Color - Texture based Image Retrieval System

Rahul Mehta
M.Tech (CTA)
SOIT, RGPV
Bhopal (M.P.), India

Nishchol Mishra
Asst. Professor
SOIT, RGPV
Bhopal (M.P.), India

Sanjeev Sharma
Professor
SOIT, RGPV
Bhopal (M.P.), India

ABSTRACT
Content Based Image Retrieval (CBIR) is an interesting and most emerging field in the area of 'Image Search', in which similar images for the given query image searched from the image database. Current systems use color, texture and shape information for image retrieval. In this paper we propose a method in which both color and texture features of the images are used to improve the retrieval results in terms of its accuracy. Color extraction and comparison are performed using Conventional color histograms (CCH) and the Quadratic Distance Metric (QDM)and the texture extraction and comparison are performed using the concept of Pyramid Structure Wavelet Transform Model (PSWTM) and the Euclidean distance. Color and texture based image retrieval computes image features more accurately which are used to retrieve similar images from the database.

Keywords
CBIR, Color based Search, Texture based Searching, Color Histogram, Pyramid Structure Wavelet Transform model, Euclidean Distance, Quadratic Distance Metric

1. INTRODUCTION
Due to the exponential growth of image data there is a compelling need for innovative tools which can easily manage, retrieve and visualize images from the large multimedia databases. In recent year, there are many advanced techniques which have been emerged in the field of Content-Based Image Retrieval (CBIR). The main goal of the CBIR is to find images which are similar to the query image visually without using any textual descriptions for the image. CBIR is the field and collection of technology and algorithms, whose detail study enable the user to query image databases by using image content such as color, texture, objects and their geometries not by using textual attributes such as image name or other keywords [3, 6]. In CBIR various techniques have been developed over the past few years based on various features of images. Among these features, color is one of the effective features which can express visual information, as well as invariant on the complexity. By the use of color features images can be grouped together very easily. Texture analysis is another major component of image processing and work as the backbone for many applications such as remote sensing, quality inspection, medical imaging, etc. [2].

Hence we proposed a method in which both color and texture features are used for efficient image retrieval because it has been observed so many times that when color and texture both features are used for retrieval process then the result obtained through these retrieval processes are more efficient. A Typical Content based Image Retrieval system is drawn below representing the basic principle used by CBIR:

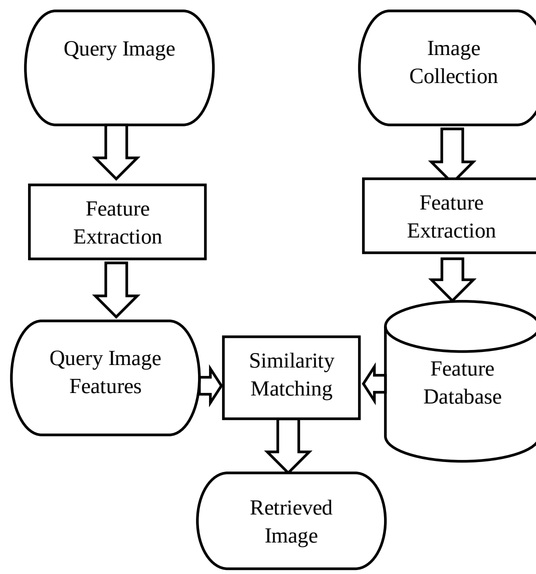

Fig.1: Typical Content Based Image Retrieval System

2. PREVIOUS WORK
Image retrieval is very interesting and vast field. Since 1970, research on the advance image retrieval is started. In 1979, a conference on Database Techniques for Pictorial Applications was held in Florence due to which many researchers attracted towards the field of image database management. After that several researches had been done on features based image retrieval, later a system [7] was proposed which uses the concept of texture based image retrieval system combines with the wavelet decomposition [12] and gradient vector [18]. In that system every image is associated with a coarse feature descriptor and a fine feature descriptor was derived through the use of wavelet coefficients related to the original image. In the first stage coarse feature descriptor is used so that non-promising images can be quickly separated, so that searching for similar images can be done efficiently. Another image retrieval system was introduced [4] which was based on the principle of motif co-occurrence matrix (MCM), which can easily find out the basic difference between pixels and can also convert them into a basic graphic. It can compute the probability of occurrence of pixels in the adjacent area and work as an image feature. Another system was proposed which utilizes the properties like contrast [14], coarseness and

directionality models [17, 13] to achieve texture classification and recognition. After that, a texture-based image retrieval method was proposed which was based on two-stage content-based image retrieval system by using texture similarity[5] which enhanced the image retrieval technique.

To obtain better result, we have proposed a better retrieval technique which integrated color and texture features in order to improve image retrieval. Color histogram [19] is one of the common techniques used in image retrieval systems hence for our proposed system Conventional Color Histogram (CCH) are used to combined with Quadratic distance in order to classify similar images for further enhancing the system texture features are extracted by using the concept of Pyramid Structure Wavelet Transformation combined with Euclidean Distance from those images which was previously classified through color model. The result obtained by using this system is better than other convention systems which only use color, texture features individually. Hence we can say that combination of color and texture feature for finding similar image retrieval makes system more efficient and effective.

3. COLOR& TEXTURERETRIEVAL

In color-based image retrieval system mostly conventional color histogram (CCH) [15] and the fuzzy color histogram (FCH) [8] are used for ensuring the similarity between the two images by calculating similarity distance through distance function. Below we described the four methods which are used by most of the color retrieval methods:

(1) Firstly a proper color space must be selected.
(2) Then Color Space Quantization will be done.
(3) Next step is to extract the color features.
(4) After that calculation of distance function is done to get Similar images [15].

Color attribute are used for object identification and color feature extraction is performed in image retrieval system, it can also help by enabling system to perform multiple measurements on the single pixel of the image, it can also classify complex object without their segmentation. Color histograms are used to represent image color information in various CBIR systems. A color histogram is a type of bar graph, in which each bar represents a particular color of the color space being used. The bars in a color histogram are referred to as bins and they represent the x-axis. The number of bins will be totally dependent on the number of colors in an image. Color histogram y-axis denotes the numbers of pixels of each bin. Some color histogram feature spaces can occupy more than one hundred dimensions [1].For our system we generate the histogram of an image in matlab, which provide basic functions to calculate histogram of an image. A typical Color histogram of an image is drawn below:

Fig. 2: Image

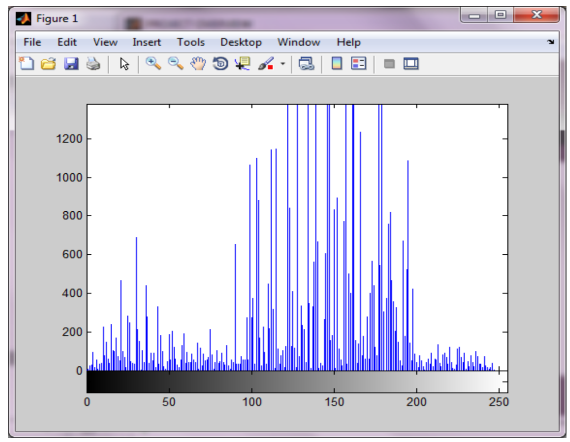

Fig.3: Image Histogram

There are two types of color histograms, Global color histograms (**GCH**s) and Local color histograms (**LCH**s). A GCH represents one whole image with a single color histogram while the LCH divides an image into fixed blocks and takes the color histogram of each of those blocks [10]. Figure 3.1 shows the GCH of an image.

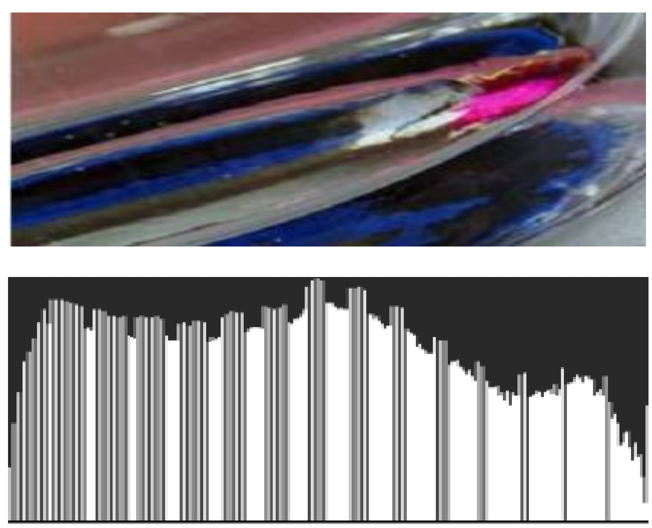

Fig.4: Image and its Global Color Histogram.

Many times it has been seen that color-based image retrieval systems are not able to retrieve the images successfully in the different condition even they have taken from similar surfaces. Surface of the object can be depicted by using texture. Texture defines the properties of an object like smoothness, coarseness, and regularity. From the presence of a single color structural homogeneity cannot be depicted and may require various intensities interaction within a region. In order to distinguish similar color objects like sky, grass etc. texture similarity is the better option. But successful texture analysis is very difficult to achieve. Texture retrieval is performed by segmenting an image whose texture analysis has to be done into number of texture regions and then perform each region analysis separately but proper segmentation is not easier to achieve and due to this image retrieval process may affect. Also texture can't describe fully as it depend upon person to person perception which may differ for the same scenario.

Hence even after so much research satisfactory quantitative definition of texture still not available at this time. The texture depends upon three factors [11] firstly it depends over a region some local "order" of large order's size can be repeated. Second, if elementary parts are arranged nonrandom may affects the order. Third, entities which are not properly uniform can occupy the whole textured region.

3.1 Quadratic Distance Metric

We use Quadratic Distance Metric for image classification. The Quadratic Distance Metric is used to describe the distance between two color histogram [19]:

$$\mathbf{d^2(Q, I) = (H_Q - H_I)^t A (H_Q - H_I)}$$

Quadratic distance has been used in many retrieval systems for color histogram-based image retrieval. The above equation defines three terms. The first term defines the difference between two color histograms or we can say that it indicates number of pixels differences in each bin. The number of bins in a histogram is defined by number of vector column. Vector transpose is denoted by the third term. Similarity matrix is defined by the middle term. Color distance between the two images is represented by **d.** If the images are similar in color features then the distance will be closer to zero. Less color similarity images will have greater distance from zero.

3.2 Similarity Measure

Instead of depending upon exact matching, we calculate visual similarities between a query image and images in a database. Hence the retrieval result is not a single image but a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years. Performance of an image retrieval system is dependent on the type of similarity ***measures*** used. Below is the similarity matrix equation obtained through a complex algorithm.

$$\boldsymbol{A = [a_{q,i}]}$$

$$\boldsymbol{a_{q,i} = 1 - 1/\sqrt{5}\Big[(v_q - v_i) + \big(s_q \cos(h_q) - s_i \cos(h_i)\big)^2 + \big(s_q \sin(h_q) - s_q \sin(h_i)\big)^2\Big]^{\frac{1}{2}}}$$

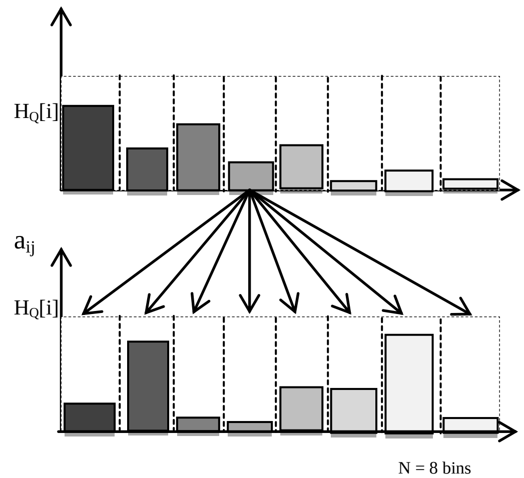

Fig.5: Quadratic Distance Approach

4. PROPOSED COLOR FEATURE EXTRACTION & SIMILARITY MEASUREMENT

In order to extract color feature of image and get consistent result we firstly resized all the images in our database to 256x256 BMP images. We use 500 image databases for evaluation of our proposed method. The problem with 24-bit JPEG is that they are compressed and the compression algorithm affects the way the histograms are derived. We use matlab function **imhist** to get indexed color maps of the images. Indexing quantize the color map by letting the user specify the number of bins. This obviously reduces the processing time in terms of calculating color distances. The equalized histogram image is split into four fixed bins in order to extract more distinct information from it. The frequencies of 256 values of images are split into four bins carrying 64 values each (0~63, 64~127, 128~191, and 192~255). This is done by turning off the gray values of image which do not lie between the four bins. This provides a better illustration of image segments and simplifies the computation of features for the distinct portion of image.

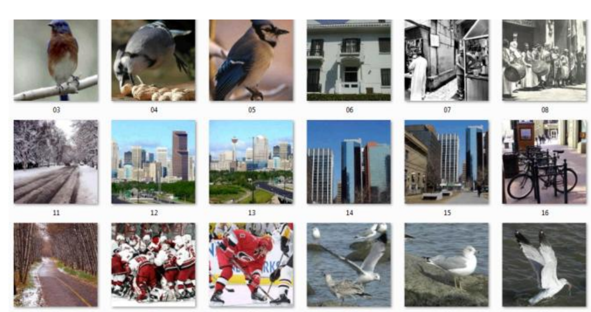

Fig.6: Image Database.

After generating Color Histogram and calculating the value the next step is to calculate Color Histogram of every image in the database. Then we compute the Euclidean distance between the query image and the image in database. Then the result is calculated and according to their distance result is displayed.

4.1 Experimental Results

After obtaining all the necessary terms, similarity matrix, and color histogram differences, for a number of images in our database, we calculated the results by using quadratic distance metric**.**According to their quadratic distances result is sorted which are shown below:

Query Image

Fig.7: Result obtained through Color Retrieval

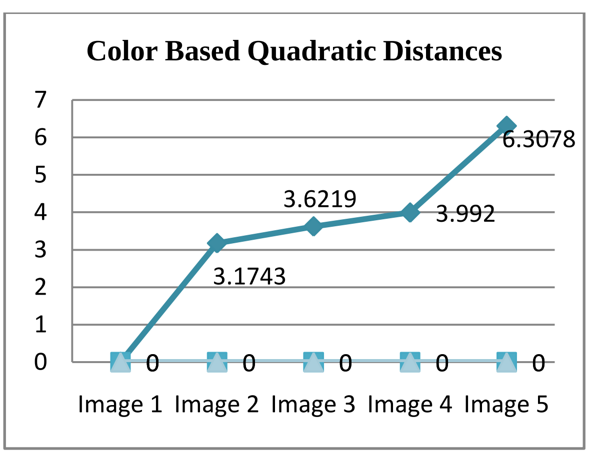

Fig.8: Color distance between query and result

5. PROPOSED COLOR-TEXTURE EXTRACTION & SIMILARITY MEASUREMENT

After calculating the color features of query and other database images we further decompose these images into *four* sub-images so that we can calculate the energy of all decomposed images at the same scale [16]:

$$\mathbf{E} = \frac{\mathbf{1}}{\mathbf{MN}} \sum_{\mathbf{i=1}}^{\mathbf{m}} \sum_{\mathbf{j=1}}^{\mathbf{n}} |\mathbf{X}(\mathbf{i}, \mathbf{j})|$$

Dimensions of the image are represented by M and ***N***, image map column i and j represent intensity of the pixel. After decompose, query image energies of the first dominant ***k*** channels is calculated. The Euclidean distance between the two sets of energies is calculated, using:

$$\mathbf{D_i} = \sum_{\mathbf{k=1}}^{\mathbf{k}} (\mathbf{x_k} - \mathbf{y_{i,k}})^2$$

We used a method called the pyramid-structured wavelet transform for texture classification as it recursively decomposes sub signals in the low frequency channels. Textures with dominant frequency channels can be classified easily. For this reason, it is mostly suitable for signals consisting of components with information concentrated in lower frequency channels [9]. Due to the innate image properties that allows for most information to exist in lower sub-bands, the pyramid-structured wavelet transform model is highly sufficient. Using the pyramid-structured wavelet transform model, the texture image is decomposed into four sub images, in low-low, low-high, high-low and high-high sub-bands. Then energy level of each sub-band is calculated. It is the first level decomposition further decomposition done by using the low-low sub-band till fifth level of decomposition, in our project. This has been done because many times it has been seen that energy of an image is concentrated in the low-low band.

5.1 Proposed Color - Texture Algorithm

Step 1: Calculate Color Histogram of Query Image.

Step 2: Then Calculate Histogram of images in the database whose result stored in the file.

Step 3: Calculate Quadratic distance of each image in the database to query image and store the result.

Step 4: Use the result obtain in step 3 then decompose the query image into *4* sub-images. (This is done to calculate the energy at same scale.). Then calculate the energy using equation 3.

Step 5: After calculating the first dominant k channels energies of query and other images from obtain from color feature extraction calculate the Euclidean distance between the two sets of energies, using equation 4.

5.2 Experimental Results

By using the above color texture algorithm, the query image is searched in the image database. For finding similar images first Color Histogram and Quadratic distance is calculated between the query image and images in the database one by one. Then the result obtained through color retrieval is stored in a file which is used for further classification performed through wavelet transformation and Euclidean distance is calculated. This process is repeated until all the images in the color retrieval file have been compared with the query image. After the completion of Euclidean distance algorithm, we finally have an array of images according to their Euclidean distances, which is sorted and displayed.

Fig.9: Query Image

Fig. 10: Result obtained through Color- Texture Retrieval

The above results are sorted according to the Euclidean distances which are represented through Table - 2.

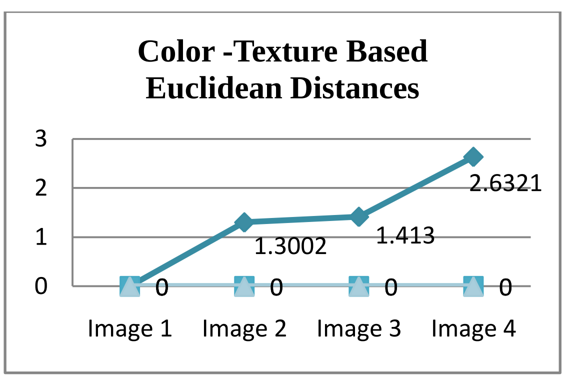

Fig.11: Euclidean distances of images retrieved through color and texture

6. CONCLUSION

In this paper, we propose a Color- Texture Based Image Retrieval System (CTBIRS). The application performs a simple color-based search in an image database for an input query image, using Conventional Color Histograms (CCH). It then finds the similar images from image database by Quadratic Distance Metric (QDM). In order to further enhance the search, the application performs a texture-based search on color results, by using Pyramid Structure wavelet transform model (PSWTM) and energy level calculation. It then compares the texture features using the Euclidean Distance Equation. The experimental results show that the proposed similarity measure method is better than the traditional ones, which use color-texture feature individually for image retrieval.

7. REFRENCES

[1] Xiang-Yang Wang, Jun-Feng Wu1 and Hong-Ying Yang "Robust image retrieval based on color histogram of local feature regions" Springer Netherlands, 2009 ISSN 1573-7721.

[2] D.Goswami, S.K.Bhatia, "RISE: A Robust Image Search Engine", Electro/information Technology, 2006 IEEE International Conference on, 7-10 May 2006, pp. 354-359.

[3] A.Ahmadian, A.Mostafa, M.D.Abolhassani, Y.Salim- pour, "A texture classification method for diffused liver diseases using Gabor wavelets", Engineering in Medicine and Biology Society, 2005.

[4] N. Jhanwar, S. Chaudhurib, G. Seetharamanc, B. Zavidovique, Content based image retrieval using motif co-occurrence matrix, Image and Vision Computing 22 (2004) 1211–1220.

[5] P. W. Huang and S. K. Dai, "Design of a two-stage content-based image retrieval system using texture similarity," Information Processing and Management, vol. 40, no.1,pp.81–96,2004.

[6] S.K.Saha, A.K.Das, B.Chanda, "CBIR using Perception based Texture and Color Measures", Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 23-26 Aug.2004,pp.985-988Vol.2.

[7] P.W. Huang, S.K. Dai, Image retrieval by texture similarity, Pattern Recognition36 (3) (2003)665–679.

[8] J. Han and K.-K. Ma, "Fuzzy color histogram and its use in color image retrieval," IEEE Transactions on Image Processing, vol. 11, no. 8, pp. 944–952, 2002.

[9] SharminSiddique, "A Wavelet Based Technique for Analysis and Classification of Texture Images," Carleton University, Ottawa, Canada, Project Report 70.593, April 2002.

[10] Shengjiu Wang, "A Robust CBIR Approach Using Local Color Histograms," Department of Computer Science, University of Alberta, Edmonton, Alberta, Canada, October 2001, Found at: http://citeseer.nj.nec.com/wang01robust.html

[11] W. K. Pratt, Digital Image Processing, Wiley-Interscience, New York, NY, USA, 3rd edition, 2001.

[12] I. Daubechies, Orthonormal bases of compactly supported wavelets,Communications on Pure and Applied Mathematics 41 (1998) 909–996.

[13] W. Niblack, X. Zhu, J. L. Hafner, et al., "Updates to the QBIC system," in Storage and Retrieval for Image and Video Databases VI, vol. 3312 of Proceedings of SPIE, pp. 150–161, San Jose, Calif, USA, January 1998.

[14] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: wold features for image modeling and retrieval," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 7, pp. 722–733, 1996.

[15] J. R. Smith and S. F. Chang, "Automated image retrieval using color and texture," Tech. Rep. CU/CTR 408-95-14, Columbia University, New York, NY, USA, July 1995.

[16] W. Niblack et al., "Querying images by content, using color, texture, and shape", SPIE Conference on Storage and Retrieval for Image and Video Database, Volume 1908, pp.173-187, April 1993.

[17] W. Niblack, R. Barber, W. Equitz, et al., "The QBIC project: querying images by content, using color, texture, and shape," in Storage and Retrieval for Image and Video Databases II, vol. 1908 of Proceedings of SPIE, pp. 173–187, 1993.

[18] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, vol. I, Addison-Wesley,Reading, MA, 1992.

[19] M.J. Swain, D.H. Ballard, Color Indexing, International Journal of Computer Vision 7 (1991) 11–32.