Content based Image Retrieval based on Histogram

Mohammed Hamzah Abed
Iraq, University of Kufa
Faculty of Education for women
Computer Science Department

Dawood Salman Jasim Al-Farttoosi
Iraq, University of Kufa
Faculty of Administration and Economics

ABSTRACT

Content based image retrieval (CBIR) systems analyze the visual content and find images in the database. this paper has been propose a novel approach to CBIR system based on prime's histogram of the color image .This paper describe a project that implements by using Matlab and test three techniques depend on the simple color histogram based search and retrieve algorithm for images. The study find the technique to be effective between the three techniques that used.

Keywords

Content Based Image Retrieval, Primes, Euclidean distance, and histogram.

1. INTRODUCTION

In last few years with the development of computer technology and network communication technology, specially the emergence and popularization of Internet, and social media like facebook ,Instagram and etc, rapid expansion of the size of the multimedia database such as digital image, every day hundreds of millions of military or civilian image data is stored in the database. The main method of image files is to establish Keywords or text description of the title as well as some additional information, and then establish a link between storage path and the keywords of the image, which is text-based image retrieval. However, with the storage capacity of images to start using GB or TB, the own shortcomings of text-based image retrieval technology led to two difficulties in the retrieval: First, it has been impossible to note each image; second, the subjectivity and non-precision of image annotation may lead to the adaptation in the retrieval process. In order to overcome these problems, Content-Based Image Retrieval (CBIR) was proposed in the 1990's [1].Contents Based Image Retrieval (CBIR) is a very important and prominent area in image processing due to its diverse applications in internet, multimedia, companies image archives and medical image archives. Improved demand for image databases has increased the need to store and retrieve digital images. And the most important features or image component that used in Content Based Image Retrieval (CBIR) are texture, color and shape [1, 2].

2. IMAGE RETRIEVAL BASED ON COLOR

Content-Based Image Retrieval (CBIR) According to the user-supplied bottom characteristics, directly find out images containing specific content from the image database. The basic process: First of all, do appropriate preprocessing of images, and then extracted image features needed from the image according to the contents of images to keep in the database. When retrieval has been occur to identify the image, extract the corresponding features from a known image and then retrieve the image database to identify the images which are similar with it, also it can give some of the characteristics and features based on a queried requirement, then retrieve out the required images based on the given eigenvalues. In all retrieval process features extraction is the most sensitive thing, because depend on the features and got the good matches image from the image's database [1, 3].

3. VISUAL FEATURES USED IN RETRIEVAL

Visual features were classified in into primitive features such as color or shape, logical features such as identity of objects shown and abstract features such as significance of scenes depicted [4].Color has been the most useful and effective feature and almost all systems employ colors. Although most of images that used through internet and social media are in the RGB (Red, Green, Blue) color space, this space is only rarely used for indexing and querying as it does not correspond well to the human color perception. It only seems reasonable to be used for images taken under exactly the same conditions each time such as trademark images. Other spaces such as HSV (Hue, Saturation, and Value) spaces are much better with respect to human perception and are more frequently used. This means that differences in the color space are similar to the differences between colors that humans perceive [4].Texture partly due to the imprecise understanding and definition of what exactly visual texture actually is, texture measures have an even larger variety than color measures. Some of the most common measures for capturing the texture of images are wavelet sand Gabor filters. These texture measures try to capture the characteristics of the image or image parts with respect to changes in certain directions and the scale of the changes. This is most useful for regions or images with homogeneous texture [3, 4]. Shape is an important visual feature and it is one of the basic features used to describe image content. However, shape representation and description is a difficult task. This is because when a 3-D real world object is projected onto a 2-D image plane, one dimension of object information is lost [5].

3.1 Color Histogram

Content The intensity histogram shows how individual brightness levels are occupied in an image; the image contrast is measured by the range of brightness levels. The histogram plots the number of pixels with a particular brightness level against the brightness level. For 8-bit for each color band in all pixels, the brightness ranges from zero (black) to 255 (white) [6]. Color histogram represents the joint distribution of three color channels in an image (Red, Green and Blue). Therefore it characterize the global color information in an image [7]. In this proposal work will used RGB color space and for each band will calculate the set of features depend on the histogram. The histogram is used a model of the probability distribution of gray level. The first order histogram probability P (g) is defined as follow:

$$p(g) = \frac{N(g)}{M} \qquad ..1$$

Where P(g) a probability of a gray level g in image

N (g) number of pixel with gray level g in image

M total number of pixel in image

3.2 Histogram of Prime Value

A histogram for 8 bits per each color of the pixel 24 bit per pixel start from black   color with 0 value into 255 (white) and between 0 and 255 there is 54 prime number. In some cases in this proposal work the retrieval based on prime's values will depend on these values of primes and their frequency to compare between query image and database's images.

3.3 Color Image Moment

Image moments and their functions have been used as features in many image processing applications, viz., pattern recognition, image classification, target identification, and shape analysis. Moments of an image are treated as region-based shape descriptors [2]. One of the moment that paper considered it in this work is mean is the average value, so tells us something about the general brightness of the image. The mean can be defined as follows:

$$Mean = \frac{1}{N} * \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{M} f(i,j) \quad \dots 2$$

And the second moment has been used it is a standard deviation which also known as the square root of the variance tell something about the contrast. It describes the spread in the data, image with high contrast should have a high standard deviation and it is defined as

$$\sigma = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{M} (x(i,j) - \mu)^2} \quad \dots 3$$

And the third and fourth features that consider it is variance and entropy is a measure of how evenly energy is distributed in an image or data, entropy provides a measure of the amount of energy that cannot be used to do work and it can be formulated

$$e = -\sum_{i=1}^{L} P(zi) \log P(zi) \quad \dots 4$$

4. PROPOSED WORK

In this proposal work have been used three ways to retrieve image from database of images all of them depend on color histogram with add some extra features to increase the efficiency of work. In general the algorithm that followed it in this work as the following outline steps with in each time different features used.

1. Read images in database and extract RGB format pixel information from images.

2. Extract the features from the images in database.

3. Read in a query image and extract RGB format pixel information

4. Create histograms for each of the RGB components of the query image.

5. Compute a Euclidean distance by comparing the query image histograms to that of each image in the database.

6. Sort images in database in order of ascending Euclidean distance to query image and return as result, and the image with smallest Euclidean distance is the optimal match.

Note the different techniques that used it depend on the features and how can extract it from the images. And the figure (1) show the proposal work.
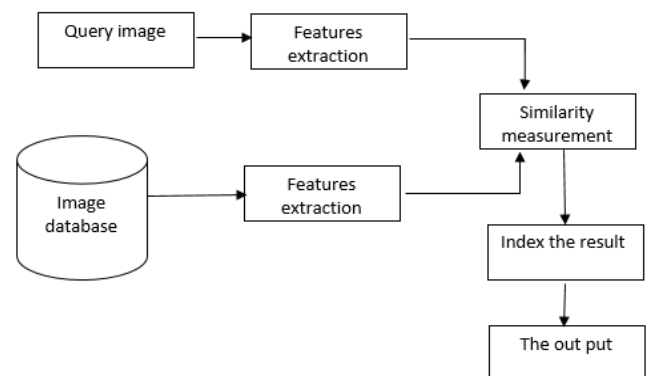

Fig 1: image retrieval proposal work

And the features have been used in the proposal work in the paper as following:

4.1 Image Retrieval using Color Histogram

A color image histogram with 24 bits per pixel starting from 0 to 255 from black to white for each color band (Red , Green and Blue) and because the changing in color value + 15 or -15 not visible with human eyes. So in the first Experiment work in this paper divided histogram into sub block histogram with 17 block each block contain 15 gray scale color for each of the RGB component of each image read from database. Thus each image will have 3 histogram associated with it.

4.2 Image Retrieval using Prime value of Color Histogram

The second experiment in our work is take the frequency of the primes number between 0-255 for all images in database and query image and apply the Euclidian distance between two features vector for each color (R,G and B) so in the actual work will compare 162 value of each vector and the smallest Euclidian value select it as the optimal match.

4.3 Image Retrieval using Color Moment

The third experiment depend on the histogram value for three band (Red, Green and Blue) will compute the set of moment (mean , Entropy , Variance and Standard Deviation)of the color image for each band and apply the measurement distance to check which image is more closer to the query image from database.

5. SIMILARITY MEASUREMENT

In the retrieval process , features of the query image specification are compare with features of the image database to determine which image match correctly (similar) with the given features. The matching process is based on similarity measure between query image and images in database. There are several similarity measures that are proposed and used for varied purpose. For retrieval base on similarity of two features distance between the two can be defined as a match measure [8]. It is the way that compute the similarity measure between query image vector and the vector of database's images, where each vector contain the features that have been selected it in different proposal work in this paper have been used the Euclidean distance to select which images are closer to the query image.

5.1 Euclidian Distance

It's the most common metric for measuring between two vectors and is given by the square root of the summation of the squares of the different between vectors component

$$\text{Euclidian Distance (A. B)} = \sqrt{\sum_{i=1}^{n}(ai - bi)^2} \quad \ldots 5$$

Where A = [a1, a2, a3… an]

B = [b1, b2, b3,…,bn]

Smallest Euclidian Distance between two vectors of features means the most probable matching and will be selected.

6. IMAGE DATABASE COLLECTION

In this paper work used a large number of image collections are being used for image retrieval research, this collection typically consist of a rose , animal ,plant ,.. etc. our database consist of images and their compression by Using SPIHT Algorithm [9] with 2 ,4 and 8 level of compression also cropping image and some editing like add some text on the image and some noise.

7. EXPERIMENTS RESULT

In the following section will discuss the experiment result with tables to show the Euclidian distance and figure to show the result that implement it by using Matlab.

7.1 First Experiment Block Color Histogram

The first experiment result as describe before in 4-1, the histogram sub divided into 17 block histogram for each color (Red, Green and Blue) of the image in database ,and for query image and apply the Euclidian distance between these two vector and get the smallest distance as the perfect result. Table 1 show the example of period of the histogram and there frequency, note this frequency for each color that means will compare 17 values for each color band.

Table 1. Show example of block histogram and frequency

Range	Frequency	Range	Frequency
0-15	454	136-150	856
16-30	2324	151-165	45
31-45	345	166-180	454
46-60	903	181-195	355
61-75	133	196-210	31
76-90	563	211-215	4546
91-105	123	216-230	456
106-120	67	231-255	3456
121-135	124		

And Table.2 show the result of the Euclidian of the first experiment result

Table 2. Result of first Experiment

| Image name	Euclidian distance	Description
1.jpg	0	Same the original image
2. jpg	292.513247563251	Same original image with text
5. jpg	1407.51412071069	original image with compression SPIHT 2level
4. jpg	2468.50865908953	original image with compression SPIHT 4level
3. jpg	8862.9185373668	original image with compression SPIHT 8level
7. jpg	21372.4232598926	Original image with resize and add text
55. jpg	21492.9940445718	Original image with resize
63.jpg	21504.8009988467	Original image with resize and rotate with 90 degree
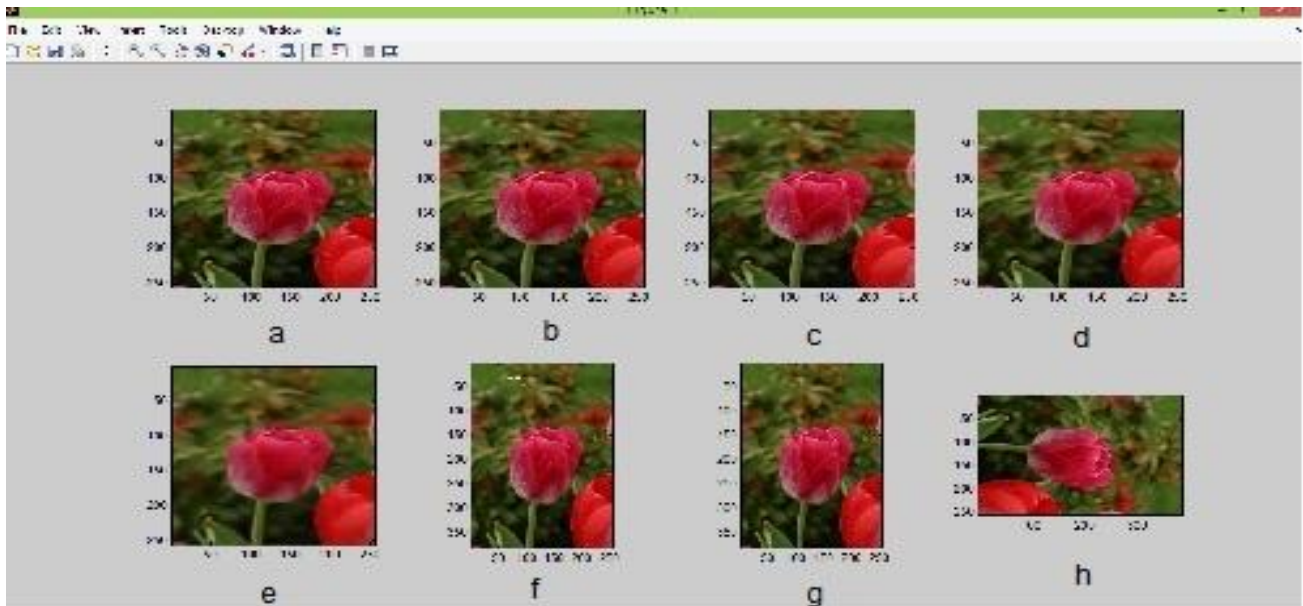
Fig 2: show the result of the first experiment (a) original image (b) original with text (c) SPIHT compression 2level (d) SPIHT compression 4level (e) SPIHT compression 8level (f) original image with resize and text (g) original image with resize (h) original image with resize and 90 rotate

7.2 Second Experiment Primes Histogram

In the Second experiment the features that will use as a image's features is histogram of prime's number to find the matching image from database as describe in section 4-2.and table 3 show the result of Euclidian distance for the second experiment.

Table 3. Result of Second Experiment

Image name	Euclidian distance	Description
1.jpg	0	Same the original image
2. jpg	292.513247563251	Same original image with text
5. jpg	1407.51412071069	Same original image with compression SPIHT 2level
4. jpg	2468.50865908953	Same original image with compression SPIHT 4level
7. jpg	21372.4232598926	Original image with resize
63.jpg	2882.03868815115	Original image with resize and rotate with 90 degree
73.jpg	2939.99642856926	Tree
55.jpg	3063.08831083924	Original image with resize

And the figure 3 show the image result of the second experiment

7.3 Third Experiment

The third experiment result in this paper work is to retrieve image depend on the moment that extract it from image in database and the query image , table 4 show the result of the third experiment.

Table 4. Result of Third Experiment

Image name	Euclidian distance	Description
1.jpg	0	Same the original image
2. jpg	292.513247563251	Same original image with text
5. jpg	1407.51412071069	Same original image with compression SPIHT 2level
4. jpg	2468.50865908953	Same original image with compression SPIHT 4level
3.jpg	17857.1717533517	Same original image with compression SPIHT 8level
66.jpg	82975.787131424	Red rose picture
78.jpg	82975.787131424	Red rose picture
27.jpg	98613.2724870459	Tree with small rose

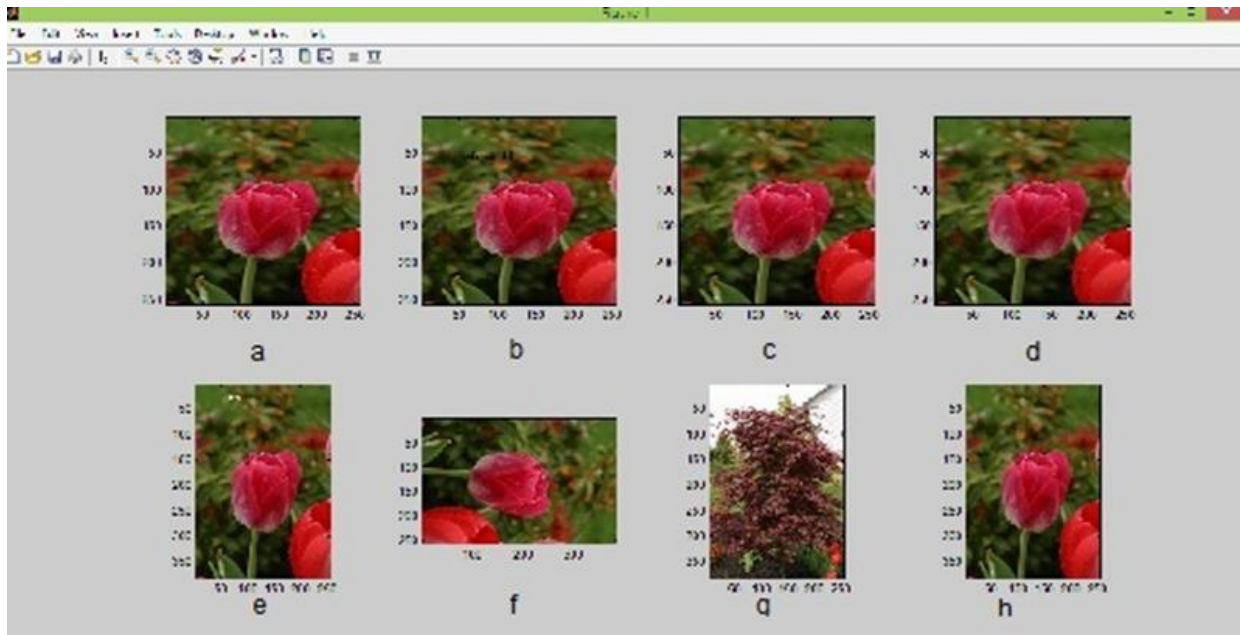And the figure 4 show the result of the third experiment.

Fig 3: second experiment result (a) original image (b) original with text (c) SPIHT compression 2level (d) SPIHT compression 4level (e) original image with resize and text (f) original image with resize and 90 rotate (g) tree image (h) original image with resize
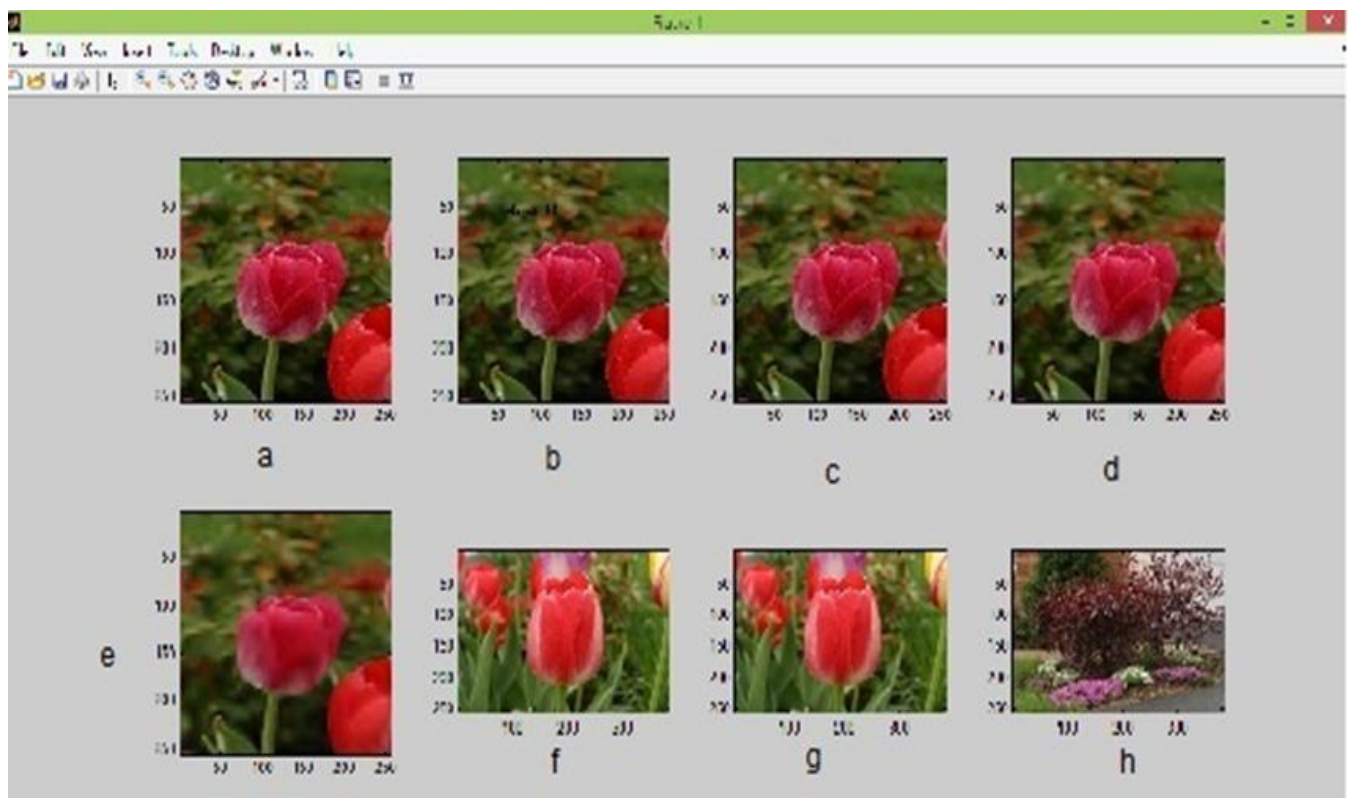

Fig 4: result of the third experiment a) original image (b) original with text (c) SPIHT compression 2level (d) SPIHT compression 4level (e) SPIHT compression 8level (f) rose image1 (g) rose image 2 (h) tree image

And figure 5 show the compare between three experiment results by compare the result of Euclidian distance of each one
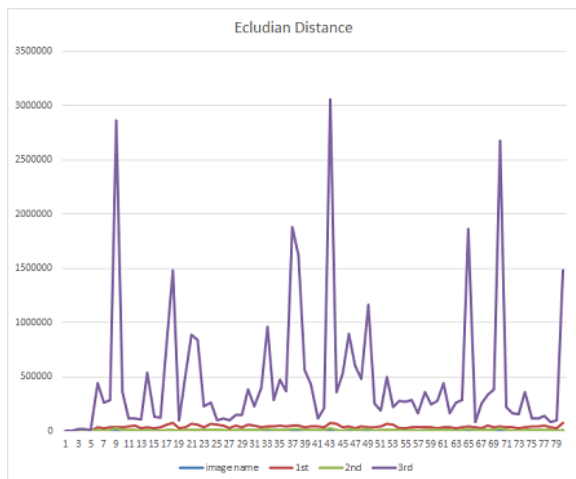

Fig 5: Compare between three experiment result by using Euclidian Distance

8. CONCLUSION

Our study implementation of a color histogram based image retrieval system identified numerous strengths in the algorithm performance as an image retrieval system. Firstly, the algorithm is relatively easy and intuitive to implement from a coding stand point. In addition, the method allows retrieval of images that have been transformed in terms of their size as well as translated through rotations and flips and compression using SPIHT. During testing, the basic algorithm produced some good results in that it was able to retrieve many relevant images from the large database, all these techniques that used in this paper and other like genetic algorithm and an ant colony optimization plan to use in indexing and retrieval and search in cloud environment.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] Bai Xue, Liu Wanjun, Research of Image Retrieval Based on Color, International Forum on Computer Science-Technology and Applications, 2009.

[2] Ch.Srinivasa Rao1 , S.Srinivas Kumar2 and B.Chandra Mohan3 , Content based image retrieval using exact Legendre moments and support vector machine , The International Journal of Multimedia &Its Applications (IJMA),Vol.2,No.2,May 2010.

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

[4] Henning Muller, Nicolas Michoux, David Bandon and Antoine Geissbuhler , A Review of Content Based Image Retrieval Systems in Medical Applications Clinical Benefits and Future Directions, Division for Medical Informatics, University Hospital of Geneva.

[5] Reshma Chaudhari, A. M. Patil , Content Based Image Retrieval Using Color and Shape Features, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 2012.

[6] Mark S. Nixon and Alberto S. Aguado , Feature Extraction and Image Processing, British Library Cataloguing in Publication Data ,2002 first edition.

[7] Xiuqi Li ,Shu Ching Chen Mei Ling Shyu Borko Furht , an effective content based visual image retrieval system , Department of Electrical and computer Engineering ,University of Miami ,2002.

[8] Sridhar R. Avula ,Jinshan Tang ,Scott T. Acton , Image retrieval using segmentation , Proceedings of the 2003 Systems and Information Engineering Design Symposium ,2003.

[9] Mohammed Hamzah Abed, C.Namrata Mehandir, SAR image Compression Using SPIHT Algorithm, international journal of computer Applications, 2011.

[10] Jianlin Zhang , Wensheng Zou , Content-Based Image Retrieval Using Color and Edge Direction Features , Institute of Computer Technology Engineering Nanchang University.

[11] Liangbin Zhang , Lifeng Xi ,Bishui Zhou , Image retrieval method based on entropy and fractal coding, WSEAS TRANSACTIONS on SYSTEMS ,2008.

[12] N.S Vassilieva , St.Petersburg , Content-based image retrieval methods , Programming and Computing Software,2009.