ABSTRACT
This paper gives an overview idea of retrieving images from a large database. CBIR is used for automatic indexing and retrieval of images depending upon contents of images known as features. The features may be low level or High level. The low level features include color, texture and shape. The high level feature describes the concept of human brain. The difference between low level features extracted from images and the high level information need of the user known as semantic gap. A Single feature can represent only part of the image property. So multiple features are used to enhance the image retrieval process. This paper has used color histogram, color mean, color structure descriptor and texture for feature extraction. The feature matching procedure is based on their Euclidean distance.

Keywords – Content Based Image Retrieval (CBIR), Conventional Color Histogram(CCH), Color Structure Descriptor,(CSD), Text Based Image Retrieval(TBIR)

INTRODUCTION
With the development of the Internet, and the availability of image capturing devices such as digital cameras, image scanners, the size of digital image collection is increasing rapidly. It is very important to efficiently store and retrieve images for different application such as fashion design, crime prevention, medicine, architecture, etc. For this purpose, many general purpose image retrieval systems have been developed. They are text-based and content-based.

The idea of text-based approach was originated at 1970s.In this approach the images are manually annotated by text descriptors, which are then used by a database management system (DBMS) to perform image retrieval. It has lead to two disadvantages. First one is that a considerable level of human labor is required for manual annotation. The second is the annotation inaccuracy due to the subjectivity of human perception. To overcome the above disadvantages in text-based retrieval system, content-based image retrieval (CBIR) was introduced in the early 1980s. In CBIR, images are indexed by their visual content, such as color, texture, shapes. The CBIR mainly consists of two steps. One is the feature extraction and another one is the similarity matching.

The difference between the user’s information need and the image representation is called the semantic gap in CBIR System. The system is said to be efficient if this semantic gap is minimum.

Section II describes the different methodology used for retrieval of images. Section III gives the overview of proposed system. Experimental Result is discussed in Section IV. Section V describes the conclusion of the paper.

II. METHODOLOGY
There are various approaches are present for Content-Based Image Retrieval. Some of the important literature which covers the more important CBIR System is discussed below.

Chin-Chin Lai et.al. [1] have proposed an interactive genetic algorithm (IGA) to reduce the gap between the retrieval results and the users’ expectation .They have used color attributes like the mean value, standard deviation, and image bitmap .They have also used texture features like the entropy based on the gray level co-occurrence matrix and the edge histogram. They compared this methods with others approaches and achieved better results.

Meenakshi Madugunki et.al.[2] have published a paper on detailed classification of CBIR Systems. They have used the Global color histogram, Local Color histogram, HSV method for extracting the color feature and matched the result by using Euclidean distance, Canberra distance and city block
distance. They have also used discrete wavelet transform (DWT) for Texture Feature extraction and compared the result obtained by using different features.

Gwenole Quellec et.al.[3] have presented a novel method to adapt a multidimensional wavelet filter bank to any specific problem. They have applied this method for content based image retrieval. The performances of the adapted wavelet filter bank over the nonadapted wavelet filter bank are higher for every database.

Nhu-Van Nguyen et.al.[4] have proposed Clustering and Image Mining Technique for fast Retrieval of Images. The main objective of the image mining is to remove the data loss and extracting the meaningful information to the human expected needs. The clustering-repeat gives good result when the number of examples of feedback is small.

A. Kannan et.al.[5] have proposed Clustering and Image Mining Technique for fast retrieval of Images. The main objective of the image mining is to remove the data loss and extracting the meaningful information to the human expected needs. The images are clustered based on RGB Components, Texture values and Fuzzy C mean algorithm. Entropy is used to compare the images with some threshold constraints.

In this paper [6] Hua Yuan et.al. have presented a new statistical model-based image feature extraction method in the wavelet domain and a novel Kullback divergence-based similarity measure. The Gaussian Mixture Models (GMM) and Generalised GMM are presented to help extract new image features. Compared with conventional norm-based distances (City-block or Euclidean), the Kullback divergence is more appropriate and efficient in the similarity measure and achieved a higher retrieval rate with the same level of computational complexity in a CBIR system.

Zhang Xu-bo et.al.[7] have published a paper on improved K-means clustering and relevance feedback to re-rank the search result in order to remedy the rank inversion problem in content based image retrieval. Experimental results show that the re-ranking algorithm achieves a more rational ranking of retrieval results and it is superior to Reranking via partial Grouping method.

Lijun Zhao et.al.[8] have proposed a multi-round relevance feedback (RF) strategy based on both support vector machine (SVM) and feature similarity to reduce the gap between query and retrieve result. From the experimental results it is seen that SVM and feature similarity based relevance feedback using best feature combination can greatly improve the retrieval and precision rate.

Sharadh Ramaswamy et.al.[9] have published a paper on a fast clustering-based indexing technique. In this method relevant clusters are retrieved till the exact nearest neighbors are found. This enables efficient clustering with low preprocessing storage and computation costs.

Sung-Bae Cho et.al. [10] have published a paper on image retrieval system based on human preference and emotion by using an interactive genetic algorithm (IGA). They used wavelet transform to extract image features and IGA to search the image that the user has in mind. They have conducted several experiments to evaluate the performance of this system. These results showed that their approach allows one to search not only an explicitly expressed image, but also an abstract image such as “cheerful impression image,” “gloomy impression image,” and so on.

From the literature survey it is concluded that a wide variety of CBIR algorithms have been proposed in different papers. The selection feature is one of the important aspects of Image Retrieval System to better capture user’s intention. It will display the images from database which are the more interest to the user.

III. PROPOSED WORK

It is proposed to implement “Content-Based Image Retrieval using different Features. The proposed system will produce the output as images which are relevant to the query Image. The proposed block diagram is given below.

Fig. 1 Proposed system Flow chart
The proposed system has the following structure.

1) Collection of Database: A database containing no of images with any one of the formats .bmp, .jpg, .tiff. is required.

2) Query: The user provides a sample image or sketched figure as the query for the system.

3) Feature Extraction: There are various kinds of low-level visual features to represent an image, such as color, texture, shape, and spatial relationship. Since one type of features can only represent part of the image properties, a lot of work done on the combination of these features. Detailed about all the features are given below:

   a) Color: Color feature is one of the most widely used features in image retrieval. Colors are defined on a selected color space. Variety of color spaces include, RGB, LAB, LUV, HSV (HSL), YCrCb and the hue-min-max-difference (HMMD). Common color features or descriptors in CBIR systems include, color-covariance matrix, color histogram, color moments, and color coherence vector. Storing, filtering and retrieving audiovisual data. The emerging MPEG-7 is a new multimedia standard, which has improved content-based retrieval by providing a rich set of standardized descriptors and description schemas for describing multimedia content. MPEG-7 has included dominant color, color structure, scalable color, and color layout as color features. In my paper I have used csd as color feature. The Color Structure Descriptor (CSD [11]) represents an image by both the color distribution of the image or image region (similar to a color histogram) and the local spatial structure of the color. The extra spatial information makes the descriptor sensitive to certain image features to which an ordinary color histogram is blind. CSD used a $8 \times 8$ structure to scan the total image. This descriptor counts the number of times a particular color is contained within the structuring element while the image or image region is scanned by this structuring element. It has used hmmd color space.

   b) Texture: The texture feature is another type of important and useful visual information for image retrieval. There exist different approaches to extract and represent textures. They can be classified into space-based, frequency-based models, and texture signatures. Some popular techniques i.e wavelet transform, co-occurrence matrix, and Gabor filters are applied to express texture features for image.

   c) Shape: It is the most obvious requirement at the primitive level. It is seen that natural objects are primarily recognized by their shape. Two main types of shape feature are commonly used – global features such as aspect ratio, circularity and moment invariants and local features such as sets of consecutive boundary segments.

4) Similarity Matching: This involves matching these features to yield a result that is visually similar. The commonly used similarity measure method is the Distance method. There are different distances available such as Euclidean distance, City Block Distance, Canberra Distance.

5) Retrieval: The System retrieves and presents a sequence of images ranked in decreasing order of similarity or with the minimum distances is returned to the user.

To evaluate the effectiveness of the proposed system precision and recall rates are to be calculated.

Where,

\[
\text{Precision} = \frac{\text{IR}}{\text{IT}} \tag{1}
\]

IR=No Of Relevance Images Retrieved

\[
\text{Recall} = \frac{\text{IR}}{\text{IRB}} \tag{2}
\]

IR= No of relevance Images Retrieved

IRB=Total no of relevant images in the database.

IV. EXPERIMENTAL EVALUATION

A database consists of different types of images has implemented in the system. The relevant images are retrieved in the screen using different features of the color images. The detailed explanation is given below.

A. Image Retrieval Using Gray Histogram:

   Histogram represents the distribution of intensity of the color in the image. The image retrieval consists of the following stages.

   1) Query image is given from the user.
   2) Color Image is converted to gray image. Histogram of the image is calculated.
   3) Gray Histogram of the database images are calculated.
   4) Euclidean Distance is calculated by the

\[
D = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \tag{1}
\]

5) Sorted the distance in ascending order and Top K images are displayed on the screen.
We have given the query image as scanner and we got some images which are related to query image. But the result which has obtained is not satisfactory. So we go for the next method.

B. Image Retrieval Using Color Histogram:
Histogram represents the distribution of intensity of the color in the image. The image retrieval consists of the following stages. 
1) Query image is given from the user.
2) Histogram of the Color image is calculated.
3) Color Histogram of the database images are calculated.
4) Euclidean Distance is calculated.
5) Sorted the distance in ascending order and Top K images are displayed on the screen.

C) Image Retrieval Using Color Mean:
The mean of pixel colors states the principal color of the image. The mean $\mu$ is given by as follows.
$$\mu = \frac{1}{N} \sum_{i=1}^{N} X_i$$ (2)

Xi indicates ith pixel of image. The steps for Image Retrieval are given below.
1) Query is given from the user.
2) Color Mean of the image is calculated.
3) Color Mean of the database images are calculated.
4) Euclidean Distance is calculated.
5) Sorted the distance in ascending order and Top K images are displayed on the screen.

D: Image Retrieval Using Color & Texture:
For retrieving the images color, edge & texture features are considered. The detailed steps are given below.
1) Query is given from the user.
2) Color & texture features are extracted calculated and these are stored in a matrix. This is called as Feature Vector.
3) Feature vector is also formed for the images present in the database.
4) Euclidean Distance is calculated between the feature vector of query image and database images.
5) Sorted the distance in ascending order and Top K images are displayed on the screen.

When we have used color, edge & texture features we have got better result.
V. CONCLUSION

Contentment Based Image Retrieval has overcome all the limitation of Text Based Image Retrieval by considering the contents or features of image. A query image can be retrieved efficiently from a large database. CBIR technology has been used in several applications such as fingerprint identification, biodiversity information systems, digital libraries, crime prevention, medicine, historical research.

A Database consists of different types of images has implemented on the system. Different Features such as histogram, color mean, Color structure descriptor-texture is taken into consideration for extracting similar images from the database. From the experimental result it is seen that combined features can give better performance than the single feature. So selection of feature is one of the important issues in the image retrieval. The system is said to be efficient if semantic gap is minimum. The result can be improved in future by introducing feedback and user’s choice in the system.

REFERENCES


