**International Journal of Technical Research and Applications e**-ISSN: 2320-8163,
www.ijtra.com Volume 2, Issue 3 (May-June 2014), PP. 56-58

# SEARCH ENGINE FOR IMAGE RETRIEVAL

**Tanvi Ranadive, Shruti Shirodkar, Varsha Jain and Nikitha Chettiar**
Computer Department,
K. J. Somaiya College of Engineering Vidyavihar,
Mumbai, India

***Abstract*— The content based image retrieval (CBIR) technique is one of the most popular and evolving research areas of the digital image processing. The goal of CBIR is to extract visual content like colour, texture or shape, of an image automatically. This paper proposes an image retrieval method that uses colour and texture for feature extraction. This system uses the query by example model. The system allows user to choose the feature on the basis of which retrieval will take place. For the retrieval based on colour feature, RGB and HSV models are taken into consideration. Whereas for texture the GLCM is used for extracting the textural features which then goes into Vector Quantization phase to speed up the retrieval process.**

***Key words*— Content based image retrieval, Texture Feature, Colour Histogram, Grey Level Co-Occurrence matrix, Vector Quantization, Euclidean Distance, k-means.**

## I.INTRODUCTION

With the growing technology, the use of multimedia data like images has increased rapidly, due to popularity of smart devices and personal computers. Large amounts of data stored in database leads to tedious retrieval of the required image from the data base when required. Text based retrieval is non-standardised because different users use different keywords for searching. Text descriptions are inefficient because they cannot depict the complicated features of an image very well. Content based image retrieval solves this problem of retrieval to a great extent.

The goal of CBIR systems is to support image retrieval based on content i.e. information of the image e.g., shape, colour, texture leading to faster and efficient retrieval of images. Content-based means that the search makes use of the information of the image itself, rather than relying on human-input metadata as captions or keywords. CBIR technique supports search of database comprising of a wide range of images.

The input to the search engine will be an image itself. Search engine later processes the database by using an algorithm which the user can choose. The relevant images are displayed to the user once the algorithmic steps are executed.

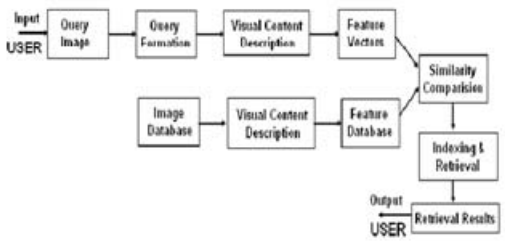


**Fig. 1. Illustration of CBIR system**

## II.FEATURE EXTRACTION METHODS

Feature extraction is an important step for extracting the information related to the image which can later be used to retrieve the image based on a content search. Image content is mainly embodied in colour, texture, shape. A good extraction technique helps in acquiring more features, thereby providing more information on the image content and hence a more successful search.

### *A. Colour Feature Extraction*

Colour is the most commonly used feature for retrieval of an image as it provides the most intuitive information. For extraction of colour feature from an image, proper colour space should be determined. The proposed system uses RGB and HSV colour model.

The RGB model is a colour model in which the three colours red, green and blue are intermixed to produce a wide range of colours. The HSV model is a cylindrical coordinate representation the of points in an RGB color model. The HSV model is more efficient to work with due it non Cartesian representation.

The algorithm for colour based image retrieval has the following steps:

1. Read images from the database and extract RGB information from the images.
2. Create histograms for the RGB/HSV components of each image. Hence each image will have three histograms associated with it.
3. The RGB/HSV information is extracted for the query image also and correspondingly histogram is generated for the image.
4. Euclidean distance is computed by comparing the query image histograms to that of the histogram of each image in the database.
5. Images which have less Euclidean distance are retrieved from the database and displayed as result.

Recall rate and precision rate are calculated to evaluate the performance of the proposed system.

$$\text{Precision rate}=\frac{\text{Number Of relevant images retrieved}}{\text{Total no of images retrieved.}}$$

$$\text{Recall rate}=\frac{\text{Number Of relevant images retrieved /}}{\text{total no of images in database.}}$$

### *B. Texture Feature Extraction*

Texture represents a particular pattern of pixels over a region. The system uses GLCM for extracting the texture features of the image.

Grey Level Co-occurrence matrix comes is a method which comes under the statistical approach. GLCM is a matrix of how often different combinations of grey levels occurring in an

image. GLCM texture considers the relation between two pixels at a time, called the reference and the neighbor pixel. After computation of Gray level co-occurrence matrix, a number of statistical texture measures based on GLCM are derived.

The algorithm for calculating texture properties is as follows:

1. Read the input image.
2. Extract a 3×3 window image from the input image and compute the co-occurrencetexture measure.
3. Estimate the texture parameters for theobtained texture image.
4. Repeat the step3 and step4 by moving the window till the end of the image.
5. Display various texture parameters by normalizing them.

## III.VECTOR QUANTIZATION

Vector Quantization is an efficient method for image compression. For the purpose of compression, Vector Quantization involves dividing an image into several vectors and each vector can be mapped with a codeword from codebook to encode or decode an image. Image compression helps in avoiding the redundancies in an image thereby improving the efficiency in storing and transforming the data. KFCG is used for vector quantization.

To summarize there are 3 steps involved in Vector Quantization :

1. Generating Codebook : For the codebook generation, size of the vector must be decided. And divide the image into vectors. After this apply codebook generating algorithm.
2. Encoding Image : Encoding means find the best match codeword for each vector and replace the codeword with the index of the matched codeword.
3. Decoding Image: Decoding means using the same codebook, for each index, replace it with the respective codeword. And assemble all code words.

### *A. Kekre's Fast Codebook Generation (KFCG) algorithm:*

1. Windows of size 2x2 pixels is obtained by dividing the image.(each pixel consisting of red, green and blue components).
2. These are arranged in a row to acquire 12 values per vector. Collection of these vectors is a training set (initial cluster).
3. Calculate the centroid (codevector) of the cluster.
4. Compare the first element of the training vector with the first element of the codevector and divide the above cluster into two.
5. Calculate the centroids of both the clusters obtained in step 4.
6. Split both the clusters by comparing second element of training vectors with the second element of the codevectors.
7. Repeat the process till we obtain codebook of size 16.
8. Codebook is then converted to 1-Dimension of size 16x12=192 and DCT is applied on this to get the feature vector of size 192.
9. The result is stored as the feature vector for the image. Thus the feature vector database is generated.

### *B. Using Colour With Texture*

K-means algorithm: The k-means algorithm uses both colour and texture features to find images in the database close to the database. The steps involved are as follows:

1. Read all images from the database and extract the colour features.
2. Generate three clusters based on RGB the colour properties.
3. Extract the RGB colour properties for the query image and classify it into one of the three clusters.
4. Extract the texture features of the images in the cluster into which the query image has been classified.
5. Use k-means algorithm to further classify the cluster into two depending on the texture properties.
6. Calculate the centroid for each cluster.
7. Extract the texture features from query image and compare with the centroids to determine the cluster to which it belongs.
8. Using Euclidean distance retrieve the closely related images.

## IV.CONCLUSION

The proposed system is initiated by uploading an image which serves as the query and the user selects the algorithm on the basis of which the search has to be done. The GUI for the system is as shown below:

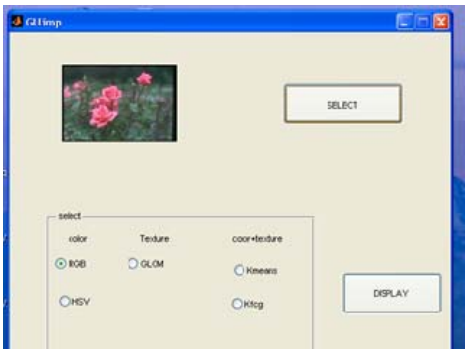


**Fig. 2. The GUI of the system**

On the completion of the input details the system proceeds further by using the algorithm as selected by the user. On completion of the procedure of search, the results are displayed to the user depending on the Euclidean distance of each image and the closely related images are displayed.

The output is as shown below:

**Fig. 3. Output for the above query**

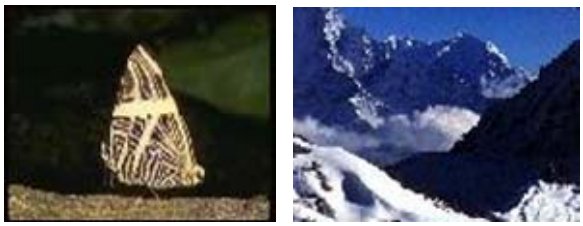

**Fig. 4. Query image 2.jpg and 100.jpg**

**Fig. 5. Query image 705.jpg and 8078.jpg**

**Fig. 6. Query image 7822.jpg**

The graphs of recall vs. precision using three algorithms are shown below:

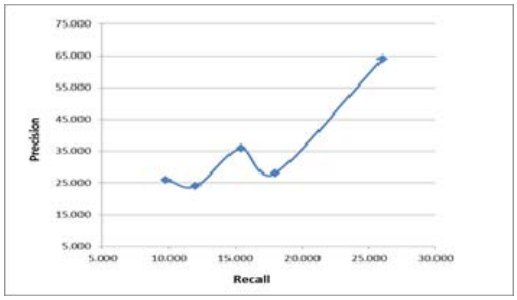


**Fig. 7. Graph of recall vs. precision using GLCM**

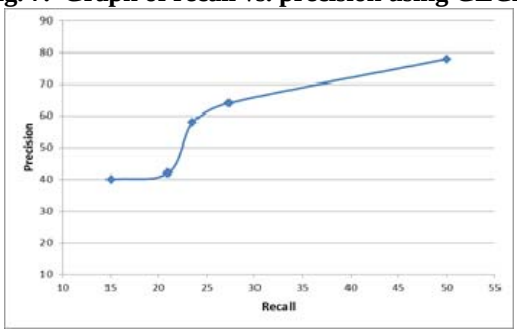


**Fig. 8. Graph of recall vs. precision using HSV**

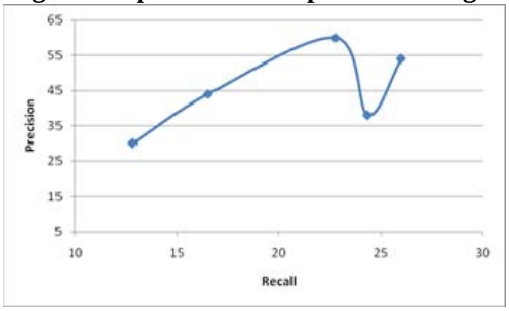


**Fig. 9. Graph of recall vs. precision using k-means**

The first graph i.e. Fig 7 is the recall versus precision graph for GLCM. The graph is a plot of recall and precision values for algorithm implementing GLCM for 5 different query images. It is the highest for the query image 705.jpg. The second graph i.e. Fig 8 is a plot of recall and precision values for the algorithm implementing HSV color model. It is the highest for the image 7822.jpg. The third graph i.e. Fig 9 is a plot of recall and precision values for k-means algorithm. It is highest for the query image 705.jpg.

The table for recall and precision of two images using each of the given three algorithms is shown below:

| | HSV | | GLCM | | k-means | |
|---|---|---|---|---|---|---|
| Recall | 50 | 23.58 | 15.38 | 26.01 | 26 | 22 |
| Precision | 78 | 58 | 36 | 64 | 54 | 60 |

**TABLE I**

V.ACKNOWLEDGMENT

We would sincerely like to thank Prof. Archana Gupta for her constant support and encouragement. Her guidance in the research work was very valuable and helpful.

REFERENCES

[1] H.B.Kekre, Tanuja K. Sarode, Sudeep D., “Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre’s Fast Codebook Generation”, ICGST-GVIP Journal, Volume 9, Issue 5, September 2009, ISSN: 1687-398X.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Dr. H. B. Kekre, Dr. Sudeep D. Thepade, Dr. Tanuja Sarode, Varun K. Banura,”Performance Comparison of Image Retrieval using KFCGwith Assorted Pixel Window Sizes in RGB and LUV Color Spaces”, IJECT Vol. 3, Issue 1, Jan. - March 2012.

[3] H.B.Kekre, Sudeep D. Thepade, Tanuja K. Sarode, Shrikant P. Sanas,”IMAGE RETRIEVAL USING TEXTURE FEATURES’ EXTRACTEDUSING LBG, KPE, KFCG, KMCG, KEVR WITH ASSORTEDCOLOR SPACES”, ISSN: 2231-1963.

[4] Ramamurthy, B. and K.R. Chandran, "Content Based Medical Image Retrieval with Texture Content Using Gray Level Co-occurrence Matrix and K-Means Clustering Algorithms", Journal of Computer Science 8 (7): 1070-1076, 2012,ISSN 1549-3636© 2012 Science Publications.

[5] Rishav Chakravarti, Xiannong Meng, "A Study of Color Histogram Based Image Retrieval", 2009 Sixth International Conference on Information Technology: New Generations.

[6] Carlos W.D. de Almeida, Renata M.C.R. de Souza, Ana L´ucia B. Candeias, “ Texture Retrieval using Co-occurrence Matrix andSymbolic Interval Data under Scale and Rotation Invariance” ,2012 IEEE International Conference on Systems, Man, and CyberneticsOctober 14-17, 2012, COEX, Seoul, Korea.