International Journal of Science Engineering and Advance Technology

Adaptive Query Image Searching Methods Using Improved Hash Coding and Unified Color and Intensity level Matrices

[1] **Brahmaiah Naik Jatothu** [2] **Sitaramanjaneyulu.P**

[1]Assistant Professor, Department of Electronics & Communication Engineering, Vignan's Lara Institute of Technology & science, Guntur-522615, (Research scholar, JNTU, Kakinada

[2]Assistant Professor, Department of Electronics & Communication Engineering, Pragati Engineering College, Surampalem-533437

Email: [1]*brahmaiahnaik@gmail.com*,[2] *sriram128@gmail.com*

Abstract: In this thesis we had implemented Hash coding with bit wise hamming distance for similarity measurement between the feature vectors of query and database, which will be more secured and real time based application. By using the hash coding method the computational time is high and the accuracy will be much lower. Hence, hash coding will be extended with inclusion of gray scale matrices (GSM), which will represent the features of gray scale texture values in the image, but it will not consider the color information in the image. To incorporate both gray scale and color information, we proposed a novel scheme in which both color and intensity variations are represented in a single composite feature known as Unified Color and Intensity level Matrices (UCILM). The proposed scheme has been merged with hash coding to improve the efficiency of the retrieval system. This addition work had improved the system accuracy as well as the precision time with almost 88% of relevant image retrieval.

I. INTRODUCTION

Recent years there is a rapid growth in searching engines such as Bing image search: Microsoft's CBIR engine (Public Company), Google's CBIR system, note: does not work on all images (Public Company), CBIR search engine, by Gazopa (Private Company), Imense Image Search Portal (Private Company) and Like.com (Private Company), image retrieval has become a challenging task. The interest in CBIR has grown because of the retrieval issues, limitations and time consumption in metadata based systems. We can search the textual information very easily by the existing technology, but this searching methods requires humans to describe each images manually in the database, which is not possible practically for very huge databases or for the images which will be generated automatically, e.g. images generated from surveillance cameras. It has more drawbacks that there is a chance to miss images that use different equivalent word in the description of images. The systems based on categorizing images in semantic classes like "tiger" as a subclass of "animal" can debar the miscatergorization problem, but it will requires more effort by a use to identify the images that might be "tigers" , but all of them are categorized only as an "animal". Content-based image retrieval (CBIR) is a application of methods of acquisition, pre-processing, analyzing, representation and also understanding images to the image retrieval problem, that is the problem of exploring for digital images from large databases. The CBIR system is opposed to traditional approaches, which is known on concept based approaches i.e., concept based image indexing (CBII) [1].

II.Related Work

In the past decades several CBIR systems have been proposed, and still the researchers are focusing on developing extended CBIR systems with more effective results. The letter proposed in [4] gives a comparison of different approaches of CBIR based on similarity measures and image features to identify the similarity between the images, which provides accurate information for retrieving the relevant images from large database. Wan Siti *et.al* proposed in [5] compares the several medical image retrieval systems based on the feature extraction and to improve the effectiveness of the CBIR system for medical images such as magnetic resonance (MR) images and computed tomography (CT) images [10]. The major concept proposed in [5] is to help in the diagnosis such as to find the similar disease and monitoring of patient s progress continuously. B. S. Manjunath *et.al* presented in [6] is the combination of color, texture with inclusion of edge compactness for Motion Picture Expert Group (MPEG)-7 standards. Another approach proposed in [7] used different color spaces such as HSV and YCbCr explains a similar approach based on color and texture analysis. The work proposed in [8] introduces a new retrieval system which has done by using wavelet transformation with both color and texture features together and will perform better than existed state of art algorithms.

Recently, retinal image retrieval system called CBIR for retinal and blood vessels extraction [9] has been analyzed by the histogram features of RGB color components. The multi resolution analysis has applied to the image to acquire the texture information. In addition to improve the performance, morphological operations are applied to study the shape of object. Swati Agarwal has proposed a new CBIR system in [11], which is by using discrete wavelet transform and edge histogram descriptor (EHD). Here the retrieval is based on color and texture features not by using color information in the image, input image first decomposes the input query image into several sub bands i.e., approximation coefficients and detail coefficients, where detail coefficients consists of horizontal (LH), vertical (HL) and also the diagonal information (HH) of the image. Afterwards, EHD is used to gather the information of dominant edge orientations. This mixture of 3D-DWT and EHD will improve the efficiency of the CBIR system.

In this paper, a novel method for CBIR system "An adaptive query image searching using hash coding and gray scale matrices" has been proposed to improve the system efficiency and precision time. In this letter three methods have been proposed first, Hash coding. Second hash code and GSM. Third, hash coding and CCG for extracting the both color and grayscale texture information from the images. First, has code of the image will be calculated by using the different hash code algorithms such as MD2, MD5, SHA-1, SHA-256, SHA-384, SHA-512 with the format of double, hex, uint8, or base64. After calculating the hash codes then generate a feature vector and then create feature vectors of database images. Now, calculate the distance between feature vector of the query image and data base images, if the distance is small enough then the corresponding image will be matched with the query image. Distance between the feature vectors has been calculated by bitwise hamming distance, which gives more effective performance rather than Euclidean distance.

II. PROPOSED CBIR SYSTEM

For image retrieval, classification and indexing both color and texture have been used widely in various applications. Histogram of a image is a graphical analysis of a image, which represents the color information of image. It is a first order statistical measure. The major drawback of this histogram based approaches is that the spatial distribution and local variations will be ignored. Local spatial variation of pixel intensity is commonly used to capture texture information in an image. Grayscale Matrix (GSM) is a well-known method for texture extraction in the spatial domain. A GSM stores the number of pixel neighborhoods in an image that have a particular grayscale combination. The GSM matrix will be generated for each and every possible neighborhood direction in an image, namely, 0, 45, 90 and 135. By combining all these four matrices, Average and range of 14 features like correlation, contrast, angular second moment (ASM), energy etc., are generated to get a total of 28 features. In this approach, we are considering only gray scale information which in results losing of color information. To comprise spatial or grayscale information along with the color information of image pixels, color correlogram (CCG) has been proposed, which is 3D matrix to represent the probability of finding pixels of any two given colors at a distance 'd'. CCG has been divided into two methods auto CCG and cross CCG. Auto CCG is a correlogram variation, which represents the same color probability to find the pixels. Where cross CCG represents the probability for finding the pixels with the different colors. This approach can effectively represent color distribution in an image.

The color and intensity variations are represented in a single composite feature known as Unified Color and Intensity level Matrices (UCILM). The proposed scheme has been merged with hash coding to improve the efficiency of the retrieval system. The existing systems treat both color and intensity separately. Where, the proposed scheme provides both informations in a single feature known as UCILM. The main advantage of using UCILM is that it avoids the use of weights to combine individual color and texture features. We use UCILM feature in an image retrieval application from large image databases.

a. HSV Quantization

Here in the proposed algorithm we used HSV color space to distinguish the different ranges of colors for human perception. HSV is known as Hue (H), Saturation (S) and Value (V), which is used to differentiate the different range of colors in image. 'H' defines the purity of color, 'S' is the amount of white color added to the true color and 'V' is the intensity value of the pixel. Due to the large range of color components, it is very difficult to calculate the characteristics for the retrieval of images from database. This results in high computational time. Hence, it is essential to quantify HSV space component to reduce computational time and to improve the efficiency of CBI retrieval system. We divided color into sixteen parts. Saturation and intensity is divided into three parts separately in accordance with the human eyes to distinguish. Hue (H) has been divided into twelve parts, saturation (S) into four parts and intensity value (V) into three parts respectively. In accordance with the quantization level above, the H, S, V three-dimensional feature vector for different values of with different weights to form one dimensional feature vector and is given by the following equation:

$$CCG = Q_s * Q_v * H + Q_v * S + V \quad (1)$$

Where Q_s is the quantized series of S and Q_v is the quantized series of V.

Now by setting $Q_s = 4$ and $Q_v = 3$, Then

$$CCG = 12H + 3S + V \quad (2)$$

$$H = \begin{cases} 0 \; if \; h \in [316,20] \\ 1 \; if \; h \in [21,40] \\ 2 \; if \; h \in [41,75] \\ 3 \; if \; h \in [76,110] \\ 4 \; if \; h \in [111,145] \\ 5 \; if \; h \in [146,173] \\ 6 \; if \; h \in [174,195] \\ 7 \; if \; h \in [196,220] \\ 8 \; if \; h \in [221,245] \\ 9 \; if \; h \in [246,270] \\ 10 \; if \; h \in [271,295] \\ 11 \; if \; h \in [296,315] \end{cases}$$

$$S = \begin{cases} 0 \; if \; s \in [0,0.2) \\ 1 \; if \; s \in [0.2,0.5) \\ 2 \; if \; s \in [0.5,0.7) \\ 3 \; if \; s \in [0.7,1) \end{cases}$$

$$V = \begin{cases} 0 \; if \; v \in [0,0.2) \\ 1 \; if \; v \in [0.2,0.7) \\ 2 \; if \; v \in [0.7,1) \end{cases} \quad (3)$$

In this way, 3 component vector of HSV form one dimensional vector, which quantizes the entire color space for the 163 kinds of main colors. So, we can handle 163 bins of one dimensional histogram. By doing this we can reduce the effects of light intensity and computational time even complexity of the system. The proposed scheme has been described briefly in fig.1; it explains how the proposed scheme will be superior to the existing schemes. The steps involved in proposed scheme are as follows:

Step1: Select and read the query image from the database.

Step2: Apply Hash coding to the query image by using 'MD5' hash code method and count the number of bits in the hash code then store them in a feature vector 'f1'.

Step3: Now, apply GSM to the query image and then normalize these values to store it in the feature vector 'f2'

Step4: Then apply CCG to the query image and also normalize these values to make a feature vector as 'f3'

Step5: Read all the database images and apply all the steps which have been applied to the query image to find the feature vectors for hash coding, GSM and CCG

Step6: Now combine the GSM and CCG to get the both gray scale and color information in a single feature called "Unified color and intensity level matrices (UCILM)" then normalize the UCILM values to form a feature vector

Step7: Combine all the feature vectors and make it into a single feature vector for query image features and database images as well

Step8: Calculate the Hamming distance for the similarity check, when the similarity between the query and database images will be more than 85%

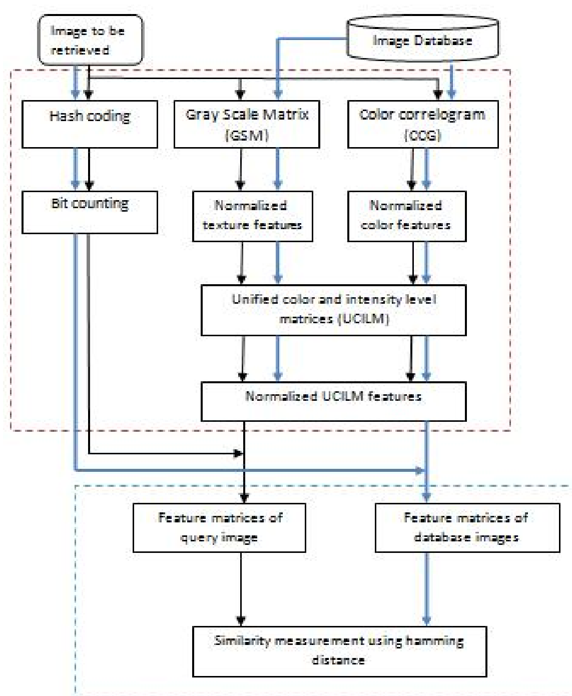
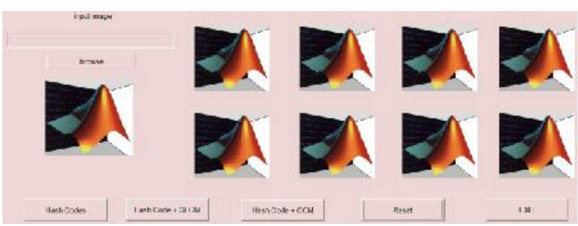


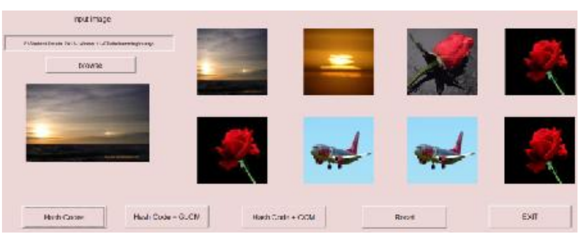
Fig.1. Proposed scheme of CBIR system

III. SIMULATION RESULTS AND DISCUSSION

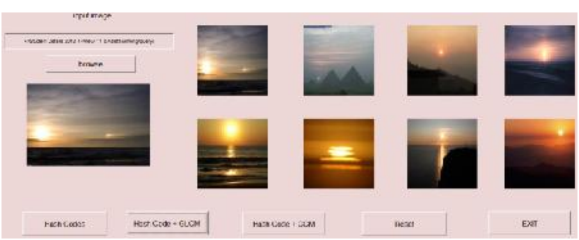
In this section we discussed the simulation results of CBIR system based on the proposed and conventional schemes such as hash codes, GSM and CCG. The proposed algorithm has been tested with few databases and displayed the outputs in the below figures. Fig2 (a) shows that the initial GUI figure of proposed model, 2(b) shows that the output of Hash coding scheme, which will be used as retrieving the images from the given database. GSM retrieval system has been shown in fig2(c) and proposed model will be shown in 2(d), which has got more precision in time and accuracy. This CBIR system has been tested with few more images such as tiger, Rose, and sky etc. the remaining figures shows that the retrieval of the given input relevant images.



(a)



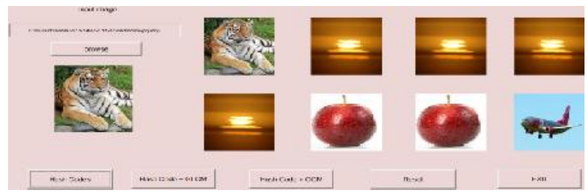
(b)



(c)

(d)

Fig.2 (a) Proposed scheme with GUI model (b) using Hash coding (c) using GSM+Hash (d) using UCILM



IV. CONCLUSION

In this letter we had proposed an adaptive CBIR scheme for large database systems. By using the hash coding and unified color and intensity level matrices (UCILM), the performance of the CBIR system had improved. UCILM is used to achieve both color and gray level information from the image pixels and also we calculated hash codes for improving the accuracy and to reduce the computational complexity while improving the system efficiency. The proposed system has proven that this approach has got superior performance than the existing CBIR schemes.

REFERENCES

1. Chua, T.-S., Hung-ICeng Pung; Guo-Jun Lu; Hee-Sen Jong, "A concept based image retrieval system" *Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences, Jan. 1994.*
2. Eakins, John; Graham, Margaret, "Content-based Image Retrieval", University of Northumbria, New castle
3. X Wang, T S Chua and Al-Hawamdeh, "Probabilistic and semantic based retrieval in hypertext", *Proc. of the South-East Asia Regional Computer Conf.*, pp. 25.01 -25.17 1992
4. Khan, S.M.H., Hussain, A. ; Alshaikhi, I.F.T., "Comparative Study on Content Based Image (CBIR)," International Conference in Advanced Computer Science Applications and Technologies (ACSAT), 2012.
5. Wan Siti, H. Munirah, W. Ahmad, M. Faizal and A. Fauzi, "Comparison of Different Feature Extraction Techniques in Content Based Image Retrieval For CT Brain Images," 10Th IEEE workshop on Multimedia Signal Processing, pp. 503-508, 2008.
6. B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, A.Yamada, "Colour and Texture Descriptors," IEEE Transactions on Circuits and Systems for Video Technology, 1998
7. P. S. Hiremath, S. Shivashankar, J. Pujari, "Wavelet Based Features for Color Texture classification with Application to CBIR," International Journal of Computer Science and Network Security, Vol. 6, No.9A, September 2006
8. Tian Yumin, Mei Lixia, "Image Retrieval Based on Multiple Features Using Wavelet," 5th IEEE International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'03), 2003.
9. M. R. Zare, R. N. Ainon, W. C. Seng, "Content-based Image Retrieval for Blood Cells," Third Asia International Conference on Modeling & Simulation, 2009
10. Prasad, B.G., Krishna, A.N., "Statistical Texture Feature Based Retrieval and Performance Evaluation of CT Brain Images" 3rd International Conference on Electronics Computer Technology (ICECT), Vol. 2, April 2011
11. Swati Agarwal, A.K. Verma, Preethvanti Singh, "Content Based Image Retrieval using Discrete Wavelet Transform and Edge Histogram Descriptor" *International conference on Information Systems and Computer Networks, 2013*
12. D. Ashok Kumar, J. Esther, "Comparative Study on CBIR based by Color Histogram, Gabor and Wavelet Transform," International Journal of Computer Applications (0975 – 8887) Volume 17– No.3, March 2011.