for a MMM.

3.2 Formulation of the Model Parameters

Each MMM has three probability distribution matrixes: \mathcal{A} , \mathcal{B} , and Π . These matrixes are critical for the stochastic process, and can be obtained from the training data set.

3.2.1 Matrix \mathcal{B} : the observation symbol probability distribution

The observation symbol probability \mathcal{B} denotes the probability of observing an output symbol from a state. Here, the observed output symbols represent the distinct features of the images and the states represent the images in the database. Since an image has one or more features and one feature can appear in multiple images, the observation symbol probabilities show the probabilities that a feature is observed from a set of images.

In this study, we consider the following features: color information and object location 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 [Ma99]. Humans perceive a color as a combination of three stimuli, R (red), G (Green), and B (Blue), which form a color space. Separating chromatic information and luminance information can generate more color spaces such as RGB, YIQ, YUV, CIE LAB, CIE LUV, and HSV. None of them can be used for all applications [Androutsos99, Aslandogan99, Cheng01a, Cheng01b, Ma99, Rui99]. RGB is the most commonly used color space primarily because color image acquisition and recording hardware are designed for this space. However, the problem of this space is the close correlation among the three components, which means that all three components will change as the intensity changes. This is not good for color analysis. YIQ and YUV are used to represent the color information in TV signals in color television broadcasting. CIE LAB and CIE LUV are often used in measuring the distance between two colors because of its perceptual uniformity. However, its transformation from the RGB space is computationally intensive and dependent on a reference white. In our CBIR system, color information is obtained for each image from its HSV color space. The whole color space is divided into twelve sub-spaces according to the combinations of different ranges of hue (H), saturation (S),

and intensity values (V). 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 [Androutsos99, Cheng01a, Cheng01b]. Secondly, the benchmark results in [Ma99] shows that the color histogram in the HSV color space performs the best. For information of object locations, the SPCPE algorithm proposed in [Sista99, Chen00] is used. The minimal bounding rectangle (MBR) concept in R-tree [Guttman84] is adopted so that each object is covered by a rectangle. The centroid point of each object is used for space reasoning so that any object is mapped to a point object. When these features are integrated into the queries, the semantic level meaning in the users' queries can be captured.

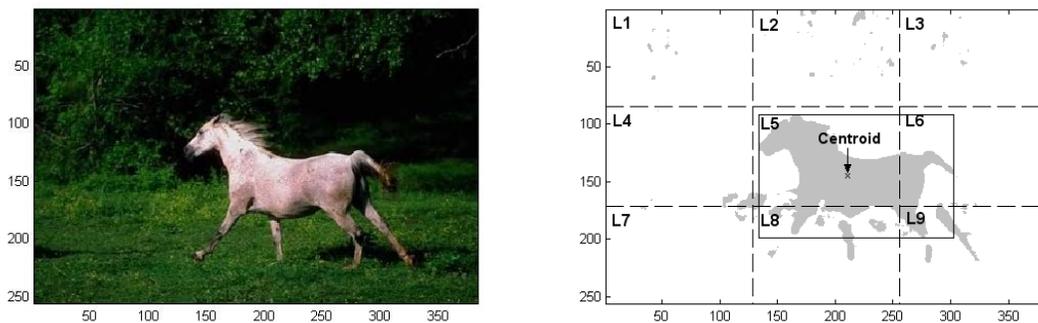


Figure 1: Object locations and their corresponding regions

In our experiments, each image has a feature vector of twenty-one elements. Within the twenty-one features, twelve are for color descriptions and nine are for location descriptions. The color features considered are ‘black’, ‘white’ (w), ‘red’, ‘red-yellow’ (ry), ‘yellow’ (y), ‘yellow-green’ (yg), ‘green’, ‘green-blue’ (gb), ‘blue’ (b), ‘blue-purple’ (bp), ‘purple’ (p), and ‘purple-red’ (pr). 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 value 0. Otherwise, we put the corresponding percentage of that color component to that position. As for the location descriptions, each image is divided into 3×3 equal-sized regions. The image can be divided into a coarser or finer

set of regions if necessary. As shown in Figure 1, the 9 regions are ordered from left to right and top to bottom: L1, L2, L3, L4, L5, L6, L7, L8, and L9. When there is an object in the image whose centroid falls into one of the nine regions, the value 1 is assigned to that region. Objects with their areas less than 8% of the total area are ignored.

In order to capture the appearance of a feature in an image, we define a temporary matrix (\mathcal{BB}) whose rows are all the distinct images and columns are all the distinct features, where the value in the $(p, q)^{th}$ entry is greater than zero if feature q appears in image p , and 0 otherwise. Then the observation symbol probability distribution \mathcal{B} can be obtained via normalizing \mathcal{BB} per row. In other words, the sum of the probabilities that the features are observed from a given image should be 1.

Matrix \mathcal{BB} consists of image feature vectors for all images. Figure 2 gives three example images and Table 1 illustrates their associated feature vectors. The observation symbol probability distribution \mathcal{B} can be obtained via normalizing \mathcal{BB} per row as shown in Table 2. In other words, the sum of the probabilities that the features are observed from a given image should be 1. We consider that the color and location information are of equal importance, such that the sum of observed probability of color features should be equal to that of location features (0.5 each).



Figure 2: Three sample images (Img1 – Img3)

Table 1: \mathcal{BB} matrix - Image feature vectors of sample images

	black	w	red	ry	y	yg	green	gb	b	bp	p	pr	L1	L2	L3	L4	L5	L6	L7	L8	L9
img1	0.107596	0	0.82744	0.06497	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
img2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
img3	0.53452	0	0	0.29148	0	0	0.174	0	0	0	0	0	0	0	0	0	1	0	0	0	0

Table 2: \mathcal{B} matrix – Normalized image feature vectors of sample images

	black	w	red	ry	y	yg	green	gb	b	bp	p	pr	L1	L2	L3	L4	L5	L6	L7	L8	L9
img1	0.053798	0	0.41372	0.032485	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0.25	0
img2	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0
img3	0.26726	0	0	0.14574	0	0	0.087	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0

3.2.2 Training data set

A set of training data is used to generate the training traces that are the central part of the stochastic process. Definition 2 gives the information that is available in the training data set. Based on the information in the training data set, we calculate the relative affinity measurements of the images in the image database (as shown in Definition 3).

Definition 2: The training data set consists of the following information:

- The value n that indicates the number of images in the image database d .
- A set of queries $Q = \{q_1, q_2, \dots, q_q\}$ that are issued to the database in a period of time. Each query shows the access patterns and access frequencies of the images. 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 0 otherwise. The value of $access_k$ denotes the access frequency of query q_k per time period.

Definition 3: The relative affinity measurements indicate how frequently two images are accessed together. The relative affinity relationship between two images m and n is defined as follows.

$$aff_{m,n} = \sum_{k=1}^q use_{m,k} \times use_{n,k} \times access_k \quad (1)$$

Table 3 gives eight example queries issued to the image database with their corresponding access frequencies. The access patterns of the three sample images in the database versus the eight example queries are shown in Table 4. In this table, the entry $(i, j) = 1$ indicates that the i^{th} image is accessed by the j^{th} query (i.e., q_j).

Table 3: Eight example queries and their frequencies ($access_k$)

Query Type	Feature Required	Frequency
q1	black / L1	1200
q2	blue	1500
q3	white / red/ L5	2500
q4	yellow / L5	1750
q5	green / gb	1250
q6	purple / L9	2220
q7	ry / L5	1870
q8	bp	1345

Table 4: The access patterns of the sample images

	q1	q2	q3	q4	q5	q6	q7	q8
img1	1	0	1	1	0	0	1	0
img2	0	1	1	1	0	0	1	0
img3	1	0	1	1	1	0	1	0
...

3.2.3 Matrix \mathcal{A} : the state transition probability distribution

The state transition probability distribution (matrix \mathcal{A}) is constructed by having $a_{m,n}$ be the element in the $(m, n)^{th}$ entry in \mathcal{A} , where

$$a_{m,n} = \frac{f_{m,n}}{f_m} \quad (2)$$

$$f_{m,n} = \frac{aff_{m,n}}{\sum_{m \in d} \sum_{n \in d} aff_{m,n}} \quad (3)$$

$$f_m = \sum_n f_{m,n} \quad (4)$$

In this formulation, $f_{m,n}$ is the joint probability that refers to the fraction of the relative affinity of images m and n in the database d with respect to the total relative affinity for all the images in d , f_m is the marginal probability, and $a_{m,n}$ is the conditional probability that refers to the state transition probability for a MMM.

3.2.4 Matrix Π : the initial state probability distribution

The preference of the initial states for queries can be obtained from the training traces. For any image $m \in d$, the initial state probability is defined as the fraction of the number of occurrences of image m with respect to the total number of occurrences for all the images in d from the training traces.

$$\Pi = \{\pi_m\} = \frac{\sum_{k=1}^q use_{m,k}}{\sum_{l=1}^n \sum_{k=1}^q use_{l,k}} \quad (5)$$

3.3 Stochastic Process for Information Retrieval

The need for efficient information retrieval from databases is strong. Usually the cost for query processing is expensive and time-consuming; meanwhile the results may not be very satisfactory. Probabilistic models offer a way to perform the searching process more efficiently and accurately. We capture the most matched images through a dynamic programming algorithm that conducts a stochastic process in calculating the current edge weights and the cumulative edge weights.

Assume N is the total number of images in the databases, and each query is denoted as $query = \{o_1, o_2, \dots, o_T\}$, where T is the total number of features requested in the query. We define the edge weights and the cumulative edge weights as follows.

Definition 4: $W_t(i, j)$ is defined as the edge weight of the edge $S_i \rightarrow S_j$ at the evaluation of the t^{th} feature (o_t) in the query, where $1 \leq i, j \leq N$ and $1 \leq t \leq T$.

Definition 5: $D_t(i, j)$ is defined as the cumulative edge weight of the edge $S_i \rightarrow S_j$ at the evaluation of the t^{th} feature (o_t) in the query, where $1 \leq i, j \leq N$ and $1 \leq t \leq T$.

Based on Definitions 4 and 5, the *dynamic programming algorithm* is given as follows.

At $t = 1$, we define

$$W_1(i, j) = \begin{cases} \pi_{S_i} b_{S_i}(o_1) & i = j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$D_1(i, j) = W_1(i, j) \quad (7)$$

The values of $W_{t+1}(i, j)$ and $D_{t+1}(i, j)$, where $1 \leq t \leq T-1$, are calculated using the values of $W_t(i, j)$ and $D_t(i, j)$ as follows.

$$W_{t+1}(i, j) = \max_k (D_t(k, i) a_{S_i, S_j}) b_{S_j}(o_{t+1}) \quad (8)$$

$$D_{t+1}(i, j) = (\max_k D_t(k, i)) + W_{t+1}(i, j) \quad (9)$$

As we mentioned before, $\mathcal{A} = \{a_{S_i, S_j}\}$ denotes the states transition probability distribution, $\mathcal{B} = \{b_{S_j}(o_k)\}$ is the observation symbol probability distribution, and $\Pi = \{\pi_{S_i}\}$ is the initial state probability distribution.

The image retrieval steps using the dynamic programming algorithm in the stochastic process are shown in Table 5. As can be seen from the result, our method can give a good ordering using this stochastic process.

Table 5: Image retrieval steps using the stochastic model

- 1) Given the query image q , obtain its feature vector $query = \{o_1, o_2, \dots, o_T\}$, where T is the total number of non-zero features of the query image q .
- 2) Upon the first feature o_1 , calculate $W_1(i, j)$ and $D_1(i, j)$ according to Equations (6) and (7).
- 3) Move on to calculate $W_2(i, j)$ and $D_2(i, j)$ according to Equations (8) and (9).
- 4) Continue to calculate the next values for the W and D matrices until all the features in the query have been taken care of.
- 5) Upon each feature in query, we can obtain a pair of matrices: $W_t(i, j)$ and $D_t(i, j)$. We then sum up each column in matrices $W_t(i, j)$ and $D_t(i, j)$. Namely, we calculate $sumW_t(j) = \sum_i W_t(i, j)$ and $sumD_t(j) = \sum_i D_t(i, j)$.
- 6) Find the candidate images by sorting their corresponding values in $D_T(q, j)$, $sumD_T(j)$, $sumD_{T-1}(j)$, ..., or $sumD_1(j)$. First, an image is ranked according to its value in $D_T(q, j)$. If there exist several images that have the same value, then $sumD_T(j)$ values are used for ranking. If several images have the same $sumD_T(j)$ value, then $sumD_{T-1}(j)$ values are used and the process continues until $sumD_1(j)$.
- 7) Select the top ranked images from the output of Step 6, and rank them to the user based on their values in $W_T(q, j) + W_T(j, q)$.

In Step 3, since we already obtained matrices $W_1(i, j)$ and $D_1(i, j)$ from Step 2, and the second feature o_2 is known, the content of $W_2(i, j)$ and $D_2(i, j)$ can be determined. The value of $D_i(i, j)$ (obtained in Step 4) represents the cumulative edge weight for the joint event that $\{o_1, o_2, \dots, o_T\}$ is observed. In the filtering step (Step 6), $D_T(q, j)$ together with $sumD_i(j)$ (where $1 \leq i \leq T$) are used as the filter to retrieve the candidate images with respect to query image q . In this step, when there are images with identical $D_T(q, j)$ values, we go to the matrix $sumD_T(j)$ to find different values to order them. If we fail to order them by the values in $sumD_T(j)$, we have to trace down to the next matrix $sumD_{T-1}(j)$ and continue the process until we reach the first matrix $sumD_1(j)$. We also take into consideration the characteristics of $sumW_i(j)$. From our observations, if the j^{th} image does not have the i^{th} feature in the query, the value of $sumW_i(j)$ would be zero. Taking advantage of this characteristic, we can exclude some of the images that do not have any feature desired in the query from the final result. Therefore, in Step 7, we use $W_T(q, j) + W_T(j, q)$ to reflect the possibility that the j^{th} image matches the issued query. In other word, it indicates the matching percentage of the j^{th} image in the image database to the query image q with respect to the features $\{o_1, o_2, \dots, o_T\}$.

4. Experiments

4.1 Experimental Image Database System

In our image database, there are 1,500 color images of various dimensions that are used to carry out the experiments. With the purpose of supporting semantic level meaning in the users' queries, both the color information and object location information are considered in our experiments. In addition, the query-by-example strategy is used for query issuing in our experiments. Based on the training data set of this image database, first we need to construct the model parameters for the MMM mechanism for the database.

4.2 Constructions of the Model Parameters

Each MMM has three probability distributions (\mathcal{A} , \mathcal{B} , and Π). The state transition probability distribution \mathcal{A} can be obtained according to Equations (1) to (4) given in Section 3.2.2-3.2.3. In order to calculate \mathcal{B} , first we need to construct $\mathcal{B}\mathcal{B}$ based on the images and their features in the experimental database. Based on $\mathcal{B}\mathcal{B}$, \mathcal{B} can be obtained using the procedure illustrated in Section 3.2.1. The initial state probability distribution for experimental database can be determined by using Equation (5). The constructions of these model parameters can be performed off-line.

Once the model parameters of the MMM for the image database is constructed, the stochastic process shown in Table 5 is used for image retrieval.

4.3 Stochastic Process for Example Queries

For a given query image issued by a user, the stochastic process with the proposed dynamic programming algorithm will be carried on to dynamically find the matched images for the user's query. A series of W_t and D_t matrices are generated according to Equations (6) to (9). The qualifying degrees of the images with respect to the certain query image are estimated by the values in the resulting W_t and D_t matrices according to rules described in Table 5.

In this section, we use a set of example queries to demonstrate the effectiveness of our stochastic model. For each set of query results, the qualifying possibilities of the images are in the descending order from the *top left* to the *bottom right*. The searching results are listed and analyzed as well. As can be seen from the experimental results, our method effectively extracts the images that contain the features specified in the query image and ranks them appropriately.

◆ Query I:



In this query example, the query image #2265 has one color feature and one location feature which are ‘blue’ (b) and ‘L5’. Below gives the corresponding \mathcal{B} matrix entry for it:

	black	w	red	ry	y	yg	green	gb	b	bp	p	pr	L1	L2	L3	L4	L5	L6	L7	L8	L9
img2265	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0

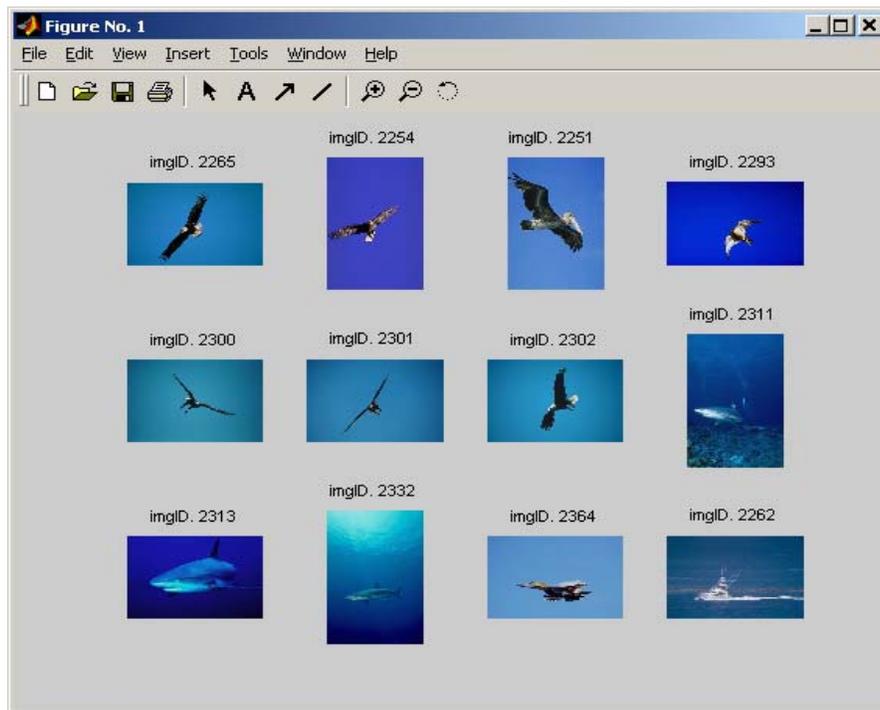
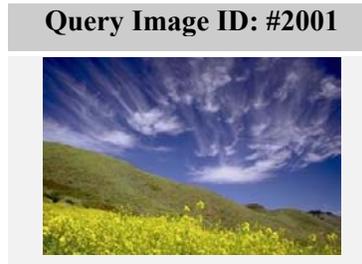


Figure 3: Snapshot of Query I

Since there are two features in the query image, two W matrices and two D matrices will be generated. The system is supposed to return those images that have the desired features in the ordering of similarity. The snapshot of the retrieval result screen containing the top 12 images is shown in Figure 3. As can be seen from this figure, the ‘blue’ color is the dominating color in all the

retrieved images. Moreover, they all have an object located at ‘L5’ (the centre location). It should also be noted that some images are excluded from the query results because their values in $sumW_T(j)$ are zeros, which means they do not have the corresponding features (‘blue’ or ‘L5’).

◆ **Query II:**



In this query, the major features of this query image include four components: ‘blue’ (b), ‘white’ (w), ‘yellow-green’ (yg), and ‘L5’ (the centre location) with the following \mathcal{B} matrix:

	black	w	red	ry	y	yg	green	gb	b	bp	p	pr	L1	L2	L3	L4	L5	L6	L7	L8	L9
img2001	0	0.1686	0	0	0	0.173	0	0	0.1584	0	0	0	0	0	0	0	0.5	0	0	0	0

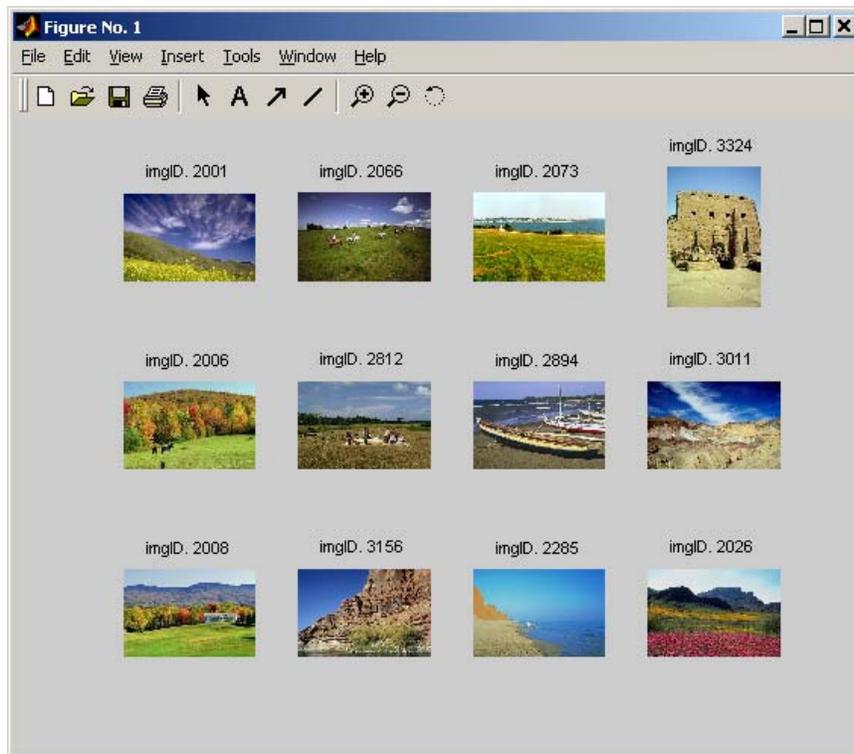


Figure 4: Snapshot of Query II

The most qualified images to this query are those that have all of the above three color features and an object at the L5 location (i.e., the centre location). Images that have only one of the desired features are less satisfactory. The snapshot of this example query is given in Figure 4. All the retrieved top 12 images have the above mentioned color features, and have the object(s) at location ‘L5’.

◆ **Query III:**

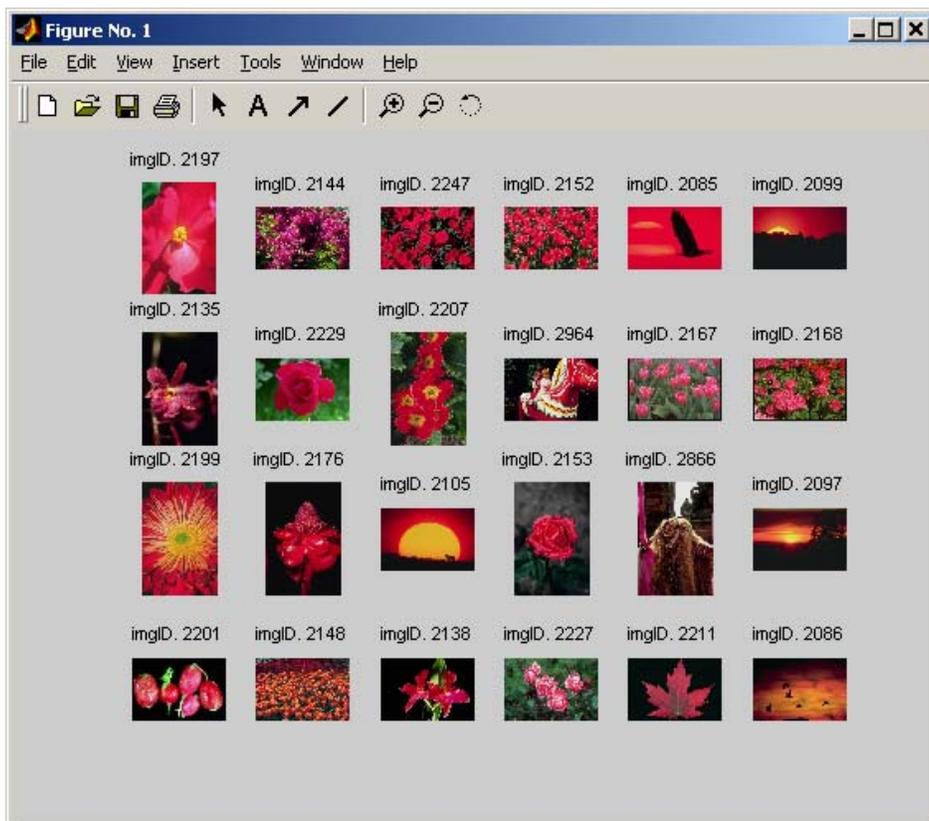


Figure 5: Snapshot of **Query III**

Similar to the previous query example, there are four features which are ‘red’, ‘purple-red’, ‘L1’, and ‘L5’. Two of them are the color features, and the other two are the object location descriptions. Figure 5 exhibits the top 24 images retrieved from the image database. The final query results are good, and the ranking is reasonable.

5. Future Trends

The current CBIR systems are promising and reflect the increasing interest in the field of content-based image retrieval. However, there are still a number of open research issues to be addressed. For example, it is critical to develop the suitable evaluation criteria and benchmarks for CBIR systems. Some future trends include:

- ◆ *Automatic or semi-automatic methods of object extraction for image retrieval:* It has been recognized that the searching for images in large databases can be greatly facilitated by the use of semantic information such as object location and type. However, using the current technique of computer vision cannot extract the object information easily. For example, even though the unsupervised image segmentation is applied in our framework to obtain the object information, we still have to find a trade-off between the accuracy and performance. Considering that users may upload their own query image (which is not in the database) during the content-based retrieval, the real-time feature extraction could be a big issue to affect the users’ will to use the system. There is an increasing need to derive more efficient yet good enough methods of segmenting images to distinguish objects of interest from their background. The main goal is to bridge the semantic gap, bringing some measure of automation to the processes of indexing and retrieving images by the type of object or scene depicted.

- ◆ *Indexing standard and query language support for image databases:* Currently, there are neither the standards in indexing the features and the subject taxonomies for image databases, nor the standard for hierarchical representation of an image when taking into consideration the objects inside the image. These standards are essential in expressing complex semantics and supporting the manipulation of content-based queries on image objects. Once the feature representations of image objects follow the same standards, it becomes possible to develop a suitable query language exclusively designed for image databases.
- ◆ *Support for automatic video indexing and retrieval:* Recently, lots of research work has been done on automatic video segmentation. After video being segmented into smaller units such as shots or scenes, each unit is represented by its key frames. Then, the complex spatial-temporal semantics can be obtained by parsing the content of these key frames, which forms the basis for supporting advanced video retrieval techniques such as query-by-motion facilities. MPEG-7 (Moving Picture Experts Group) standard [MPEG7] is the first to take into consideration the issue of multimedia content representation seriously. This standard will define a standard for describing every aspect of the content of a multimedia object, including the specifications of a video's image features. MPEG-7 will definitely have an impact on CBIR and will probably guide the development of future CBIR systems.
- ◆ *Better user interaction, especially the improved techniques for collecting users' feedback:* User's relevance feedback has been adopted in most of recent efforts towards the research of CBIR. In order to get better results, the user may be asked to browse a bunch of images through iterations and to provide the detailed ranking for similarity for the images. The fact is that a heavy and unnecessary burden of responsibility is brought to the user. In addition, it is highly probable that this burden will have a negative effect on user's perception of the effectiveness and efficiency of the system.

6. Conclusion

Currently, Content-Based Image Retrieval (CBIR) technology is still immature but with great potential. In this chapter, 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 efficiency concern of the searching process. In response to this issue, in this chapter, the Markov model mediator (MMM) mechanism is applied to the image databases for content-based image retrieval. A stochastic process based on the MMM mechanism is proposed to traverse the database and find the similar images with respect to the query image. This approach performs similarity comparison based on not only the relationship between the query image and the target image, but also the relationships among all the images within the database, such as their access patterns and access frequencies. Joint color/layout similarity measurement is supported in this system to offer more complete distinction descriptions of the images and better retrieval effectiveness. Several experiments were conducted and the experimental query-by-example image query results to the proposed retrieval system were reported. The fact that the proposed stochastic content-based CBIR system utilizes the MMM mechanism and supports both spatial and color information offers a more flexible and accurate results for user queries. The experimental results exemplify this point, and the overall retrieval performance of the presented system is promising. Besides modelling the relationship of image objects within a single database, the MMM mechanism also has the capability (although not shown in this chapter) to model the relationships among distributed image databases so as to guide the efficient search across the distributed image databases.

Acknowledgement

For Shu-Ching Chen, this research was supported in part by NSF EIA-0220562 and NSF CDA-9711582.

References

- [Androutsos99] D. Androutsos, "Efficient Indexing and Retrieval of Color Image Data Using a Vector-Based Approach," Ph.D. Dissertation, University of Toronto, 1999.
- [Aslandogan99] Y. A. Aslandogan, C. T. Yu, "Techniques and Systems for Image and Video Retrieval," *IEEE Trans. On Knowledge and Data Engineering*, Vol.11, No.1, Jan/Feb 1999.
- [Carson97] C. S. Carson, et al., "Region-based image querying," in *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, San Juan, Puerto Rico, pp. 42-49, 1997.
- [Chen00] S.-C. Chen, S. Sista, M.-L. Shyu, and R.L. Kashyap, "An Indexing and Searching Structure for Multimedia Database Systems," *IS&T/SPIE Conference on Storage and Retrieval for Media Databases 2000*, pp. 262-270, 2000.
- [Cheng01a] H.D. Cheng, X.H. Jiang, etc. "Color Image Segmentation: Advances and Prospects," *Pattern Recognition*, 2001.
- [Cheng01b] H.D. Cheng and Y. Sun, "A Hierarchical Approach to Color Image Segmentation Using Homogeneity," *IEEE Transactions on Image Processing*, Dec. 2001.
- [Faloutsos94] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic and W. Equitz, "Efficient and Effective Querying by Image Content," *Journal of Intelligent Information Systems*, 3, 3/4, pp. 231-262, July 1994.
- [Flickner95] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. steele, and P. Yanker, "Query by image and video content: The QBIC system," *IEEE Computer*, Vol. 28, No. 9, pp. 23-31, September 1995.

- [Frank86] O. Frank and D. Strauss, "Markov graphs," *Journal of the American Statistical Association*, 81, pp. 832-842, 1986.
- [Guttman84] A. Guttman, "R-tree: A Dynamic Index Structure for Spatial Search," *Proc. ACM SIGMOD*, pp. 47-57, June 1984.
- [Hafner95] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner and W. Niblack, "Efficient Color Histogram Indexing for Quadratic Form Distance Functions," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 17, no. 7, pp. 729-736, July, 1995.
- [Kaplan98] L. M. Kaplan, et al., "Fast texture database retrieval using extended fractal features," in *Storage and Retrieval for Image and Video Databases VI* (Sethi, I K and Jain, R C, eds), *Proc SPIE* 3312, 162-173, 1998
- [Li00] Z.-N. Li, Z. Tauber, and M. S. Drew, "Locale-Based Object Search Under Illumination Change Using Chromaticity Voting and Elastic Correlation," *Proc. ICME2000*.
- [Lin97] H. C. Lin, L. L. Wang and S. N. Yang, "Color Image Retrieval Based On Hidden Markov Models," *IEEE Transactions on Image Processing*, vol. 6, no. 2, pp. 332-339, 1997.
- [Ma99] W.-Y. Ma and H.J. Zhang, "Content-Based Image Indexing and Retrieval," *Handbook of Multimedia Computing*, CRC Press, 1999.
- [Ma98] W.-Y. Ma and B. S. Manjunath, "A texture thesaurus for browsing large aerial photographs," *Journal of the American Society for Information Science*, 49 (7) pp. 633-648, 1998.
- [MPEG7] <http://www.wileyurope.com/cda/cover/0,,0471486787|excerpt,00.pdf>
- [Naphade00] M. R. Naphade and T. S. Huang, "A probabilistic framework for semantic indexing and retrieval in video," *IEEE Transactions on Multimedia*, vol. 3, no. 1, March 2001.
- [Pentland94] A. Pentland, R.W. Picard, and S. Sclaroff, "Photobook: Tools for Content-Based Manipulation of Image Databases," *Proc. Storage and Retrieval for Image and Video Databases II*, Vol. 2185, pp.34-47, SPIE, Bellingham, Washington, 1994.

- [Rabiner86] L. R. Rabiner and B. H. Huang, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, Vol. 3, No. 1, pp. 4-16, January 1986.
- [Rui99] Y. Rui and T. S. Huang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues," *Journal of Visual Communication and Image Representation*, 1999.
- [Rui97] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based Image Retrieval with Relevance Feedback in MARS," *Proceedings of the 1997 International Conference on Image Processing (ICIP '97)* (3-Volume Set).
- [Shyu01a] M.-L. Shyu, S.-C. Chen, and C. Haruechaiyasak, "Mining User Access Behavior on the WWW," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1717-1722, October 7-10, 2001, Tucson, Arizona, USA.
- [Shyu01b] M.-L. Shyu, S.-C. Chen, C. Haruechaiyasak, C.-M. Shu, and S.-T. Li, "Disjoint Web Document Clustering and Management in Electronic Commerce," the *Seventh International Conference on Distributed Multimedia Systems (DMS'2001)*, pp. 494-497, September 26-28, 2001, Tamkang University, Taipei, Taiwan.
- [Shyu00a] M.-L. Shyu, S.-C. Chen, and R. L. Kashyap, "A Probabilistic-Based Mechanism For Video Database Management Systems," *IEEE International Conference on Multimedia and Expo (ICME2000)*, pp. 467-470, July 30-August 2, 2000, New York City, USA.
- [Shyu00b] M.-L. Shyu, S.-C. Chen, and C.-M. Shu, "Affinity-Based Probabilistic Reasoning and Document Clustering on the WWW," the *24th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC)*, Taipei, Taiwan, pp. 149-154, October 25-27, 2000.
- [Shyu00c] M.-L. Shyu, S.-C. Chen, and R. L. Kashyap, "Organizing a Network of Databases Using Probabilistic Reasoning," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1990-1995, October 8-11, 2000, Nashville, Tennessee, USA.

- [Sista99] S. Sista and R. L. Kashyap, "Unsupervised Video Segmentation and Object Tracking," *IEEE International Conference on Image Processing*, Japan, 1999.
- [Smith96] J. R. Smith and S. F. Chang, "VisualSEEK: A Fully Automated Content-based Image Query System," *In Proceedings ACM Intern. Conf. Multimedia*, pp. 87-98, Boston, Nov 1996.
- [Stricker96] M. Stricker and A. Dimai, "Color Indexing with Weak Spatial Constraints," in *Storage and Retrieval for Image and Video Databases IV*, (Sethi, I K and Jain, R C, eds), Proc SPIE 2670, pp. 29-40.
- [Virage] <http://www.virage.com>
- [Wiederhold92] G. Wiederhold, "Mediators in the architecture of future information systems," *IEEE Computers*, pp. 38-49, March 1992.
- [Wolf97] W. Wolf, "Hidden Markov model parsing of video programs," presented at the *International Conference of Acoustics, Speech and Signal Processing*, 1997.
- [Zhong00] Y. Zhong, A.K. Jain, and M.-P. Dubuisson-Jolly, "Object Tracking Using Deformable Templates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(5):544--549, May 2000.

Biography:

1st Author: Mei-Ling Shyu

Dr. Mei-Ling Shyu has been an Assistant Professor at the Department of Electrical and Computer Engineering, University of Miami since January 2000. She received her Ph.D. from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA in 1999. She also received her M.S. in Computer Science, M.S. in Electrical Engineering, and M.S. in Restaurant, Hotel, Institutional, and Tourism Management from Purdue University, West Lafayette, IN, U.S.A. in 1992, 1995, and 1997, respectively. Her research interests include multimedia networking, wireless communications and networking, data mining, multimedia database systems, multimedia information systems, and database systems. An expert in these fields, she has authored and/or co-authored more than 60 technical papers published in various prestigious journals and conferences in networking, data mining, database, and multimedia areas. She served on the program committee of the 3rd International Conference on Information Reuse and Integration (IRI-2001). She also serves currently on the program committee of ACM GIS 2002. She is listed in International Who's Who on Information Technology. She is a member of the Institute of Electrical and Electronics Engineers (IEEE), IEEE Women in Engineering Society, Association for Computing Machinery (ACM), and ACM SIGMOD.

2nd Author: Shu-Ching Chen

Dr. Shu-Ching Chen received his Ph.D. from the School of Electrical and Computer Engineering at Purdue University, West Lafayette, Indiana, USA in December 1998. He also received Masters degrees in Computer Science, Electrical Engineering, and Civil Engineering from Purdue University, West Lafayette, IN. He has been an Assistant Professor in the School of Computer Science (SCS), Florida International University (FIU) since August 1999. He is currently the director of Distributed Multimedia Information System Laboratory and the Associate Director of Center for Advanced Distributed System Engineering (CADSE). Dr. Chen's research interests include distributed multimedia database systems and information systems, data mining, databases, and multimedia communications and networking. Dr. Chen is authored and co-authored more than 70 research papers in journals and refereed conference proceedings. In addition, he is co-authored a book entitled "Semantic Models for Multimedia Database Searching and Browsing" and published by Kluwer Academic Publishers on September 2000. He is the general co-chair of the IEEE 2003 International Conference Information Reuse and Integration. He is the program co-chair of the 10th ACM International Symposium on Advances in Geographic Information Systems (ACM GIS-2002), the 2nd International Conference on Information Reuse and Integration (IRI-2000), and the International Symposium on Information Systems and Engineering (ISE2002) and will be the program co-chair of the 9th International Conference on Distributed Multimedia Systems (DMS'2003).

3rd Author: Chengcui Zhang

Chengcui Zhang is a research assistant at the school of Computer Science in Florida International University, pursuing a Ph.D. in Computer Science. Her research interests are in multimedia data mining, video segmentation and indexing, and image retrieval. She received her BS and MS in Computer Science from Zhejiang Univeristy in China.