Abstract—An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Image retrieval is one of the most exciting and efficient growing research area. In this paper, we present a highlight of recent research for image retrieval. The choice of a color system is of great importance for the purpose of proper image retrieval. This paper presents the functionality of a system which retrieve images on the basis of similarity measures and indexes.

Keywords—Color Histogram, Image retrieval, Minkowski-Form Distance and Quadratic Form (QF) Distance.

I. INTRODUCTION

Image retrieval is the field of study concerned with searching and retrieving digital images from a collection of database. Content-based image retrieval plays a central role in the application area such as multimedia database systems in recent years [1]. The emergence of multimedia technology and the rapidly expanding image and video collections on the Internet have attracted significant research efforts in providing tools for effective retrieval and management of visual data [2]. The fundamental idea of this approach is to generate automatically image descriptions directly from the image content by analyzing the content of the images [3].

Textual information is linear while images are bi-dimensional, and videos are three dimensional (one dimension is time). More precisely, text is provided with an inherent starting and ending point, and with a natural sequence of parsing. Such a natural parsing strategy is not available for images and videos. Problems with text-based access to images and videos have prompted increasing interest in the development of feature-based solutions. The main features used for image retrieval are color, texture and shape.

Color is the first and most straightforward visual feature for indexing and retrieval of images (Swain and Ballard, relatively robust and simple to represent). It is also the most commonly used feature in the field. Color has been an active area of research in image retrieval, more than in any other branch of computer vision. Color makes the image take values in a color vector space.

The choice of a color system is of great importance for the purpose of proper image retrieval. An important criterion is that the color system is independent of the underlying imaging device. This is required when images in the image database are recorded by different imaging devices such as scanners, camera's and cam recorder (e.g. images on Internet). Another prerequisite might be that the color system should exhibit perceptual uniformity meaning that numerical distances within the color space can be related to human perceptual differences. This is important when images are to be retrieved which should be visually similar (e.g. stamps, trademarks and paintings databases).

II. IMAGE RETRIEVAL SYSTEM

In image retrieval system, each image is stored in the database has its feature extracted and compared to the features of the query image. Then indexing techniques are utilized. In order to help the users retrieve the correct image they seek, relevance feedback techniques have been implemented [4].

Fig. 1 Diagram for content-based image retrieval system

III. FEATURE EXTRACTION

Feature extraction is the basis of content-based image retrieval. This involves extraction of the image features at a distinguishable extent. For color based image retrieval, color can also be represented by numerous of ways [5]. Most commonly used color descriptors are: Color moments, color histograms, color coherence vector, color correlogram.

A. Color Moments
Color moments are the statistical moments of the probability distributions of colors. Color moments used especially when image contain just the objects. The first order (mean), the second (variance) and the third order (skewness) color moments have been proved to be effective and efficient in representing color distribution of images.

\[ \mu_i = \frac{1}{N} \sum_{j=1}^{N} f_{ij} \]

\[ \sigma_i = \left( \frac{1}{N} \sum_{j=1}^{N} (f_{ij} - \mu_i)^2 \right)^{1/2} \]

\[ s_i = \left( \frac{1}{N} \sum_{j=1}^{N} (f_{ij} - \mu_i)^3 \right)^{1/2} \]

Where \( f_{ij} \) is the value of the \( i \)-th color component of the image pixel \( j \), and \( N \) is the total number of pixels in the image.

B. Color Histogram

Color histogram serve as an effective representation of the color content of an image. The color histogram is easy to compute and effective in characterizing both the global and local distribution of colors in an image. A histogram is the distribution of number of pixels for each quantized bins. However, more number of bins not only increases the computational costs. To address this problem, opponent color space can be used. Another way is to adopt clustering algorithm or to use bins that have largest pixel number. Also, color histogram does not take the spatial information of pixels into consideration. So, image is divided into sub-parts and the histogram is obtained for each sub-part. The division can be as simple as a rectangular partition and complex as a region or even object segmentation.

C. Color coherence vector

In [6] a different way of incorporating spatial information into color histogram, color coherence vectors, was purposed. Each histogram bin is partitioned into two types, i.e., coherent, if it belongs to a large uniformly-colored region, or incoherent, if it does not. Let \( a_i \) is a value for coherent pixels and \( b_i \) is the value of non-coherent pixels, then CCV of the image is defined as \( <a_1, b_1>, <a_2, b_2>, \ldots, <a_N, b_N> \) where color histogram of the image is \( <a_1+b_1, a_2+b_2, \ldots, a_N+b_N> \). Due to additional spatial information, CCV provides better results than the color histogram.

D. Color Correlogram

The color correlogram [7] was purposed to characterize not only the color distribution of pixels, but also the spatial correlation of pairs of colors. The first and the second dimension of the 3-D histogram are the colors of any pixel pair and the third dimension is their spatial distance. Color correlogram is a table indexed by color pairs, where k-entry for \((i,j)\) specifies probability of finding a pixel of color “j” at distance “k” from pixel “i” in the image. Let \( I_c(i) \) represent the entire set of image pixels and \( I_c(j) \) represent the set of pixels.

Whose colors are \( c(i) \). Then, the color correlogram is defined as:

\[ \gamma_{ij} = \Pr \{ p2 \in I_c(j) \mid p1 - p2 = k \} \]

Where \( i, j \in \{1,2,\ldots,N\}, k \in \{1,2,\ldots,d\} \), and \( p1 - p2 \) is the distance between pixels \( p1 \) and \( p2 \). Its simplified version called color autocorrelogram, is often used. The color autocorrelogram only captures the spatial correlation between identical colors and thus reduces the dimension.

Compared to the color histogram and CCV, the color autocorrelogram provides better retrieval results. but is also the most computational expensive due to its high dimensionality.

IV. SIMILARITY MEASURES

Once features of images in the database are extracted and the user’s query is formed, the search results are obtained by measuring the similarity between the pre-extracted features of the image database and the

\[ \delta(I,J) = \frac{\sum_{i=1}^{N} \min(f_i(I), f_i(J))}{\sum_{i=1}^{N} f_i(J)} \]

analyzed user’s query. Instead of exact matching, content-based image retrieval calculates visual similarities between a query image and images in a database. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years. We denote \( D(I,J) \) as the distance measure between the query image \( I \) and the image \( J \) in the database; and \( f_i(I) \) as the number of pixels in bin \( i \) of \( I \).

Minkowski-Form Distance

Minkowski-form distance is the most widely used metric for image retrieval. If each dimension of image feature vector is independent of each other and is of equal importance, the Minkowski-form distance \( L_p \) is appropriate.

\[ D(I,J) = \left( \sum_{i=1}^{N} |f_i(I) - f_i(J)|^p \right)^{1/p} \]

when \( p = 1, 2, \) and \( \infty \), \( D(I,J) \) is the \( L1 \), \( L2 \) (also called Euclidean distance), and \( L\infty \) distance respectively. The Histogram intersection can be taken as a special case of \( L1 \) distance, which is used by Swain and Ballard [8] to compute the similarity between color images.

\[ S(I,J) = \frac{\sum_{i=1}^{N} \min(f_i(I), f_i(J))}{\sum_{i=1}^{N} f_i(J)} \]

Quadratic Form (QF) Distance

The Minkowski distance treats all bins of the feature histogram entirely independently and does not account for the fact that certain pairs of bins correspond to features which are perceptually more similar than
other pairs. To solve this problem, quadratic form distance is introduced:

\[
D(i,j) = \sqrt{(F_i - F_j)^T A (F_i - F_j)}
\]

where \(A=[aij]\) is a similarity matrix, and \(aij\) denotes the similarity between bin \(i\) and \(j\). \(F_i\) and \(F_j\) are vectors that list all the entries in \(fi(I)\) and \(fi(J)\). Quadratic form distance can lead to perceptually more desirable results than Euclidean distance.

**Mahalanobis Distance** The Mahalanobis distance metric is appropriate when each dimension of image feature vector is dependent of each other and is of different importance. It is defined as:

\[
D(i,j) = \sqrt{(F_i - F_j)^T C^{-1} (F_i - F_j)}
\]

where \(C\) is the covariance matrix of the feature vectors. The Mahalanobis distance can be simplified if feature dimensions are independent. In this case, only a variance of each feature component, \(Ci\), is needed.

\[
D(i,j) = \frac{\sum_{k=1}^{N} (F_i - F_j)^2}{Cl}
\]

**Kullback-Leibler (KL) Divergence and Jeffrey-Divergence (JD)** The Kullback-Leibler (KL) divergence measures how compact one feature distribution can be coded using the other one as the codebook. The KL divergence between two images \(I\) and \(J\) is defined as:

\[
D(i,j) = \sum_{l} F_l(I) \log \frac{F_l(I)}{F_l(J)}
\]

The Jeffrey-divergence (JD) is defined by:

\[
D(i,j) = \sum_{l} F_l(I) \log \frac{F_l(I)}{F_l(I)} + F_l(J) \log \frac{F_l(J)}{F_l(I)}
\]

where \(F_l=[fi(I)+fi(J)]/2\). In contrast to KL-divergence, JD is symmetric and numerically more stable when comparing two empirical distributions.

**Weighted-Mean-Variance** The Weighted-Mean-Variance was proposed in (Manjunath and Ma, 1996). This distance is defined by:

\[
\text{Dist WMV} = \frac{\mu_1 - \mu_2}{\sigma(\mu)} + \frac{\sigma_1 - \sigma_2}{\sigma(\sigma)}
\]

where \(\mu_1, \mu_2\) are the empirical means and \(\sigma_1, \sigma_2\) are the standard deviations of the two histogram \(h_1, h_2\). \(\sigma(.)\) denotes an estimate of the standard deviation of the respective entity.

V. **INDEXING SCHEME**

Another important issue in content-based image retrieval is effective indexing and fast searching of images based on visual features. The feature vectors of images tend to have high dimensionality and therefore dimension reduction is usually used before setting up an efficient indexing scheme [5].

For dimension reduction, Principal component analysis (PCA) is commonly used. It is an optimal technique that linearly maps input data to a coordinate space such that the axes are aligned to reflect the maximum variations in the data.

In addition to PCA, many researchers have used Karhunen-Loeve (KL) transform to reduce the dimensions of the feature space. Although the KL transform has some useful properties such as the ability to locate the most important sub-space, the feature properties that are important for identifying the pattern similarity may be destroyed during blind dimensionality reduction [9].

Apart from PCA and KL transformation, neural network has also been demonstrated to be a useful tool for dimension reduction of features [10].

After dimension reduction, the multi-dimensional data are indexed. A number of approaches have been proposed for this purpose, including R-tree (particularly, R*-tree [11]), linear quad-trees [12], K-d-B tree [13] and grid files [14]. These indexing schemes assume that the underlying feature comparison is based on the Euclidean distance, which is not necessarily true for many image retrieval applications. One attempt to solve the indexing problems is to use hierarchical indexing scheme based on the Self-Organization Map (SOM) proposed in [15]. In addition to benefiting indexing, SOM provides users a useful tool to browse the representative images of each type.

VI. **RELEVANCE FEEDBACK**

Human perception of image similarity is subjective, semantic, and task-dependent. Although content-based methods provide promising directions for image retrieval, generally, the retrieval results based on the similarities of pure visual features are not necessarily perceptually and semantically meaningful. In addition, each type of visual feature tends to capture only one aspect of image property and it is usually hard for a user to specify clearly how different aspects are combined. To address these problems, interactive relevance feedback, a technique, was introduced.

The iterative and automatic refinement of a query is known as relevance feedback in information retrieval literature [16]. Relevance feedback can be seen as a form of supervised learning to adjust the subsequent queries using the information gathered from the user’s feedback [17].

In implementing relevance feedback, three minimum requirements need to be fulfilled. First, the system must show the user a series of images, remember what images have already been shown, and not to display them again. Second, the user must somehow indicate which images are to some extent relevant to the present query and which are not. We call them here positive and negative seen images, respectively. The third requirement, the system must change its behavior depending on which images are included in the positive and negative image
sets. During the retrieval process more and more images are accumulated in these two image sets and the system has increasing amount of data to use in retrieving the succeeding image sets. The art of relevance feedback is finding the ways which use this information most efficiently [16].

VII. CONCLUSION

Color is usually represented by the color histogram, color correlogram, color coherence vector, and color moment under a certain color space. There are various ways to calculate the similarity distances. Up to now, the Minkowski and Quadratic form distance are the most commonly used distances for image retrieval. To set up an indexing scheme, dimension reduction is usually performed to reduce the dimensionality of the visual feature vector. Query results can be refined through the relevance feedback of users.

Although color-based image retrieval provides an intelligent and automatic solution the majority of current techniques are based on low level features. In general, each of these low level features tends to capture only one aspect of an image property. Neither a single feature nor a combination of multiple features has explicit semantic meaning. Although relevance feedback provides a way of filling the gap between semantic searching and low-level data processing, this problem remains unsolved and more research is required.

REFERENCES


