Content-Based Image Retrieval Using Relevance Feedback

Jing Xin

Student ID: 3039107

Supervisor: Dr. Jesse Jin

Department of Software Engineering
School of Computer Science and Engineering
Telephone: 9385.6206
Email: jingx@cse.unsw.edu.au

Aug 2003

1. Introduction and Motivations

With the advances in computer technologies and the advent of the World Wide Web, there has been an explosion in the amount and complexity of digital data being generated, stored, transmitted, analyzed, and accessed. Much of this information is multimedia in nature, including digital images, video, audio, graphics, and text data. However, we cannot access or make use of the information unless it is organized so as to allow efficient browsing, searching and retrieval. In order to make use of this vast amount of data, efficient and effective techniques to retrieve multimedia information need to be developed. Among the various media types, images are of prime importance. Not only is it the most widely used media type besides text, but it is also one of the most widely used bases for representing and retrieving videos and other multimedia information.

Image Retrieval has been a very active research area since the 1970's, with the push from two major research communities, Database Management and Computer Vision. These two research communities study Image Retrieval from different angles, one being text-based and the other visual-based.

Text-based image retrieval is the traditional image retrieval paradigm. In this approach, the images are first annotated manually by some keywords, then they can be retrieved by using text-based database management systems (DBMS). However, there are there major difficulties with this approach, especially when the size of the image collections is large. One is the vast amount of labour required in manual image annotation. The other difficulty, which is crucial, results from the rich content in the images. Different people may perceive the same image content differently. The third is the inconsistency of the keyword assignments among different indexers (Faloutsos et al. 1993).

In the early 1990s, because of the emergence of large-scale image collections, these there difficulties faced by the manual annotation approach became increasingly more acute. To overcome these difficulties, content-based image retrieval (CBIR) was proposed. That is, instead of being manually annotated by text-based keywords, images would be indexed by their own visual content, such as color, texture, shape, etc. Since then, many techniques in this research direction have been developed and many image retrieval systems, both commercial and research, have been built.

Many early CBIR systems only use the one-shot approach during retrieval where a query is given in the form of a feature vector and the result is calculated based on the similarity of feature vectors. The main problem with these systems was that they have not accounted for the subjectivity of human perception. It is undeniable that different people under identical circumstances or the same person under different circumstances may perceive and interpret the same visual content in different ways. This is known as *human perception subjectivity*. In order to capture the user's perception subjectivity and allow the user to refine the query result, recent research has moved to an interactive mechanism that involves a human as part of the retrieval process. A natural way of interaction between a human and a computer is to ask the user to give feedback about the relevance of previously retrieved objects. Through successive feedback iterations, the system can capture the user's need more accurately and present more relevant images to the user.

However, despite the extensive research, the retrieval techniques used in CBIR systems lag behind the corresponding techniques of today's best text search, and many research issues remain to be solved in order to make CBIR systems more practical. One problem, which is maybe the most difficult problem, is the semantic gap that exists between the high level semantics in the human minds and the low level features computable by machines. Another is to resolve user inconsistency during feedback. That is, the user may give conflicting feedback during iterations during the whole process of relevance judgments.

The purpose of this thesis is to implement a CBIR system with the further improvement of usability and performance. The main research area in this thesis is the field of user interaction. We will focus on techniques for interactive image retrieval based on relevance feedback. We hope to design a CBIR system with more precisely relevance feedback management and much potential to shorten the semantic gap. The user interaction area also includes the aspects such as resolving the user inconsistency.

The motivations of this thesis come directly from the current research issues in the field of content-based image retrieval. Many problems in this area have been identified but much more effort is needed to work out solutions.

2. Review of Literature

Content-based image retrieval has attracted many researchers from various fields, including computer vision, image processing, information science, database management systems, etc. The major bases for content-based image retrieval lie in four areas: visual feature extraction, multidimensional indexing, retrieval techniques and system design. Here we present a comprehensive survey of what has been achieved in recent years in the research domains within CBIR.

2.1. Feature Extraction

Feature (content) extraction is the basis of content-based Image Retrieval. Visual features can be classified as color, texture, shape etc. Because of perception subjectivity, there does not exist a single best presentation for a given feature. For any given feature, there exist multiple representations which characterize the feature from different perspectives.

2.1.1. Color

The color feature may be one of the most straight-forward features utilized by humans for visual recognition and discrimination. It is relatively robust to background complication and independent of image size and orientation. Some representative studies of color perception and color spaces can be found in (Miyahara 1988) and (Wang et al. 1997).

In image retrieval, a color histogram is the most commonly used color feature representation. Statistically, it denotes the joint probability of the intensities of the three-color channels. Swain and Ballard (1991) proposed Histogram Intersection, an L1 metric, as the similarity measure for the color histogram. To take into account the similarities between similar but not identical colors, Ioka (1989) and Niblack et al. (1994) introduced an L2-related metric in comparing the histograms. Furthermore, considering that most color histograms are very sparse and thus sensitive to noise, Stricker and Orengo (1995) used the cumulated color histogram and their research results demonstrated the advantages of the proposed approach over the conventional histogram approach.

Besides color histogram, several other color feature representations have been applied in image retrieval, including color moments and color sets. To overcome the quantization effects as in color histogram, Stricker and Orengo (1995) proposed to use color moments approach. The mathematical foundation of this approach is that any color distribution can be characterized by its moments. Furthermore, since most of the information is concentrated on the low-order moments, only the first moment (mean), and the second moment (variance) were extracted as the color feature representation. The weighted Euclidean distance was used to calculate the color similarity.

To facilitate a fast search over large-scale image collections, Smith and Chang (1995a, 1995b) proposed color sets as an approximation to color histogram. They first transformed the (R, G, B) color space into a perceptually uniform space, such as HSV, and then quantised the transformed color space into M bins. A color set is defined as a selection of the colors from the quantized color space. Because color set feature vectors were binary, a binary search tree was constructed to allow fast search. The relationship between the proposed color sets and the conventional color histogram was further discussed in (Smith and Chang 1995a, 1995b). Both of these two approaches are widely used in color feature representation.

2.1.2. Texture

Texture refers to the visual patterns that have the properties of homogeneity that can not result from the presence of only a single color or intensity (Smith and Chang 1996a). Texture is an innate property of virtually all surfaces. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. In the early 70's, Haralick et al. (1973) proposed the co-occurrence matrix representation of the texture feature. This approach explored the gray level spatial dependence of texture. It first constructed a co-occurrence matrix based on the orientation and distance between image pixels and then extracted meaningful statistics from the matrix as the texture representation.

Motivated by the psychological studies in human visual perception of texture, Tamura et al. (1978) explored the texture representation from a different angle. They developed computational approximations to the visual texture properties found to be important in psychology studies. The six visual texture properties were coarseness, contrast, directionality, linelikeness, regularity, and roughness. One major distinction between the Tamura texture representation and the co-occurrence matrix representation is that all the texture properties in Tamura representation are visually meaningful whereas some of the texture properties used in co-occurrence matrix representation may not (for example, entropy). This characteristic makes the Tamura texture representation very attractive in image retrieval, as it can provide a friendlier user interface.

In early 90's, after Wavelet transform was introduced and its theoretical framework established, many researchers began to study the use of Wavelet transform in texture representation (Kundu and Chen 1992; chang and kuo 1993; Smith and Chang 1994; Gross et al. 1994; Thyagarajan et al. 1994). Smith and Chang (1994, 1996a) used the statistics (mean and variance) extracted from the Wavelet subbands as the texture representation. This approach achieved over 90% accuracy on the 112 Brodatz texture images. To explore the middle-band characteristics, tree-structured Wavelet transform was used by (Chang and Kuo 1993) to further improve the classification accuracy. Wavelet transform has also been combined with other techniques to achieve better performance.

There are also quite a few review papers in the area of texture representation. In (Ma and Manjunath1995), they evaluated the texture image annotation by various Wavelet transform representations, including orthogonal and bi-orthogonal Wavelet transforms, tree-structured

Wavelet transform, and Gabor wavelet transform. They found that Gabor transform was the best among the tested candidates, closely matching the human vision study results (Smith and Chang 1996a). So currently, Gabor transform is the most promising technique for texture identification recognition.

2.1.3. Shape

In general, shape representations can be divided into two categories, boundary-based and region-based. The former uses only the outer boundary of the shape while the latter uses the entire shape region (Rui et al. 1996). The most successful representatives for these two categories are Fourier Descriptor and Moment Invariants.

The main idea of Fourier Descriptor is to use the Fourier transformed boundary as the shape feature. Some early work can be found in (Persoon and Fu 1977). To take into account the digitization noise in the image domain, Rui et al. (1996) proposed a modified Fourier Descriptor which is both robust to noise and invariant to geometric.

The main idea of Moment Invariants is to use region-based moments, which are invariant to transformations, as the shape feature. Hu (1977) identified seven such moments. Based on his work, many improved versions emerged. Yang and Algregtsen (1994) proposed a fast method of computing moments in binary images based on the discrete version of Green's theorem. Kapur et al. (1995) developed a algorithm to systematically generate and search for a given geometry's invariants. Realizing that most researchers did not consider what happened to the invariants after image digitization, Gross and Latecki (1995) developed an approach which preserved the qualitative differential geometry of the object boundary, even after an image was digitized. In (Copper and Lei 1995) and (Lei et al. 1995), a framework of algebraic curves and invariants was proposed to represent complex objects in cluttered scene by parts or patches. Polynomial fitting was done to represent local geometric information, from which geometric invariants were used in object matching and recognition.

Some recent review papers in shape representations are (Li and Ma 1994) and (Mehtre 1997). Li and Ma showed that the Geometric Moments method (region-based) and the Fourier Descriptor (boundary-based) were related by a simple linear transformation. Mehtre et al. compared the performance of boundary-based representations (chain code, Fourier Descriptor, UNL Fourier Descriptor), region based representations (Moment Invariants, Zernike moments, pseudo-Zernike moments), and combined representations (Moment Invariants & Fourier Descriptor, Moment Invariants & UNL Fourier Descriptor). Their experiments showed that the combined representations outperformed the simple representations.

2.2 High Dimensional Indexing

To make the content-based image retrieval truly scalable to large size image collections, efficient multi-dimensional indexing techniques need to be explored.

2.2.1. Dimension Reduction

Before we utilize any indexing technique, it is beneficial to first perform dimension reduction. Karhunen-Loeve Transform (KLT) and Principal Component Analysis (PCA) have been applied in dimension reduction. Faloutsos and Lin (1995) proposed a fast approximation to KLT to perform the dimension reduction, and Ng and Sedighian (1996) followed the eigenimage approach to carry out the dimension reduction. Experimental results from their research showed that most real data set could be considerably reduced in dimension without significant degradation in retrieval quality (White and Jain 1996). Recently, Chandrasekaran et al. (1997) developed a low-rank Singular Value Decomposition (SVD) update algorithm which is efficient and numerically stable in performing KLT.

Clustering is another powerful tool in performing dimension reduction. Clustering technique is used in various disciplines such as pattern recognition, speech analysis, and information retrieval. Normally it is used to cluster similar objects (patterns, signals, and documents) together to perform recognition or grouping. This type of clustering is called row-wise clustering. However, clustering can also be used in column-wise to reduce the dimensionality of the feature space.

2.2.2. Multi-dimensional Indexing Techniques

After we identify the embedded dimension of the feature vectors, we need to select appropriate multi-dimensional indexing algorithms to index the reduced but still high dimensional feature vectors. The early multi -dimensional indexing techniques include Bucketing algorithm, k-d tree, priority k-d tree, quad-tree, K-D-B tree, and hB-tree. Later, Guttman proposed the R-tree indexing structure (Guttman 1984). Based on his work, many other variants of R-tree were developed. Sellis et al. (1987) proposed R^+-tree and Beckman and Kriegel (1990) proposed the best dynamic R-tree variant, R*-tree. However, even for R*-tree, it was not scalable to dimensions higher than 20 (Faloutsos et al. 1993).

Ng and Sedighian (1996) presented a very good review and comparison of various indexing techniques in image retrieval and provided general-purpose and domain-independent indexing algorithms. Motivated by k-d tree and R-tree, they proposed VAM k-d tree and VAMSplit R-tree. They pointed out that the VAMSplit R-tree provided the best performance, but the trade-off is the loss of dynamic nature of R-tree. On their test data sets, they showed that BA-KD-tree gave the best performance.

Berchtold et al. (1996) proposed to use X-tree as the tool for constructing the tree indexing structure in image retrieval. The advantage of using X-tree is that a new split algorithm is introduced to minimise the overlap problem in R-tree-based index structures.

Aimed to alleviate the problem of overlapping, Kurniawati et al. (1997) described the SS^+-tree. The SS^+-tree uses a tighter bounding sphere for each node and makes a better use of the clustering property of data by using the variant of *k*-means algorithm. Experiments show that this is an effective approach.

In addition to the above approaches, clustering and neural nets, widely used in pattern recognition, are also promising indexing techniques (Zhang and Zhong 1995).

2.3. Retrieval with Relevance Feedback

Initially developed in document retrieval, relevance feedback as a human-computer interaction technique was introduced into content-based image retrieval during late 1990's. Image retrieval system with relevance feedback techniques can achieve a considerable performance boost. A typical scenario for relevance feedback in content-based image retrieval is as follows:

(1). Search engine provides initial retrieval results through query-by-example;

(2). User provides judgment on the currently displayed images as to whether, and to what degree, they are relevant or irrelevant to her/his request;

(3). Computer analyzes what the user's need and tries again. Go to step (2).

Base on the feedback, the system can capture the user's need more accurately and present more relevant images to the user.

In recent years, a variety of solutions for relevance feedback have been proposed. In its short history, relevance feedback developed along the path from heuristic based methods to optimization, learning, or classification techniques. The early work is inspired by term weighting feedback techniques in document retrieval (Salton 1989). These methods proposed heuristic formulation with empirical parameter adjustment, mainly along each feature axis and each dimension in the feature space (Rui et al. 1998; Peng et al. 1999; Porkaew et al. 1999; Santini and Jain 2000). The intuition of this re-weighting method is to emphasize more on the features and dimensions that have greater relevance and can better separate the positive and the negative examples.

Later on, researchers began to look at this problem from a more systematic point of view by formulating it as an optimization problem. Ishikawa et al. (1998) and Rui and Huang (2000) presented a minimization problem on the parameter estimation process. The optimal solutions of both query vector and feature weights are obtained based on the minimization of total distances of positive examples from the new query. Additionally, Rui and Huang adopted a two level weighting scheme to better cope with singularity issue due to the small number of training samples.

Assuming that the user is searching for a particular target, and the feedback is in the form of "relative judgment", the PicHunter system (Cox et al.1998, 2000) proposed the *stochastic comparison search* as its relevance feedback algorithm. In PicHunter, each image is associated with a probability of being the user's target. And it uses Bayes's rule to predict the user's target. Moreover, while most systems use most positive display strategy, Cox et al. proposed a minimum entropy display strategy to select image presenting to the user. Most positive scheme is always choosing the neighbours nearest to display as measured by the Euclidean distance between feature vectors. By contrast, the minimum entropy display strategy may present the user some dissimilar images but it is expected to get most information gain from user's feedback.

Recently there have also been attempts that present the solution of relevance feedback in the view of two classes classification problem:

MacArthur et al. (2000) cast relevance feedback as a two class learning problem, and used a decision tree algorithm to recursively partition the feature space until all instances within a partition are of the same class. The resulting decision tree classified the database. Images that fall into a relevant leaf were collected and the nearest neighbours of the query were returned.

Zhou and Huang (2000, 2001) used the intuition that all positive examples are alike in a way, each negative example is negative in its own way, and proposed asymmetric treatment for the positive and negative examples. They assumed that the positive examples have a compact low-dimensional support while the negative examples can have any configuration. A custom designed discriminant analysis, namely, biased discriminant analysis (BDA), is applied to find the transformed, reduced-dimension space where the positive examples cluster while the negative scatters away.

Hong et al. (2000) and Chen et al. (2001) incorporate support vector machine (SVM) into relevance feedback process. In this approach, images that are far away from the separating hyperplane are considered to be more important. The farther the positive examples from hyperplane, the higher they are assigned preference. However, SVM as a two-class classifier is not directly suitable for relevance feedback. The main difficulty is that the way of selecting the kernel function is heuristic.

Most relevance feedback schemes are designed to deal with global image features, which sometimes is not the best choice. A number of methods propose region-based image retrieval by decomposing each image into a set of homogeneous regions. Using region-based techniques, similarity between images is assessed by computing similarity between pairs of regions and then combining the results at the image level. Fauqueur and Boujemaa (2002) presented a schema of coarse automatic image segmentation and automatically determine the specific regions of interested. Ratan et al. (1999) used multiple-instant learning model to learn the most important sub-images from example images, which are represented as a collection of instances. The Diverse Density algorithm is adopted to find the area in feature space that are shared by all positive images while far from all negative sub-images. Vasconcelos and Lippman (2000) used image local features for image regional query. Along the same line are the works by (Ko et al. 2001) and (Jing et al. 2003).

Some Latest works treat the relevance feedback as a statistic learning problem. Different assumptions about the distribution of the image collection have been proposed. Mix Gaussian assumption is the most common one. Expectation Maximization (EM) algorithm (Wu et al. 2000; Tian et al. 2000; Huang et al. 2002) is also introduced to use examples from the user feedback as well as other unlabeled data points to estimate the classification of image database. Furthermore, D-EM algorithm was proposed to perform discriminant analysis inside the EM iterations to relax the assumption of probabilistic structure of data distributions. EM has been widely used in statistics and is a valuable algorithm in the area of relevance feedback.

2.4. Image Retrieval Systems

Since the early 1990s, many image retrieval systems both commercial and research have been built. Here we present a few well-known systems with their main characteristics.

2.4.1. QBIC

IBM's QBIC (Flickner et al. 1995; Niblack et al. 1994) is by far the most cited system and it is the first content-based image retrieval system. It is made available on the web as a demonstration and can be purchased as a commercial product.

QBIC allows query by example images, by user sketch, by colors and by textures. The user sketches are shapes based on moments, area, circularity and other measures. The color features used are average colors in different color space, and color histogram in Munsell coordinates.

Its texture features are combinations of measures such as coarseness, contrast and directionality. The user can have several methods to turn on and turn off these features. QBIC is one of the few systems that utilized a high-dimensional indexing technique to facilitate fast searches. It uses KLT to perform dimension reduction and R*-tree to construct multi-dimensional indexing. In its new system, images can be annotated and text-based keyword search can be used for query as well.

2.4.2. Virage

Virage is another very early content-based image search engine. It is also a commercial system used by reputable clients such as CNN and NBC. In (Bach et al. 1996), the framework of Virage system is described. The features used for querying include local and global color, texture and object boundary information. On step further than QBIC, Virage supports arbitrary combinations of the above features. The user can adjust the weights of the atomic features according to their own emphasis. Furthermore, Virage proposes an open framework with API (Application Programming Interface) to design specialized applications.

2.4.3. Photobook

MIT Media Lab produced a number of image searching systems. The early version is Photobook published in 1996 (Pentland et al. 1996). The system can be downloaded from the internet freely. Photobook supports visual features based on shape, texture and face features, which are called three sub-books. Users can query according to each of the three sub-books.

The later version of Photobook is FourEyes, a learning system incorporating human in the image annotation and retrieval loop (Minka and Pcard 1996; Picard et al. 1996). FourEyes uses a “society of models” to select and combine modes base on example from the user. Experiment results show that this approach is effective in interactive image annotation.

2.4.4. Blobworld

One of the well-known retrieval systems that use image regions during the query process is Bolbworld. It is described in several articles and changed significantly several times (Carson et al. 1997, 1999, 2002).

The system first segments the images into a small number of homogenous regions based on 6-8 color and texture features. Then user can select one or more regions and query with the regions based on the regions color or texture characteristics, and also based on the spatial relationships of several regions. In its first version, the ellipses are used to symbolize the image regions. The more resent versions uses the real boundaries of the regions. Another advantage of this system is that the regions responsible for the retrieval are highlighted in the results section, facilitating the use of feedback. Although the segmentation might no always conform to what a user is looking for, it provides a more exact way to specify queries.

2.4.5. Mars

MARS (Multimedia analysis and retrieval system) is developed at University of Illinois in Urbana Champaign (Huang et al. 1996; Mehrotra et al 1997; Ortega et al. 1997, 1998; Rui et al. 1997b, 1998). The system differs from other systems with respect to research scope and techniques used. Its feature is the integration of computer vision, database management system (DBMS), and information retrieval (IR). The features adopted by the system are color, texture, shape and layout features. In (Rui et al. 1998), the subjective in human perception is analysed based on the observation that there is no single feature that can best represent images. It focuses on how to organize various visual features into a meaningful structure which can dynamically adapt to different application and different users. The main contribution of MARS is that it proposes a simple solution of relevance feedback in image retrieval and integrates such technique in query vector refinement and automatic feature adoption. This relevance feedback technique has profound effects on later image retrieval researches.

2.4.6. PicHunter

PicHunter image retrieval system is designed by NEC computer vision group (Cox et al. 1997, 1998, 2000). The assumption adopted in PicHunter is different from other systems. It assumes that the user is looking for an exact image. At each step, PicHunter uses a Bayesian Learning approach to update the similarity measure of each image in the database to the query. And it selects the set of images to minimal to total number of iterations needed to reach the target under the theory of maximizing the gain of information. One of the evaluations of the system is the number of images user is looking at before finding the target he has in mind.

2.4.7. Other systems

Some other content-base image retrieval systems include the early RetrievalWare system, developed in 1993, which emphasizes on the application of Neural Nets (Dowe 1993). VisualSEEL and WebSEEK are both developed at Columbia University (Smith and Chang 1996b, 1997). Their main features are spatial relationship query of image regions and visual feature extraction from compressed domain. The *El Niño* system described in (Santini and Jain

1998) offers an interesting user interface, where the user can move images closer together that he thinks are similar and thus changes the similarity space. Imedia is another system that specializes on image segmentation retrieval based on regions. Metaseek is by far the only meta search system which combines several well- known search engines such as QBIC, Virage, VisualSEEL and WebSEEK. All engines are queried and then the results are combined. It also supports relevance feedback with all the images of different image database (Benitez et al. 1998). In (Veltkamp and Tanase 2000), an overview on more image retrieval systems can be found.

3. Proposed Methods

A content-based image retrieval system will be implemented firstly. Then the relevance feedback techniques will be the studied based on the experimental system. Fig. 1 shows the outline of system structure. From this diagram we can see that there are several major components in this system: user interface, indexing subsystem, feature extraction subsystem, and retrieval subsystem. Methods used for these four parts will be described separately. The emphasis will be put on the retrieval subsystem since it is the key component of the CBIR system and our contributions mainly come from this part.

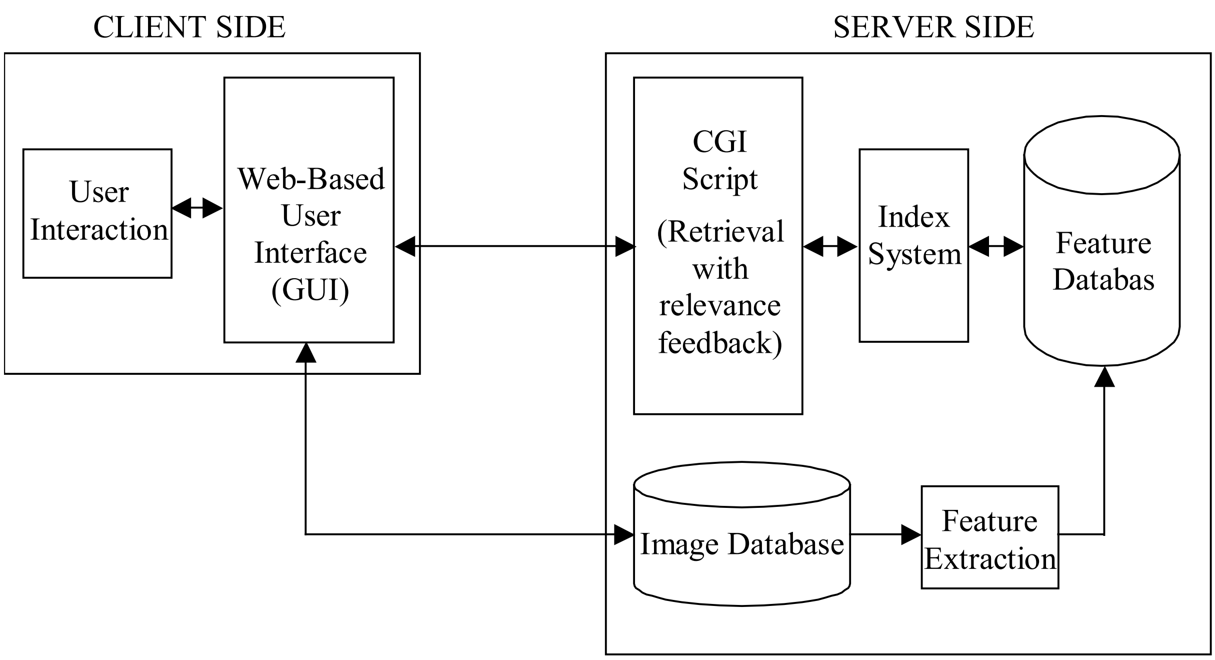

Fig1. Outline of System Structure

3.1. User Interface

This is the interface where the user can interact with the system. The implementation will be written in Java applets and can be accessible over the Internet by the technique of CGI-BIN. The user interface will support many options for query: query by example, query by feature, query by browsing, query by definition. Query by example is the most common way for the user to specify a query. The user may select a random set of images from the database, and once he finds an image of interest, he can click on the image to submit a query. Query by feature allows user pick up color, texture and shape information from different images, and the retireval can be performed based on the combination of features. Through query by definition, the user can select the color or creat arbitrary shapes on the palette, the system then retrieve similar images based on the defined shape or color. Query by browsing is combined with clustering technique. Once the user request the retrieval, all images within that cluster will be ranked by similarity and displayed to the user. The advantages of browing is that it provides users with a visual tool to narrow the search to a small region quickly.

3.2. Indexing subsystem

For large image collections, it is highly desirable to match queries without searching the entire image collections. Therefore, the efficient feature indexing is crucial for reducing the time to execute a query. In our CBIR system, we will use a clustering-based indexing approach and SS^+-tree to facilitate the fast search.

3.3. Feature extraction subsystem

Our CBIR system supports several feature representations. Color, texture and shape features are extracted from all the images in the database, and the corresponding feature database will be built.

Color: We will use the hue, saturation, value (HSV) color modal as color descriptor, since it approximates a perceptually uniform color modal, making it easier for the user to specify colors. The global color histogram of an image is computed and stored. It can be defined as $\max(\sum_{i,j\in\Omega} H(i,j,h))$, where i and j are the coordinates of each image pixel, h is the hue value and H is the hue histogram.

Texture: Many researches have shown that Gabor filter as texture extraction methods is relatively robust in image analysis. In our system, a set of thirty two-dimensional Gabor filters with five orientations and six bandwidths will be used as texture descriptor. Gabor function takes the form:

$$g(u,v) = e^{-\frac{1}{\pi}[\frac{(u-u_0)^2}{a^2}+\frac{(v-v_0)^2}{b^2}]} e^{-2\pi i|x_0(u-u_0)+y_0(v-v_0)|} \quad (3.3.1)$$

It is consistent with the human visual system.

Shape: By comparing the two categories of methods for shape representation, region-based method has more computational cost while boundary-based method can provide a much more complete description. We choose the boundary-based Fourier Transform method. Instead of using complex numbers in traditional Fourier Transform, we will use a simple Fourier descriptor as a shape representation. It is given by:

$$a(u) = 1\Big/ N\sum_{k=1}^{N-1} f(k)\exp[-j2\pi uk / N] \quad (3.3.2)$$

where $u = 0, 1,2 \ldots N-1$ and $f(k)$ is the distance from a pixel k on the shape boundary to its centroid.

3.4. Retrieval subsystem

As what we mentioned in section 1, the greatest challenge in image retrieval results from the semantic gap between the low level representation and the underlying high level concept in visual information. While the computer understands images with the low level features such as color, texture and shape, humans perceive images based on the semantic or true meaning of content. However, it is very difficult to directly extract the semantic level features from images with the current technology in computer vision and image understanding. An alternative way to bridge this gap is to gather information on semantics from the user interaction, that is the function of relevance feedback. By the user interaction, the queries are refined and the similarity metric on the visual feature space is dynamically updated. We have attempted on designing a more efficient relevance feedback technique applied in our CBIR system.

The process of relevance feedback can be divided into two procedures: information updating and display updating. The purpose of information updating is to infer the user's intention by analysing the user's feedback. The display procedure is to determine the next display image set on the basis of the update information. In the following, we will present a joint design including the information updating strategy, display updating strategy, and the consideration of resolving the problem of user inconsistency.

3.4.1 Model

Before we describe how our relevance feedback technique works, we first need to introduce the model adopted during the whole design. We treat the user's target as a statistic distribution and model its density as a Gaussian mixture of the following:

$$f(\underline{x}_i \mid \Theta) = \sum_{j=1}^{K} w_j f_j(\underline{x}_i \mid \theta_j), \quad \sum_{j=1}^{K} w_j = 1 \qquad (3.4.1.1)$$

where $\underline{x}_i = (x_i^1, x_i^2 \ldots, x_i^d)^T$ is an d-dimensional low level feature vector, $\Theta = (w_1, ..., w_k, \theta_1, \ldots, \theta_k)$ ($\theta_j = (\mu_j, \Sigma_j)$) is a parameter set of the mixture model, K is the number of components, and $f_j(\underline{x}_i \mid \theta_j) = \frac{1}{(2\pi)^{\frac{d}{2}} \left|\Sigma_j\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{x}_i - \mu_j)^T \Sigma_j^{-1} (\underline{x}_i - \mu_j)}$ is the probability density of component j.

The motivations of using this model come from two aspects. First, multi-model, compared with single-model, is a more elaborate model to represent the user's mind. It is common that the user's target is a combination of some components. For example, the user is looking for some pictures of flowers. The color feature that user is interested in is maybe multiple. Thus there is no single-model sufficient to cover the user's desire. In the most cited paper within the area of relevance feedback (Rui et al. 1998), Rui et al. uses standard deviation (variance) to adjust the weight for each dimension in the feature space. Actually, the single-modal Gaussian model is adopted in their system. Gaussian distribution is the most often adopted assumption. Here we use Gaussian mixture distribution to represent the user's target more precisely.

Second, it is inspired from the existing applications of Gaussian mixture model in the area of CBIR. Several retrieval systems (Wu et al. 2000; Tian et al. 2000; Huang et al. 2002) treat the distribution of the whole image collection as a mix Gaussian distribution. Then they use some clustering algorithm to classify the image database. After the classification, each image in the database has a class label. They assume user is interested in one class of images and the images which belong to the same class as the query image will be returned as the retrieval result. This approach is efficient when user's desire is simply one of the classes. But under most circumstances, user's desire is more complex and this assumption is too rough to describe it. As we mentioned above, user's intention is always the combination of several components. Even it consists of only one component, it is rarely exactly one of the classes of image collection. For instance, one class of image collection is pictures of car, which is what user is looking for. But user may like cars with special color (such as black and deep blue) and model (such as hatchback but not wagon). They are only some subsets of the whole car class. In this case, returning one class of image database would not satisfy the user. Therefore, we discard the idea of using Gaussian mixture model to represent the classification of image database, but use it to represent the user's target distribution. This design can better capture the semantic meaning of user's mind, thus shorten the semantic gap from high level concept to low level features.

3.4.2 Resolving User Inconsistency

Along each feedback iteration, users give valuable information to the system. We are not going to use the only information given in one pass at each iteration. Instead, we will utilize the information given throughout the retrieval process. Therefore, if the user's feedback is not consistent, it will hurt the retrieval results. Most of the relevance feedback systems assume that the user is consistent when performing the relevance judgement. But in reality, user consistency is hard to achieve. There are often conflicts during each iteration. Roughly we summarize the user inconsistency into two classes and the resolutions have been designed separately.

One class of user inconsistency is relatively simple: the user gives the contradictive feedback at different iterations. That is, user indicates a specific image as relevant at one time then select the same image as irrelevant the next time. It is maybe the user's careless mistake or the change of his mind. For this kind of inconsistency, we have a straightforward resolution. Let R and I be the set of relevant images and the set of irrelevant images specified by the user in each iteration respectively. Let $rel(\underline{x}_i)$ be the measure of how relevant an image $\underline{x}_i$ is.

$$rel_{t+1}(\underline{x}_i) = rel_t(\underline{x}_i) + 1 \qquad \underline{x}_i \in R \tag{3.4.2.1}$$

$$rel_{t+1}(\underline{x}_i) = rel_t(\underline{x}_i) - 1 \qquad \underline{x}_i \in I \tag{3.4.2.2}$$

We only consider images of $rel(\underline{x}_i) > 0$ and $rel(\underline{x}_i) < 0$. Using equation (3.4.3.1) and equation (3.4.3.2), we can solve the contradicting feedbacks given by the user.

The other kind of user inconsistency is more complex. It is the shift of user's mind especially at the early stage of feedback, when user's desire has not been very clear. At the beginning of image retrieval, it is likely that user is not quite sure what exactly he wants to find. It is only an ambiguous idea in his mind. Then with more images displayed in front of him and he may get more hints from these pictures, his idea is becoming clear and he can decide what kind of images he really wants after several iterations of feedback. In this case, two issues are essential to handle this kind of user inconsistency: First, giving a correct judgement whether the inconsistency happens. Then tracking the user's new target as quick as possible. The strategies for judgement and tracking are closely related to our information updating and display updating strategy. The detailed description of our approach will be mixed in the next two sections.

3.4.3 Information Updating

We use the Gaussian mixture model for the user's target distribution. At each iteration of feedback, we will use the Expectation Maximization (EM) algorithm to estimate the parameters of the distribution. EM is a widely used standard algorithm for parameter estimation in statistics. One of the methods EM uses is maximizing the log-likelihood function $L(\Theta) = \sum_{i-1}^{N} \log f(\underline{x}_i \mid \Theta)$ to estimate the parameters of the statistical source, where N is the number of relevant images from each iteration of feedback. However, this maximization is hard to be resolved directly. To relax this concern, the missing data are incorporated. A set of class labels $Z = \{z^{(1)},...,z^{(N)}\}$ associated with relevant images, indicating which component produced each image, are treated as missing part. The EM algorithm can estimate the probability parameters by alternatively recomputing the expected value of z, $E(z)$, and the maximize likelihood parameter given the expected value of z. This process of estimation is an iterative hill climbing procedure. Each cycle revises the value of Θ so as to increase the likelihood until the optimal parameters are achieved. Going through the derivation, the update equations for distribution parameters can be obtained:

$$E(z_{ij}) = p_j(\underline{x}_i) = \frac{w_j f_j(\underline{x}_i \mid \theta_j)}{\sum_{k=1}^{K} w_k f_k(\underline{x}_i \mid \theta_k)} \tag{3.4.3.1}$$

$$w_j^{new} = \frac{\sum_{i=1}^{N} p_j(\underline{x}_i)}{N} \tag{3.4.3.2}$$

$$\mu_j^{new} = \frac{\sum_{i=1}^{N} p_j(\underline{x}_i) \cdot \underline{x}_i}{N} \tag{3.4.3.3}$$

$$\Sigma_j^{new} = \frac{\sum_{i=1}^{N} p_j(\underline{x}_i)[(\underline{x}_i - \mu_j^{new})(\underline{x}_i - \mu_j^{new})^T]}{\sum_{i=1}^{N} p_j(\underline{x}_i)} \tag{3.4.3.4}$$

where $E(z_{ij})$ is the expected value of the probability that data $\underline{x}_i$ belong to component *j* and $\sum_{i=1}^{N} p_j(\underline{x}_i)$ is the number of data points in component *j*. The process repeats until the convergence (the likelihood no longer increases) or until the number of iterations exceeds some predetermined number.

Instead of only utilizing the information provided by user's feedback at each iteration, we need to consider the connection between each iteration. This consideration is also necessary if we want to detect the user inconsistency. In our approach, at each step after getting the user's feedback, we use EM algorithm to estimate the distribution parameters twice. The differences of the two estimations lie in two aspects: the initialization values of parameters and the sample data being used. For the first method of estimation, in order to well utilize the information given throughout the retrieval process, we use all the relevant images accumulated from the previous relevance feedback as sample data, and initialize the parameter using the estimation values obtained from last iteration. In the second kind of estimation, we only adopt relevant images specified at current feedback as sample data and start EM with parameters initialization by other methods, such as initializing randomly, or initializing by *k*-means method, which are nothing to do with the previous estimations. Then we compare the parameters obtained by these two processes. If the distance of the two mean vector sets is larger than a threshold, we deem that user's target is changing. In this case, the data before the change are ignored. And from the next iteration, the sample data used in the first method are only accumulated from current feedback. In the following iterations, if the difference between the mean values decreases gradually and then less than the threshold for several iterations, we deem the user's intention becomes clear. Afterwards, we only need to estimate parameters by the first method. Another advantage of initializing the parameter using the previous estimations in the first method is that we can treat the whole feedback process as another "big" EM procedure. Every "small" EM executed at each iteration can be considered as one step of the "big" EM as well. In this way, we fully utilize the hill climbing character of EM. This character is very suitable to be applied in the feedback process, since the whole process of relevance feedback is also an iterative procedure.

However, there still exist some problems need to be considered when using EM for parameter estimation. First, since the EM algorithm maximizes the likelihood of the observation from the source, we must regard only the relevant images as the observation and discard the non-relevant images. Therefore, the information from the non-relevant images cannot be well utilized. Second, to achieve robust and accurate estimates by the EM algorithm, the sample size from the observation should be large. However, the number of the relevant images from each iteration of the user's feedback is relatively small compared to the size of the feature space. Learning from a small data set cannot guarantee the robust performance. All these problems will be addressed in the future research.

3.4.4 Display Updating

After the distribution of user's target is updated at each iteration, the next task of the relevance feedback is to select the display image set on the basis of the estimated distribution. There are two kinds of display updating strategies used in the existing image retrieval systems: the *most probable strategy* and the *most informative strategy*.

Most probable strategy means the system always display the images with the highest similarity to the query image. The advantage of this method is that it minimizes the number of false matches, thus the user can always get the best results and can terminate the query at any time. However, this strategy selects the images in a narrow region near to the relevant images without exploring the entire feature space. It is not a good schema at the early stage of relevance feedback if the user's intention has not been very clear yet. Because it is not helpful for the user to get some hints and clarify what he really wants. By contrast, *most informative strategy* selects the images in the unexplored region on the purpose of reducing the uncertainty of the user's intention. It retrieves the images dissimilar to the relevant images on purpose, which increases the false matches but allows clarifying the user's desire. The extreme example of this scheme is displaying the images randomly. A very successful application of *most informative strategy* can be found in (Cox et al.1998, 2000).

We intend to combine these two display strategies to a hybrid one. Our criterion for the selection of the display image set is evolved from the most informative to the most probable as the retrieval process progresses. That is, the scheme relies on the most informative scheme at first and shifts the emphasis to the most probable scheme, as the iterations of the relevance feedback are progressed. By this approach, at the early stage of relevance feedback it is helpful for the user to clarify his intention and it reduces the chance of converging to a local maximum of the probability parameters. But at the final stage of estimation, if we still using most informative to search in a wide range, it may result in oscillation around the true optimum parameters estimation and reduce the speed of convergence. Therefore, at the final stage of estimation, we narrow down the selection region to accelerate the convergence and the user can get best results at the same time.

3.5. Performance Evaluation

Currently, there is no standard image collection available for all the researchers to test their system, so we can use our own image collection. First we will use a small image database with more than 200 images to test the performance of the designed system. And during the process of experiments, we may also need to use some simulation data generated by software like Matlab. Finally we will test the performance of our CBIR system by an image library which contains more than 10,000 images to present a more comprehensive performance evaluation. Many different methods for measuring the performance of a CBIR system have been used by researches. Among them, the most common one is *precision* and *recall*. We will use *precision* and *recall* graph as our measure method. *Precision* is the ratio of relevant images to the total number of images retrieved and *recall* is the percentage of relevant images among all possible relevant images. Either value alone contains insufficient information. By the graph of precision vs. *recall*, an objective performance evaluation of our CBIR system can be given.

4. Preliminary Work

I have already implemented a content-based image retrieval system with re-weighting relevance feedback technique. The current system provides a CGI-based web access interface. The main graphical interface of our image retrieval system is shown in Fig2. The system supports all the options metioned in section 3.1 to specify a query.

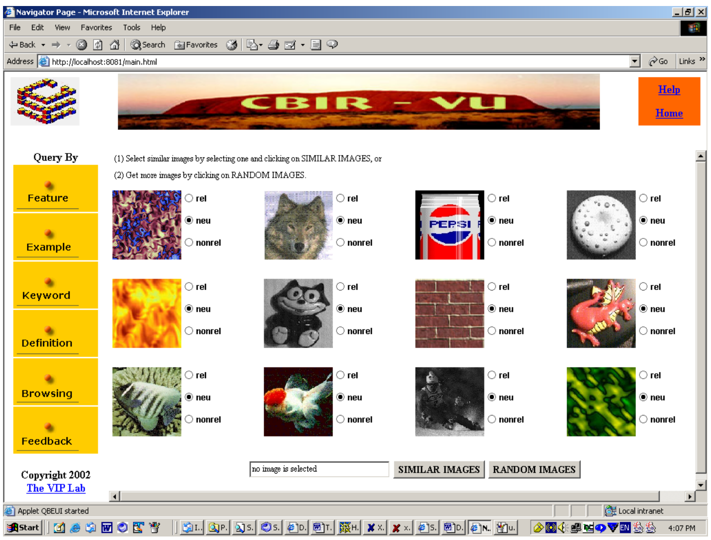

Fig2. The Main Graphical User Interface of CBIR System

This system also allows user to refine the retrieval results by adjusting the weights associated with each features and each dimension in the feature space. An image database of 220 images is applied on various queries. A typical retrieval process on our test database is given in Figs 3. and 4. The query image is displayed at the upper left corner, and the best 11 matched images are presented in order from top to bottom and from the left to right.

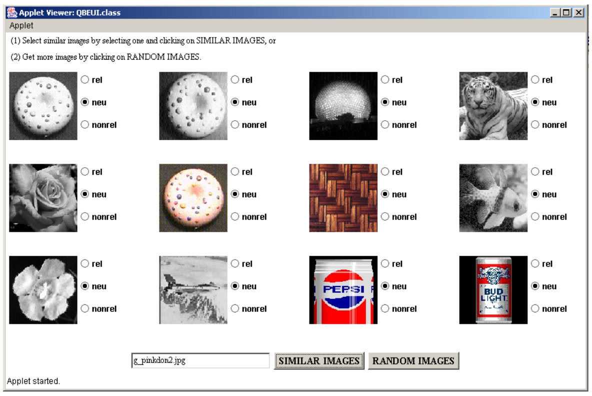

Fig 3. Retrieval results befor the relevance feedback

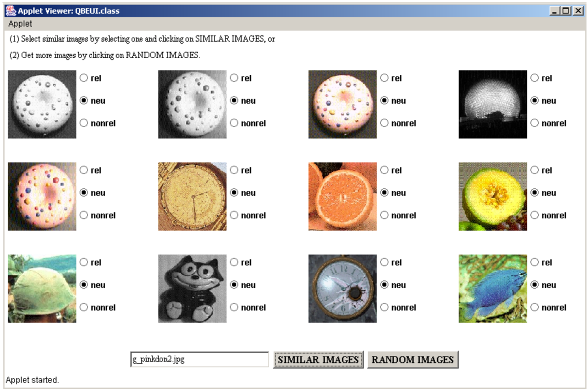

Fig 4. Retrieval results after the relevance feedback

This example assumes the user's true information need is to retrieve similar images based on their shapes. The initial retrieval results are images similar to the query image on all the color, text and shape features. After learning the relevance feedback given by the user, we can see that the system put more emphsis on the shape feature and more relevant images which match the user's need are retrieved.

5. Time-Line

Roughly the timetable for completing the thesis is as following:

- Phase-1: 07/2002 - 06/2003
  - Familiarize with the area of CBIR
  - Proposal for the future work
  - Build the basic CBIR system with existing relevance feedback technique using small image database
- Phase-2: 07/2003 – 08/2004
  - Further design and implement the proposed relevance feedback technology
  - Implement the resolution for user inconsistency
  - Test on both simulated data and small image database
- Phase-3: 09/2003 – 11/2004
  - Expand the CBIR system using large image library
  - Evaluate the effectiveness and efficiency of the whole system
- Phase-4: 12/2004 – 04/2005
  - Finish writing thesis

5. Reference

Alfred C., She A.C. and Huang T. S., 1994, Segmentation of road scenes using color and fractal-based texture classification, In Proc. IEEE Int. Conf. on Image Proc., Austin.

Bach J. R., Fuller C., Gupta A., Hampapur A., Horoqitz B., Humphrey R., Jain R. and Shu C. F., 1996, The Virage image search engine: An open framework for image management, SPIE Storage and Retrieval for Image and Video Databeses, pages 63-68.

Beckmann N., Kriegel H. P., Schneider R., and Seeger B., 1990, The R*-tree: an efficient and robust access method for points and rectangles, In Proc. ACM SIGMOD, Atlantic City.

Benitez A. B., Beigi M. and Chang S. F., 1998, Using relevance feedback in content-based image metasearch. IEEE Int. Computing.

Berchtold S., Keim D. A., Kriegel H. P., 1996, The X-Tree: An Index Structure for High-Dimensional Data, Proc. 22th Int. Conf. on Very Large Data Bases, Bombay, India.

Carson C., Belongie S., Greenspan H. and Malik J., 1997, Region-based image querying, In CVPR'97, pages 42-51.

Carson C., Belongie S., Greenspan H. and Malik J., 1999, Blobworld: A system for region-based image indexing and retrieval, Third Int. Conf. on Visual Information Systems, Amsterdam, The Netherlands.

Carson C., Belongie S., Greenspan H. and Malik J., 2002, Color- and texture-based segmentation using EM and its application to image querying and classification, IEEE Transactions on Pattern Analysis and Machine Intelligence.

Charikar M., Chekur C., Feder T., and Motwani R., 1997, Incremental clustering and dynamic information retrieval, In Proc. of the 29th Annual ACM Symposium on Theory of Computing, pages 626-635.

Chandrasekaran S., Manjunath B. S., Wang Y. F., Winkeler J., and Zhang H., 1997, An eigenspace update algorithm for image analysis, Comput. Vis., Graphics, and Image Proc.

Chang T. and Kuo C. C. J., 1993, Texture analysis and classification with tree-structured wavelet transform, IEEE Trans. Image Proc., 2(4): 429-441.

Chen Y, Zhou X. S. and Huang T.S., 2001, One-class SVM for learning in image retrieval. Int'l Conf on Image Processing, Greece.

Chua T. S., Tan K. L., and Ooi B., 1997, Fast signature-based color-spatial image retrieval, In Proc. IEEE Conf. on Multimedia Comput. and Syss.

Copper D. and Lei Z., 1995, On representation and invariant recognition of complex objects based on patches and parts. In Spinger Lecture Notes in Computer Science series, 3D Object Representation for Computer Vision, New York: Springer, pp.139-153.

Cox I. J., Ghosn J., Miller L., Papathomas V. and Yianilos P., 1997, Hidden annotation in conten based image retrieval, In IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'97), pages 76-81.

Cox I. J., Miller L., Minka P. and Yianilos P., 1998, An optimized interaction strategy for Bayesian relevance feedback. IEEE Conf. Computer Vision and Pattern Recognition, Santa Barbara, CA.

Cox I. J., Miller L., Minka P., Papathomas V., and Yianilos P., 2000, The Bayesian image retrieval system, PicHunter: Theory, Implementation and Psychophysical Experiments. IEEE Transaction on Image Processing, Vol 9, No. 1, 20-37.

Dowe J., content-base retrieval in multimedia imaging, 1993, In Proc. SPIE Storage and Retrieval for Image and Video Database.

Faloutsos C., Flickner M., Niblack W., Petkovic D., Equitz W., and Barber R., 1993, Efficient and effective querying by image content, IBM Research Report.

Faloutsos C. and Lin K. D., 1995, Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets, In Proc. of SIGMOD, pages 163-174.

Fauqueur J. and Boujemaa N., 2002, Image Retrieval by Regions: Coarse Segmentation and Fine Color Description, In Proc. of VISUAL, pages 24-35.

Flicker M., Sawhney H., Niblack W., Ashley J., Huang Q., Dom B., Gorkani M., Hafner J., Lee D., Petkkovic D., Stele D. and Yanker P., Query by Image and Video Content: The QBIC system, IEEE Computer, 28(9): 23-32.

Gevers T. and Kajcovski V. K., 1994, Image segmentation by directed region subdivision, In Proc. IEEE Int. Conf. on Image Proc.

Gross A. and Latecki L., 1995, Digital Geometric Invariance and Shape Representation. In Proc. of the IEEE International Symposium on Computer Vision (ISCV 95), Florida, USA.

Guttman A., 1984, R-tree: a dynamic index structure for spatial searching, In Proc. ACM SIGMOD.

Hansen M. and Higgins W., 1994, Watershed-driven relaxation labelling for image segmentation, In Proc. IEEE Int. Conf. on Image Proc.

Haralick R. M., Shanmugam K., and Dinstein I., 1973, Texture features for image classification, IEEE Trans. on Sys, Man, and Cyb, SMC-3 (6): 610-621.

Hong P., Tian Q. and Huang T.S., 2000, Incorporate Support Vector Machines to Content-Based Image Retrieval with Relevance Feedback, IEEE Int'l Conf. on Image Proc. (ICIP'2000), Vancouver Canada.

Hu M. K., 1977, Visual pattern recognition by moment invariants, In Computer Methods in Image Analysis, Los Angeles: IEEE Computer Society.

Huang J., Kumar S., Mitra M., Zhu W. J., and Zabih R., 1997, Image indexing using color correlogram, In Proc. IEEE Conf. on Comput. Vis. and Patt. Recog..

Huang T. S., Mehrotra S. and Ramchandran K., 1996, Multimedia analysis and retrieval system (MARS) project, in Proc. of 33rd Annual Clinic on Library Application of Data processing - Digital Image Access and Retrieval.

Huang T. S., Zhou X. S., Nakazato M., Wu Y. and Cohen I., 2002, Learning in Content-Based Image Retrieval, Proc. of International Conf. on Development and Learning (ICDL'02), Cambridge, Massachusetts.

Ioka M., 1989, A method of defining the similarity of images on the basis of color information. Technical Report RT-0030, IBM Research, Tokyo Research Laboratory.

Ishikawa Y., Subramanya R., and Faloutsos C., 1998, "MindReader: Query databases through multiple examples", in Proc. Of the 24th VLDB Conf., New York.

Jing F., Li M., Zhang H. J. and Zhang B., 2003, Learning region weighting from relevance feedback in image retrieval, Proc. International Conference on Image and Video Retrieval.

Kapur D., Lakshman Y. N., and Saxena T., 1995, Computing invariants using elimination methods, In Proc. IEEE Int. Conf. on Image Proc.

Ko B., Peng, J., and Byun H., 2001, Region-Based Image Retrieval Using Probabilistic Feature Relevance Learning, Pattern Analysis and Applications , Vol. 4, pp.174-184.

Kurniawati R., Jin J. S., Shepherd J. A., The SS$^+$-tree: An improved index structure for similarity searches in a high-dimensional feature space. Storage and Retrieval for Image and Video Databases (SPIE) 1997: 110-120

Lei Z., Keren D., and Cooper D. B., 1995, Computationally fast Bayesian recognition of complex objects based on mutual algebraic invariants, In Proc. IEEE Int. Conf. on Image Proc.

Li B. and Ma S. D., 1994, On the relation between region and contour representation, In Proc. IEEE Int. Conf. on Patt. Recog.

Li X. Q., Zhao Z. W., Cheng H. D., Huang C. M., and Harris R. W., 1994, A fuzzy logic approach to image segmentation, In Proc. IEEE Int. Conf. on Image Proc.

Lu H., Ooi B., and Tan K., 1994, Efficient image retrieval by color contents, In Proc. of the 1994 Int. Conf. on Applications of Databases.

Lybanon M., Lea S., and Himes S., 1994, Segmentation of diverse image types using opening and closing, In Proc. IEEE Int. Conf. on Image Proc.

Ma W. Y. and Manjunath B. S., 1995, Edge a framework of boundary detection and image segmentation, In Proc. IEEE Conf. on Comput. Vis. and Patt. Recog.

MacArthur S. D., Brodley C. E. and Shyu C., 2000, Relevance feedback decision trees in content-based image retrieval. IEEE Workshop CBAIVL, South Carolina.

Mehtre B. M., Kankanhalli M., Lee W. F., 1997, Shape measures for content-based image retrieval: A comparison, Information Processing & Management, 33(3): 319-337.

Mehrotra S., Chakrabarti K., Ortega M., Rui Y. and Huang T. S., Multimedia analysis and retrieval system, 1997, In Proc. of the 3rd Int. Workshop on Information Retrieval Systems.

Minka T. P. and Picard R. W., 1996, Interactive learning using a “society of models”, In Proc. IEEE Conf. on Comput. Vis. And Patt. Recog..

Miyahara M., 1988, Mathematical transform of (R,G,B) color data to Munsell (H,S,V) color data, In Proc. SPIE Visual Communications and Image Processing, volume 1001, pages 650-657.

Ng R. and Sedighian A., 1996, Evaluating multi-dimensional indexing structures for images transformed by principal component analysis, In Proc. SPIE Storage and Retrieval for Image and Video Databases.

Niblack W., Barber R., and et al., 1994, The QBIC project: Querying images by content using color, texture and shape, In Proc. SPIE Storage and Retrieval for Image and Video Databases.

Ortega M., Rui Y., Chakrabarti K., Mehrotra S. and Huang T. S., Supporting similarity queries in MARS, 1998, In Proc. of the Workshop on Digital Library Metrics at ACMDLM’98, Pittsburgh, PA.

Ortega M., Rui Y., Chakrabarti K., Porkaew K., Mehrotra S. and Huang T. S., Supporting ranked Boolean similarity queries in MARS, 1998, IEEE Transactions on Knowledge and Data Engineering, 10(6): 905-925.

Pass G., Zabih R., and Miller J., 1996, Comparing images using color coherence vectors, In Proc. ACM Conf. on Multimedia.

Persoon E. and Fu K. S., 1977, Shape discrimination using Fourier descriptors, IEEE Trans. Sys. Man, Cyb.

Peng J, Bhanu B and Qing S, 1999, Probabilistic feature relevance learning for content-based image retrieval. Computer Vision and Image Understanding 75: 150-164.

Pentland A., Picard R. W. and Sclaroff S., Photobook: Content-base manipulation of image databases, 1996, Int. J. on Comput. Vis., Vol 18, No. 3, pp. 233-254.

Picard R. W., Minka T. P. and Szummer M., 1996, Modeling user subjectivity in image libraries, In Proc. IEEE Int. Conf. on Image Proc.

Porkaew K, Mehrotra S and Ortega M, 1999, Query reformulation for content based multimedia retrieval in MARS. IEEE Int'l Conf. Multimedia Computing and Systems.

Ratan A. L., Maron O., Grimson W. and Lozano-Perez T., 1999, A framework for learning query concepts in image classification. IEEE Conf. Computer Vision and Pattern Recognition, CO.

Rickman R. and Stonham J., 1996, Content-based image retrieval using colour tuple histograms, In Proc. SPIE Storage and Retrieval for Image and Video Databases.

Rui Y., Alfred C., She A.C., and Huang T. S., 1996, Automated shape segmentation using attraction-based grouping in spatial-color-texture space, In Proc. IEEE Int. Conf. on Image Proc..

Rui Y., She A.C., and Huang T. S., 1996, Modified fourier descriptors for shape representation-a practical approach, In Proc. of First International Workshop on Image Databases and Multi Media Search.

Rui Y., Chakrabarti K., Mehrotra S., Zhao Y., and Huang T. S., 1997a, Dynamic clustering for optimal retrieval in high dimensional multimedia databases, In TR-MARS-10-97.

Rui Y., Chakrabarti K., Mehrotra S., Zhao Y., and Huang T. S., 1997b, Content-based image retrieval with relevance feedback in MARS, In Proc. of IEEE Int. Conf. On Multimedia Computing and Systems.

Rui Y., Huang T. S., Ortega M., and Mehrotra S., 1998, "Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval", IEEE Tran on Circuits and Systems for Video Technology, Vol 8, No. 5, 644-655

Rui Y. and Huang T. S., 2000, Optimizing learning in image retrieval. IEEE Conf. Computer Vision and Pattern Recognition, South Carolina.

Samadani R. and Han C., 1993, Computer-assisted extraction of boundaries from images, In Proc. SPIE Storage and Retrieval for Image and Video Databases.

Salton G., 1989, Automatic text processing. Reading, Mass., Addison-Wesley.

Santini S. and Jain R., 1998, Direct manipulation of image databases, Technical report, Department of Computer Science, University of California San Diego, San Diego California.

Santini S. and Jain R., 2000, Integrated browsing and querying for image database. IEEE Trans. Multimedia 7(3).

Sellis T., Roussopoulos N., and Faloutsos C., 1987, The R^+ tree: A dynamic index for multi-dimensional objects, In Proc. 12th VLDB.

She A. C. and Huang T. S., 1994, Segmentation of road scenes using color and fractal-based texture classification, In Proc. IEEE Int. Conf. on Image Proc., Austin.

Smith J.H. and Chang S. F., 1994, Transform features for texture classification and discrimination in large image databases, In Proc. IEEE Int. Conf. on Image Proc..

Smith J. R. and Chang S. F., 1995a, Single color extraction and image query, In Proc. IEEE Int. Conf. on Image Proc..

Smith J. R. and Chang S. F., 1995b, Tools and techniques for color image retrieval, In IS & T/SPIE Proc. Vol.2670, Storage & Retrieval for Image and Video Databases IV.

Smith J. R. and Chang S. F., 1996a, Automated binary texture feature sets for image retrieval. In Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc., Atlanta, GA.

Smith J. R. and Chang S. F., 1996b, Visualseek: A fully automated conten-based image query system, In Proc. ACM Multimedia.

Smith J. R. and Chang S. F., 1997, Visual searching the web for content, IEEE Multimedia Mag., Vol 4, pp.12-20.

Stricker M. and Markus Orengo M., 1995, Similarity of color images, In Proc. SPIE Storage and Retrieval for Image and Video Databases.

Stricker M. and Dimai A., 1996, Color indexing with weak spatial constraints, In Proc. SPIE Storage and Retrieval for Image and Video Databases.

Swain M. and Ballard D., 1991, Color indexing, Int. J. Comput. Vis., 7(1): 11-32.

Tagare H., 1997, Increasing retrieval eciency by index tree adaption, In Proc. of IEEE Workshop on Content-based Access of Image and Video Libraries, in conjunction with IEEE CVPR '97. Tamura H., Mori S., and Yamawaki T., 1978, Texture features corresponding to visual perception, IEEE Trans. on Sys, Man, and Cyb, SMC-8 (6): 460-473. Tieu K. and Viola P., 2000, “Boosting Image Retrieval”, IEEE Conf Computer Vision and Pattern Recognition (CVPR'00), Hilton Head, South Carolina.

Tamura H., Mori S., and Yamawaki T., 1978, Texture features corresponding to visual perception, IEEE Trans. On Sys, Man, and Cyb, SMC 8(6): 460-473.

Tian Q., Wu Y. and Huang T. S, 2000, Incorporate Discriminant Analysis with EM Algorithm in Image Retrieval, In Proc. IEEE Int'l Conf. on Multimedia and Expo (ICME'2000), New York.

Tieu K. and Viola P., 2000, Boosting image retrieval, IEEE Conf. Computer Vision and Pattern Recognition, South Carolina.

Vasconcelos N. and Lippman A., 2000, Learning from user feedback in image retrieval. Adv. In Neural information Processing Systems, MIT Press.

Veltkamp R. C., Tanase M., 2000, Content-based image retrieval systems: A survey, Technical Report UU-CS-2000-34.

Wang J., Yang W. J., and Acharya R., 1997, Color clustering techniques for color-content-based image retrieval from image databases, In Proc. IEEE Conf. on Multimedia Computer and Systems.

White D. and Jain R., 1996, Similarity indexing: Algorithms and performance, In Proc. SPIE Storage and Retrieval for Image and Video Databases.

Wu Y., Tian Q., Huang T. S., 2000, Discriminant EM Algorithm with Application to Image Retrieval, IEEE Conf. Computer Vision and Pattern Recognition (CVPR'00), Hilton Head Island, South Carolina.

Yang L. and Algregtsen F., 1994, Fast computation of invariant geometric moments: A new method giving correct results, In Proc. IEEE Int. Conf. on Image Proc.

Zhang H. and Zhong D., 1995, A scheme for visual feature based image retrieval, In Proc. SPIE Storage and Retrieval for Image and Video Databases.

Zhou X. S. and Huang T. S., 2000, A generalized relevance feedback scheme for image retrieval, SPIE Int’l Conf. on Internet Multimedia Management Systems, Boston, MA.

Zhou X. S. and Huang T. S., 2001, Small sample learning during multimedia retrieval using BiasMap, IEEE Conf. Computer Vision and Pattern Recognition, Hawaii.