An Introduction of Content Based Image Retrieval Process

Sunil Chavda[1] Lokesh Gagnani[2]
[1,2]Department of Information Technology
[1,2]Kalol Institute of Technology and Research Center Kalol, India

Abstract—Image retrieval plays an important role in many areas like fashion, Engineering, Fashion, Medical, advertisement etc. As the process become increasingly powerful and memories become increasingly cheaper, the deployment of large image database for a variety of applications. It has now become realisable. Content Based image retrieval is method of extracting the similar image or matched image from the image database. It has become popular for getting image from the large image database. By using low level features like colour, texture, shape the retrieval become more efficient. There are many research algorithm developed for CBIR. In this paper we have used two retrieval process query by colour and query by texture. Colour features contain colour histogram in RGB colour space and texture features involve the invariant histogram and mean and standard deviation to retrieve the image. It is observed from the experiment that query by texture is more effective than colour for retrieving images. Precision and Recall provides the performance measurement.

Key words: CBIR, Colour features, Texture Features, Histogram, Colour Space

I. INTRODUCTION

As digital image quickly increase number, retrieval of images from large database has become more important. Image retrieval procedure can be divided into two approaches: Text based and Content based. In text based, it retrieves image based on one or more keywords by user. To solve this CBIR has come in way to retrieve the image based on the content. CBIR avoids many problems with traditional way retrieval images. Content Based Image Retrieval (CBIR) is the process to retrieve images from low level features like colour, texture, shape and spatial information. The most challenge of CBIR is to determine the exact/approximate image of database to query image. It extract content based features namely colour, texture, shape from image. By comparing between query and database image determined by the difference of features of the image by distance function. A feature vector is extracted from the query image and is matched against feature vectors in indexed.

In Fig.1 how CBIR system is working .CBIR perform in two ways indexing and searching. CBIR takes image database, Query image, image database features, Query image features, Similarity matching process, and resultant or matched image from the database. Feature image of both query image and database image is firstly calculated. Then similarity measurement is done between two. The database image matches with query image is ranked as first and so on.

The features of image are represented in the feature vector form. For similarity measures various methods are used to compare the two features. Main point the retrieval result is not a single image, rather it is number of ranked images.

II. FEATURES OF IMAGE

It is defined as the characteristics of the image. There are three levels.

- Low level: It includes texture, colour, shape, Spatial information.
- Middle level: There is arrangement of particular types of objects.
- High level: There is specific content of the image. Impression, emotions are high level features.

After Extract the feature vector of an image and calculate feature vector. So the low level features are extracted and calculated.

1) Colour: Colour is the property depends on the reflection of the light on eye and processing of information in brain. Colour defines three dimensional colour spaces. They could be RGB (Red, Green, Blue), HSV (Hue, Saturation, Value) or HSB (Hue, Saturation and Brightness).The last two are depends on human perception of hue, Saturation and brightness. Colour Histogram is most traditional process for describing low level property of an image. It is type of bar graph. We can get colour histogram in RGB or HSV colour space. A Bar in colour histogram is referred as bins and represent x-axis. The number of pixel in each bins denote y-axis. Colour moments and colour vector are used to calculate colour feature vector.
2) Texture: Texture is the property of surface which describe visual pattern. It contains information about structural arrangement of the surface such as cloud, leaf, bricks etc. In short it is physical composition of surface.

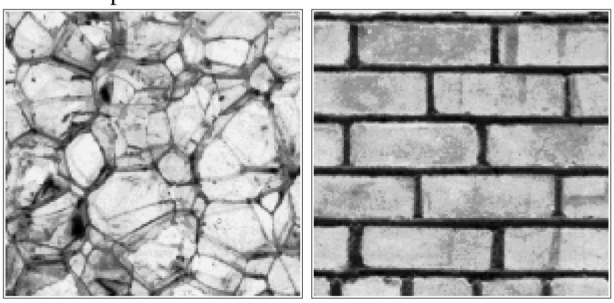

Fig. 1: Texture surface of Brick and Rocks

In Fig.1 describes the texture of brick and rock. The contrast, directionality, roughness are the texture properties.

3) Shape: Shape is also an important low level feature image retrieval process. Object can form by a set of shape.it extract the shape from images by segmentation. It has various properties like scaling, rotation, transition. In shape based image retrieval process user need to choose the reference image or sketch a desired shape. There are main two type : Boundary based and region based. The morphological operation, canny edge detection is more useful in shape based retrieval process.

4) Image Database: The image collection is obtained from world ide. The database of image is being used. Images are divided into different categories and each category has similar images.

A. Colour Feature Extraction:

In this proposed system, we retrieve image using low level features like colour and texture .While retrieving image using colour, the RGB image is converted into HSV space.

1) Colour Quantization and Colour Space:

The colour of an image is represented by colour spaces like RGB, XYZ, YIQ, YUV and HSV. The HSV colour space gives best colour histogram feature among different colour spaces.

Colour Quantization is the process to optimize the use of various colours in an image without effecting visual properties. To reduce computation time, the colour quantization can be used, without reduction in image quality and reducing the storage space and process speed.

The colour histogram used to represent colour distribution in an image. Colour histogram counts the number of each unique colour on a sample image. Image is composed of pixels and each pixel has a colour. The colour histogram of image can compute by visiting every pixel once.

Histogram search characterizes image y its colour distribution. Many histogram distances have been used to define similarity of two colour histogram representation. It is implemented using one-dimensional array.

A colour histogram for given image is represented by vector:

S = { S[0] , S[1] , S[2], S[3], ………S[i]………S[n] }

Where is the colour bin in the colour histogram, S[i] is the number of pixels of colour I in the image, and n is the total no. of bins used in colour histogram. Each pixel in an image will assign to a bin of a colour histogram. The value of each bin gives the number of pixels that has same correspondence colour. For compare image in different sizes, histogram should be normalized. The normalized colour histogram is:

S' = {S'[0], S'[1], S'[2]… ……..S'[i]… S'[n] }

Where S' [i] = S[i] / p, P is the total no. of pixels of an image.

The image in the database whose histogram compared with the query image histogram and the minimum distance will be displayed at the first ranked.

B. Texture Feature Extraction:

Like colour, Texture is a powerful low-level feature for image search. There is no unique definition for it. It is the spatial arrangement of grey levels of the pixels in image. The common texture descriptors are Tamura features, Gabor-filter, co-occurrence matrices, Wavelet transform. We will use wavelet transformation. It decomposes an image into orthogonal components, because of its better localization property.

The texture features provide the information about the properties of intensity level distribution like uniformity, smoothness, flatness, contrast and brightness. The proposed texture features are mean, skewness, standard deviation, energy, entropy and kurtosis. Which are calculated using different functions like probability distribution of intensity level in the histogram. The bins of histogram of DC, AC1, AC2 and AC3 are coefficients.

$$P(b) = \frac{H(b)}{M}$$

Where, M is the total number of blocks in the image I.

1) Mean:

It is the average of intensity values of all bins of the four quantized histogram. It describes the brightness of image.

$$\text{Mean} = \sum_{b=1}^{t} b\,P(b)$$

2) Standard Deviation:

It is defined as the distribution of intensity values about the mean in all blocks of histogram. It shows high or low contrast of histograms in images with low or high values.

3) Standard deviation

$$= \sqrt{\sum_{b=1}^{L} (b - mean)\,(b - mean) P(b)}$$

4) Energy:

The energy is used to calculate uniformity of the intensity level distribution in all bins of histogram. The energy with high values shows the intensity value is small number of bins of histograms. It is calculated as:

$$ENERGY = \sum_{b=1}^{L} [P(b)]^2$$

It measures the randomness of distribution of intensity levels in bins.

After calculation of these texture features, these values are combined to get feature vector f such as:

F = {mean, Standard deviation, Energy}

The feature vector is calculated in all histogram like F (HDC) is calculated in DC histogram, F(HAC1) is calculated in AC histogram. There are 3 AC histogram calculations. Four features of vectors are combined to get a single feature vectors.

FV = [F(HDC), F(HAC1), F(HAC2),F(HAC3)]

There is feature vector of all images constructed. And it is stored to create a feature database. The feature vector of query image is also constructed and compared with the feature vector of database for similarity and retrieve relevant images.

C. Combining the Feature:

The retrieval result using only one feature may be insufficient. Sometimes the output image is not similar to the query image. So to produce efficient results, the combination of texture and colour is done. The colour used the colour histogram and the normalized colour histogram which is linear. Texture used the different features and made the feature vector. If we combine the colour and texture features to retrieve the image the efficient result should be generated. The new feature vector is the combination of colour and texture feature.

D. Similarity Measurement:

Similarity image is the process I which the query image is compared with database image. The similarity measurement calculates the difference between the query feature vector and database feature vector using distance matrices. The small difference between the query and database feature vectors shows the larger similarity. Here we will use the Euclidean distance function for similarity measurement process.

The Euclidean distance matric is most commonly used for similar measurement. It is used to measure the distance between two vectors of images by calculating the sum of squared absolute differences.

$$Euclidean\ Distance\ (\Delta d) = \sqrt{\sum_{i=1}^{n}(|Qi - Di|)^2}$$

Where, Di is the Database image feature vector
Qi is the Query image feature vector

E. Proposed Method:

In proposed method, there is combination of colour and texture retrieval process. There is combination of colour and texture features process. First we calculate the colour histogram then calculate the mean, probability distribution, standard deviation, Energy and make the feature vector of them. Then combine colour and texture features for make the combine feature vector and all the combine values are store in the feature database.

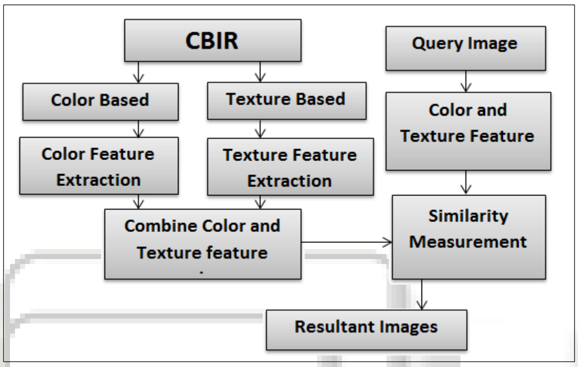

Fig. 2: Block diagram of Proposed System

For the Query image calculate the colour and texture features of the image and compare its value with the feature image database by the distance function and the minimum distance gives the best image retrieval image and the resultant image are arranged in the ascending order.

F. Performance Evaluation and Results:

For evaluate performance of CBIR, the value of Precision and Recall is considered.

1) Precision:

It is defined as the ratio of number of relevant images retrieved to total number of images.

$$\text{Precision} = \frac{\text{Number of relevant Images retrieved}}{\text{Total number of Images Retrieved}}$$

2) Recall:

It is defined as umber of relevant images retrieved to relevant images available in database. Fig3. Shows image retrieve from the query image.

$$\text{Recall} = \frac{\text{Number of relevant Images retrieved}}{\text{Total number of releant Images}}$$

In this paper the standard database is subset of 600 images. They have been manually selected from 10 classes of 100 images. It is used for clustering and classification process.

Total Images	Proposed (%)	Normal(%)
200	74	76
300	74	75
400	70	75
500	78	79
600	79	80

Table 1: Accuracy of Both the System

From the above table the proposed system provides the accuracy 78% and normal system provides the 74% accuracy. So the proposed system is accurate.

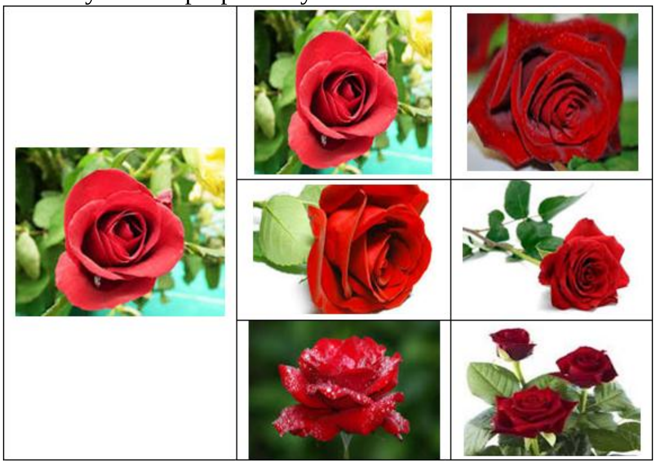

Fig. 3: Resultant Images from the Query image.

Fig 3 describes the results from the proposed method. The query image is compared with the database images which retrieve the similar images from the database image. Here the flower database is taken from there one flower is taken as the query image and it goes for search the similar image from the database image.

III. CONCLUSION

When the feature dataset is divided into similar image classes by clustering and classification the retrieval of similar images time will be decreased because there are similar images placed together. Proposed method is better than normal method because it provide better result and less time than normal method. The proposed system used in all types of image because it is not based on dataset used. The feature work is to retrieve the exact shape from the retrieval image of the proposed method.

ACKNOWLEDGEMENT

I would like to thank Professor Lokesh Gagnani working as Assistant Professor in IT Department in KITRC for guiding me in this field.

REFERENCES

[1] Sunkari Mahdi,"Content based Image retrieval process: A Quantitative comparison between the Query by color and Query by Texture" journal of Industrial and Intelligent Information Vol. 2 June 2014.
[2] Anita Sharma, Gurpreet Singh "Content Based Image Retrieval using low level features" Department of Electronic and Communication Lovely Professional University Volume 3. March-April 2013.
[3] A.Ramesh Kumar, D.Saravanan "Content Based Image retrieval using color histogram"-IJCSIT.
[4] Shankar patel " Content Based Image retrieval using color, texture and shape" International Journal of Computer Science & Engineering Technology (IJCSET).
[5] Ms. K. Arthi, Mr. J. Vijayaraghavan "Content Based Image retrieval algorithm using color models" International Journal of Advanced Research in

Computer and Communication Engineering Vol. 2, Issue 3, March 2013
[6] Nidhi Singha,Kanchan Singhb , Ashok K. Sinha "A Novel Approach for Content Based Image retrieval process" ELSEVIER C3IT-2012.
[7] Ms. Urvashi Chavan, Prof. N. M. Shahan" Content Based Image Retrieval using Clustering" IJAIEM- Volume 3, Issue 10, October 2014