IMAGE RETRIEVAL SYSTEM BASED ON COLOR (QUANTIZED HSV) AND TEXTURE FEATURE (GLCM)

Nishad Ahmed
Student, Department of Information Technology
Gauhati University Institute of Science & Technology
Guwahati-781014, India
nikuna786@gmail.com

Parismita Sarma
Assistant Professor, Department of Information Technology
Gauhati University Institute of Science & Technology
Guwahati-781014, India
parismita.sarma@gmail.com

Abstract— ***An Image Retrieval (IR) is a system which allows user to browse, search and retrieve digital images based on visual Features such as color, texture and shape. Image retrieval based on single feature cannot provide a good result for accuracy and efficiency. High level feature like object, Semantic etc describes the concept of human brain effort which will reduce the query efficiency and low level features such as color, texture and shape will reduce the query accuracy. So, it is good to use multi feature in image retrieval system. The most important visual features are Color and texture which are used in this review paper. A review on Retrieval of Digital images based on color and texture feature have been done in this paper and also a new system has been proposed for efficient Image Retrieval System using multi-feature. In this paper, three methods for image retrieval system based on color, texture and combination of both color and texture feature have been proposed. Here, color feature has been extracted by using quantized HSV Color space. In HSV (Hue, Saturation, Value) color space, first quantifying the color space in non-equal intervals, then constructing one dimension feature vector and after that representing the color feature by finding the summation of one dimension feature vector. Similarly, the work of texture feature extraction has been done by using gray-level co-occurrence matrix (GLCM) and the work of both Color and texture feature have been extracted by using the quantification of HSV color space and Gray-level co-occurrence matrix (GLCM).The technique shows the advantage of using multi-feature than the single feature in image retrieval system. In order to express the similarity between different images or feature vectors of images, Euclidian Distance is used.***

Keywords- Image Retrieval (IR), Content based image retrieval (CBIR), Color, Texture, HSV Color space, Gray Level Co-Occurrence Matrix (GLCM), Euclidian distance

I. INTRODUCTION

Images are widely used now a days. It has the advantage of visual representation and it is usually adopted to express other mediums. With the rapid development of computers and networks, the storage and transmission of a large number of images become possible. 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[1] .

A. Image Retrieval System

An Image Retrieval (IR) system is a system which allows user to browse, search and retrieve digital images based on visual Features such as color, texture and shape. They are mainly two types –Text based and Content based Image Retrieval. Content based IR are divided into three parts - color, texture and shape. Text based IR are keyword and category etc [6].These classification is shown in figure 1.

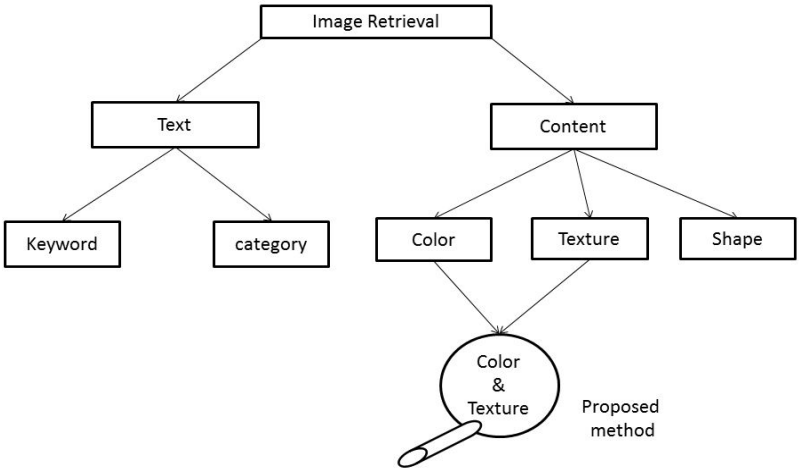

Figure 1: Two categories of image retrieval

The idea of text-based approach was originated at 1970s.In this approach, images are stored in the database associating it with a keywords, number, texts. Then we search the images based on associated metadata such as keywords, texts and other elements that defined the image [6]. This method is not only time consuming but it is dependent to the people who categorized it.

To describe an image such as in Figure 2, perception of each of the person who looks at the image will differs. One person will see the image as sunset scene and one person will interpret this image as boat scene, river or bridge. In this situation, "A picture is worth a thousand words" is being applied.

Figure 2: Example of scenery image (Sunset on Brahmaputra River)

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[7]. Block-diagram of the content-based image retrieval system is depicted in Fig.3. In general, In general, CBIR system has two steps- one is feature extraction, and the other is image matching/similar (also known as feature matching). Feature extraction means, extracting features, but semantic valuable information from images. One of the key issues with the CBIR is the need to extract useful information from the raw data to reflect the image content. The extraction task transforms the rich content of image into various features. Basically, color and texture based image retrieval techniques are the measures of similarity between the color and texture features in the images. The similarity is measured by computing the features in both query and target images[12].

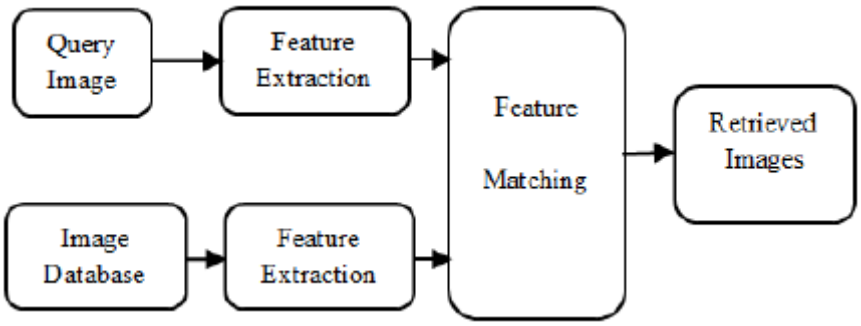

Fig 3: Block Diagram of CBIR System

B. *Different Features of Image Retrieval*

a) Color Feature

In image retrieval color feature is one of the most widely used feature. Colors are defined on different color space[5]. There are many color spaces like RGB, LAB, LUV, HSV. Color-covariance matrix, color histogram, color moments and color coherence vector are common color features or color descriptors in Content Based Image Retrieval (CBIR) systems[8].

b) Texture Feature

The texture feature is another type of important and useful visual information for image retrieval. Color features characterize the visual content of a pixel, whereas the texture features are used to characterize a region of interest[4]. The texture contains important information about structural arrangement and visual pattern of an object and it describes the relationship of the surface to the surrounding environment. There exist different approaches to extract and represent textures. Some popular techniques i.e. wavelet transform, co-occurrence matrix, and Gabor filters are applied to extract the texture features for image [12].

C. *Aim of This Paper*

The objective of this paper is mainly to develop a robust Image Retrieval System based on Color and Texture features. First the work of color feature has been extracted by using Quantized HSV color space. Similarly, the work of texture feature has been obtained by using gray-level co-occurrence matrix (GLCM). The work of both Color and texture feature have been obtained by using the quantification of HSV color space and gray-level co-occurrence matrix (GLCM). In order to express the similarity between different images based on color and texture feature, a metric distance (i.e Euclidian Distance) is employed.

D. *Summary*

In section 2, outlines Literature Survey where provides a previous study on various related work regarding Retrieval of Digital Images Using Color and Texture Feature with Quantized HSV and GLCM Algorithm. In Section 3 deals with Problem Formulation. Propose methodology and conclusion are given in section 4 and section 5.

II. LITERATURE SURVEY

This Section provides a previous study on various related work regarding Retrieval of Digital Images Using Color and Texture Feature with Quantized HSV and GLCM Algorithm .We have undergone detailed study of papers published by various authors related to my area of concern are mentioned below.

[1] The author proposed a new and efficient technique to retrieve images based on sum of the values of Local Histogram and GLCM (Gray Level Co-occurrence Matrix) texture of image sub-blocks to enhance the retrieval performance. The image is divided into sub blocks of equal size. Then the color and texture features of each sub-block are computed.

[2]The author proposed a method of CBIR(Content based image Retrieval) using color sketches ,which is one of the most popular, rising research areas of the DIP (Digital Image Processing).The Goal of CBIR is to extract visual content of an image automatically ,like color , texture and shape. The proposed method is based on free hand color sketch of image ,which use color and texture features. The combination of the color and texture features of an image provides a robust feature set for image retrieval. Euclidean distance of color and texture is used in retrieving the similar images.

[3] The author proposed the texture retrieval method using the GLCM algorithm which is compared with Gabor filters. For color feature retrieval, the two color models RGB and HSV are used where images are

converted from RGB to HSV color model and new technique called Color GLCM method is proposed for color retrieval which shows a comparatively efficient result.

[4] The author proposed a efficient image retrieval system which is used in texture feature by using grey – level co-occurrence matrix (GLCM) and Color Co-occurrence matrix (CCM). The GLCM and CCM separately combined with a color feature with the use of quantization of HSV color space. The multi-feature extraction is achieved through the Euclidean distance classifier.

[5] Images are retrieved by their contents such as color, texture, shape, or objects. Thus, the degree of similarity between query images and images in database can be measured by color feature extraction, texture feature extraction, shape feature extraction .similarity, or object presence between the two images .Using a single feature extraction for the image retrieval cannot be a better solution for the accuracy and efficiency .High-dimensional feature like object, semantic will reduce the query efficiency; low-dimensional feature will reduce the query accuracy, so that, better way is using multi features for image retrieval. Color, texture and shape are the most important visual features.

III. PROBLEM FORMULATION

In this paper, it is proposed to develop an efficient and a new Technique to retrieve images based on Color and Texture Feature using Quantized HSV color space and Gray level Co-occurrence Matrix (GLCM) Algorithm. At the initial state, color feature can be extracted by using HSV Color space. In HSV (Hue, Saturation, Value) color space, first quantifying the color space in non-equal intervals : H: 16 bins; S: 4 bins and V: 4 bins.Finally concatenate 16X4X4 histogram and get 256-dimensional vector ., then constructing one dimension feature vector and after that representing the color feature by finding the summation of one dimension feature vector. Color quantization is a process that optimizes the use of distinct colors in an image without affecting the visual properties of an image. In order to reduce the computation, the HSV color quantization can be used to represent the image, without a significant reduction in image quality, thereby reducing the storage space and increase the process speed. The complexity of the matching process can be reduced by using quantized color space. Because there are many colors, to reduce the complexity in histogram computation, the HSV color space needs to be quantized.

After that, the work of texture feature extraction can be obtained by using Gray-level co-occurrence matrix (GLCM) with four statistical features that is contrast, homogeneity, energy and correlation. The features are calculated in four directions (0, 45, 90, and 135 degree). In this retrieval system, Euclidean distance can be used to measure similarity of images. This retrieval system performance is measured in terms of its recall and precision.

The primary and secondary objectives of this paper are:

1. The first goal of this paper is to reduce the computation time and user interaction. The typical Content Based Image Retrieval (CBIR) systems also display the large amount of results at the end of the process which will force the user to spend more time to implement the output images. In this paper work, texture feature and color feature can be computed for the similarity between query and database images. This approach will reduce the output results to a certain levels.
2. The second goal is to reduce the Semantic gap problem in Content Based Image Retrieval System (CBIR). Generally the content based image retrieval systems compute the similarity between the query image and the database images. For this reason, chances for unexpected results at the end the retrieval process.
3. A third goal is to evaluate their performance with regard to speed and accuracy. These properties are chosen because they have the greatest impact on the implementation part.
4. A final goal has been to design and implement an algorithm for Image Retrieval System based on color and texture feature using Quantized HSV and GLCM. This can be done in high-level language or Matlab. The source code can be easy to understand. sion.

IV. METHODOLOGY

The aim of this paper is to retrieve images (based on color and texture) from a stored database depending on the query image. The entire algorithm for this image retrieval system is based on image processing. The proposed system uses MATLAB as a platform on which image processing algorithm has been developed and tested. Here, Different types of Digital images are used in database section. These are the following processes can be carried out for developing this image retrieval system.

A. Software and Hardware Requirement

a. Software Requirement

- Operating System : Windows 7
- Coding Language: : MATLAB
- Tool: MATLAB R2013a

b. *Hardware Requirement:*

- Processor: Intel Core i5
- Hard Disk : 500 GB
- System: 15.6 inch Laptop.
- Mouse: Logitech.
- Ram: 3GB

B. Collection of Images

Figure 4: Collection of Wang Dataset Image

The image used here has downloaded from http://wng.ist.psu.edu/docs/related[11]. Image Set consists of 1000 images. These images are grouped into 10 clusters with each containing 100 images of size 256*384 or 384*256. The images in the same cluster are considered as similar images.10 clusters are shown in figure 4 and Table 1

TABLE 1: 10 classes of Wang Dataset Images

Classes	Semantic Name
1	African people and village
2	Beach
3	Building
4	Buses
5	Dinosaurs
6	Elephants
7	Flowers
8	Horses
9	Mountains
10	Food

C. Proposed Method

a. Color Feature Extraction Using Quantized HSV color space

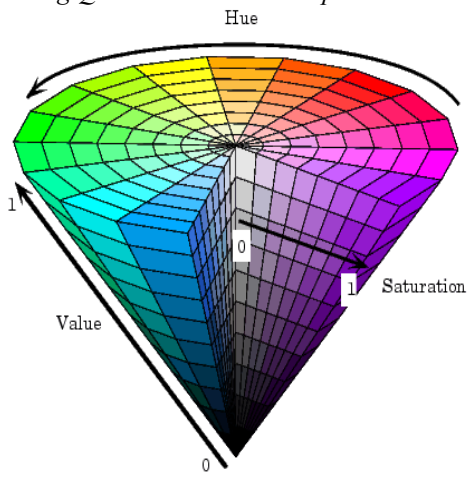

Figure 5: HSV color model

HSV defines a type of color space. The HSV color space has three components as define in Figure 5: Hue, Saturation and Value. Here Hue is used to distinguish colors. In this model, hue is an angle from 0 degrees to 360 degrees. Saturation is the percentage of white light added to a pure color. Its value is calculated from 0 to 1. When the value is '0,' the color is grey and when the value is '1,' the color is a primary color. Value is the brightness of the color and varies with color saturation. Its value is calculated from 0 to 1. When the value is 0 the color space will be totally black [5]. The advantage of HSV color space is that it is closer to human conceptual understanding of colors and has the ability to separate different components. Because of a large range of each component, if directly calculate the characteristics for retrieval, then computation will be very difficult to ensure rapid retrieval. It is essential to quantify HSV space component to reduce computation and improve efficiency. At the same time, because the human eye to distinguish colors is limited, no need to calculate all segments. According to the human color perception, Unequal interval quantization has been applied on H, S, and V components.

Algorithm for Color feature extraction using HSV color space

Step1: First convert the RGB image to HSV format
Step2: we divide color into Sixteen parts. Saturation and intensity is divided into four parts separately in accordance with the human eyes to distinguish.
Step3: On the basis of this division, it can be converted the values of Hue (H), Saturation (S) and Value (V) according to the following conditions:

$$H=\begin{cases} 0 \text{ if } € h \in [0,15] \\ 1 \text{ if } € h \in [16,25] \\ 2 \text{ if } € h \in [26,45] \\ 3 \text{ if } € h \in [46,55] \\ 4 \text{ if } € h \in [56,80] \\ 5 \text{ if } € h \in [81,108] \\ 6 \text{ if } € h \in [109,140] \\ 7 \text{ if } € h \in [141,165] \\ 8 \text{ if } € h \in [166,190] \\ 9 \text{ if } € h \in [191,220] \\ 10 \text{ if } € h \in [221,255] \\ 11 \text{ if } € h \in [256,275] \\ 12 \text{ if } € h \in [276,290] \\ 13 \text{ if } € h \in [291,316] \\ 14 \text{ if } € h \in [317,330] \\ 15 \text{ if } € h \in [331,345] \end{cases}$$

$$S=\begin{cases} 0 \text{ if } € s \in [0,0.15] \\ 1 \text{ if } € s \in [0.16,0.40] \\ 2 \text{ if } € s \in [0.41,0.75] \\ 3 \text{ if } € s \in [0.76,1] \end{cases}$$

$$V=\begin{cases} 0 \text{ if } € v \in [0,0.15] \\ 1 \text{ if } € v \in [0.16,0.40] \\ 2 \text{ if } € v \in [0.41,0.75] \\ 3 \text{ if } € v \in [0.76,1] \end{cases}$$

Step4: In accordance with the quantization level above, the H,S,V three-dimensional feature vector for different values of with different weight to form one-dimensional feature vector named

$$G = QsQvH + QsS + V \quad (1)$$

Where *Qs* is quantified series of S, *Qv* is quantified series of V.
Here, I am going to use Qs=Qv=4, applying this value in equation 1

$$G = 16H + 4S + V \quad (2)$$

(This is called color feature)
Step5: After converting 3D to 1D then find the Histogram
Step6: Apply Euclidean Distance for finding Similarity between the images.

b. Texture Feature Extraction Using GLCM

One of the most important texture analysis algorithms is the GLCM method, which is a two dimensional matrix that represents the number of co-occurrences to the pair of pixels, or two sets of pixels which have a gray scale value and are related by specific relationships [3]. GLCM is composed of the probability value it is defined by p(i,j|d,θ) which expresses the probability of the couple pixels at θ direction and d interval. When θ and d is determined p(i,j|d,θ) is showed by pi,j. Distinctly GLCM is a symmetry matrix, it's level is determined by the image gray level. Elements in the matrix are computed by the equation showed as follows:

$$p(i,j \mid d,\theta d = p(i,j \mid d,\theta) \; / \sum i . \sum j (p(i,j \mid d,\theta)) \quad (3)$$

Where,

i, j = 0,1,…..L-1 are the gray levels
θ = direction
d = interval

The algorithm consists of the following steps:

Step1: Convert the RGB image to Gray Format.
Step2: Compute four GLCM matrices (directions for θ = 0,45, 90,135 degree) as given by eq. (3)
Step3: For each GLCM matrix, compute the statistical features Energy (Angular second moment), Entropy(ENT), Correlation(COR), Contrast(CON) as follows where P(i,j) is probability density.

Energy (Angular Second Moment (ASM)): Energy measures textural uniformity (i.e. pixel pairs repetations).

$$ASM = \sum\sum p2(i,j) \quad (4)$$

Contrast (CON): Contrast indicates the variance of the gray level

$$CON = \sum\sum (i-j)2\, p(i,j) \quad (5)$$

Entropy (ENT): This parameter measures the disorder of the image. For texturally uniform image, entropy is Small

$$ENT = -\sum\sum p(i,j) \log [p(i,j)] \quad (6)$$

Correlation (COR):

$$COR = (\sum\sum ijp(i-j) - \mu x \mu y) / \sigma x\, \sigma y \quad (7)$$

Where μx, μy, σx, σy are the means and standard deviations of px and py respectively.
px is the sum of each row in co-occurance matrix
py is the sum of each column in the co-occurance matrix.

Step4: Compute the feature vector using the means and variances of all the parameters. Thus, the feature vector f={μASM, μENT, μCOR, μCON, σASM, σENT, σCOR, σCON } Where μ is mean and σ is variance of the parameters. (This is the texture feature)
Step5: Apply Euclidean Distance for finding Similarity between the images

(Euclidian Distance: The distance between two images is used to compare and find the similarity between query image and the images in the database. Finding the distance between the feature vectors is similar to that of finding the similarity between the feature vectors [9]. Here, Euclidean distance can be used. Let P(p1,p2…….pn)and Q(q1,q2…qn) are two points in an n – dimensional space. Then the distance can be calculated as follows:

$$D(P,Q) = \sqrt{(p1-q1)2+(p2-q2)2+.....+(pn-qn)2} \quad (8)$$

)

c. Color and Texture Feature Extraction Using HSV Color Space and GLCM

Algorithm:

Step1: Finding color feature using HSV, Fcolor
Step2: Finding Texture feature using GLCM, Ftexture
Step3: Euclidian distance Formula for both color and Texture feature extraction

$$D(color, texture) = (w1\sqrt{2} \times (\sqrt{2}- Fcolor)) + (w2\sqrt{2} \times (\sqrt{2}- Ftexture)) \quad (9)$$

Where, I am going to use w1=w2=0.5 (for better retrieval performance)
Block Diagram for this Proposed Method is shown in Figure 6:

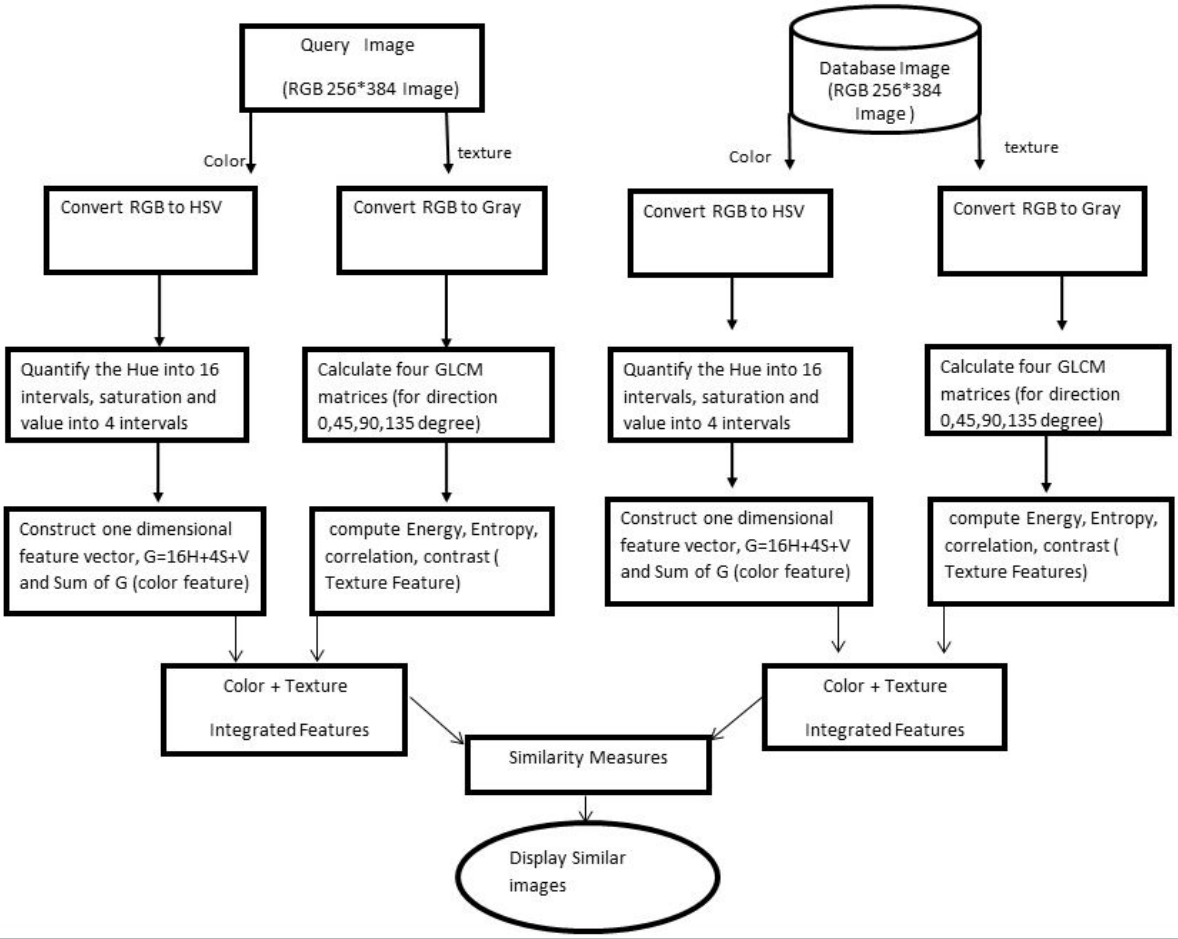

Figure 6: Proposed System Block Diagram

V. CONCLUSION

In this review paper, an efficient color and texture feature based image retrieval technique using HSV and GLCM is proposed. The basic concept of this without having any laborious work of typing keywords, we can use input as an image and can retrieve required images based on color and texture features. Here Euclidean distance can be used for similarly measurement. In the simulation part, 12 images can be used for retrieval result. The ultimate goal of this paper is to achieve higher retrieval efficiency from large database of images by improving the speed, efficiency and accuracy.

ACKNOWLEDGMENT

I, Nishad Ahmed, student of M.Tech, 4th Semester, Department of Information Technology, Gauhati University (G.U), have partially completed my review paper entitled "Image Retrieval System based on Color (HSV) and Texture(GLCM)" under the guidance of Parismita Sarma, Assistant Professor, Department of Information Technology, G.U. I would like to express my sincere gratitude and appreciation to all those who gave me the possibility to complete this paper. Special thanks to my guide Parismita Sarma Madam, our Head of the Department Mr. Mirzanur Rahman sir, and all the teachers and researchers related to this paper, whose stimulating suggestions, coordination and encouragement helped me a lot.

REFERENCES

[1] Ch.Kavitha, Dr. B.Prabhakara Rao, Dr. A.Govardhan," Image Retrieval based on combined features of image sub-blocks", Ch.Kavitha et al. / International Journal on Computer Science and Engineering (IJCSE)
[2] N Raghu Ram Reddy, Gundreddy Suresh Reddy, Dr.M.Narayana," Color Sketch Based Image Retrieval" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering ,Vol. 3, Issue 9, September 2014
[3] Benco M. and Hudec R.," Normal Method for Texture Feature extraction Based on GLCM", Department of Telecommunication, University of Zilina, Slovak Republic.
[4] Md. Baharul Islam, Krishanu Kundu, Arif Ahmed ," Texture Feature based Image Retrieval Algorithms" ,International Journal of Engineering and Technical Research (IJETR) ISSN: 2321-0869, Volume-2, Issue-4, April 2014
[5] Cheruvu Mahammad Ibraheem , Dr. G.Umamaheswara Reddy, "Content based Image retrival system using HSV color ,Shape and GLCM Texture", International Journal of Advanced Research in Computer and Communication Engineering,Vol. 4, Issue 10, October 2015
[6] Swapnalini Pattanaik, Prof.D.G.Bhalke ,"Beginners to Content Based Image Retrieval"International Journal of Scientific Research Engineering &Technology (IJSRET) Volume 1 Issue2 pp 040-044 May 2012.
[7] Ruziana Mohamad Rasli, T Zalizam T Muda, Yuhanis Yusof, Juhaida Abu Bakar, "Comparative Analysis of Content Based Image Retrieval Technique using ColorHistogram. A Case Study of GLCM and K-Means Clustering",Third International Conference on Intelligent Systems Modelling and Simulation,2012
[8] Fan-Hui Kong, "Image Retrieval Using Both Color And Texture Features", Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding, 12-15 July 2009.
[9] Pushpa B. Patil, Manesh B. Kokare," Relevance Feedback in Content Based Image Retrieval: A Review", Journal of Applied Computer Science & Mathematics, no. 10 (5) /2011, Suceava
[10] http://wng.ist.psu.edu/docs/related
[11] R. C. Gonzalez, "Digital Image Processing in Matlab". Edition, Publisher
[12] CBIR, http//en.wikipedia.org/wiki/Content-based image retrieval