

A New Method for Content based Image Retrieval using Primitive Features

Dr.S.Maruthuperumal¹

¹Professor & HOD CSE, Godavari Institute of Engineering & Technology, Rajahmundry, A.P, India.
Email: maruthumail@gmail.com

G. Rosline Nesa Kumari²

²Associate Professor, IT, Godavari Institute of Engineering & Technology, Rajahmundry, A.P, India.
Email: rosemaruthu@gmail.com

Abstract— The diminishing expenditure of consumer electronic devices such as digital cameras and digital camcorders along with ease of transportation facilitated by the internet, has lead to a phenomenal rise in the quantity of multimedia data. The need to find a desired image from a collection is shared by many professional groups, including journalists, design engineers and art historians. While the requirements of image users, it can be characterize image queries into three levels. The proposed method based on primitive features such as color and shapes. These features are extracted and used as the basis for a similarity check between images. The shape and color features are extracted through Gradient Edge Detection and color histogram the combination of these features is robust. The experiment results show that the proposed image retrieval is more efficient and effective in retrieving the user interested images.

Index Terms— Image retrieval, Color, Shape, Gradient Edge

I. INTRODUCTION

In the last little years, the fast growth of the Internet has extremely amplified the digit of image collections obtainable. The accretion of these image collections (including art works, satellite and medical imagery) is attracting extra users in a variety of specialized field [2] — for example natural features, medication, structural design, publicity, blueprint, style and publishing. in the intervening time, the study of image retrieval, which is worried with successfully and resourcefully accessing preferred images from large and varied image collections, has become more motivating and more demanding [1].

Image retrieval is concerned with techniques for storing and retrieving images both professionally and successfully [2]. In the early hours image retrieval methods establish the preferred images by corresponding keywords that are assigned to every image physically. On the other hand, as a consequence of the huge digit of images in collections, physical processing has become unrealistic. As well, since we are unlikely to foresee all the query keywords that will be used in a retrieval process, it is impractical to assign keywords to every

image, so the effectiveness of classic image retrieval is very limited.

Robust Content-based image retrieval, which is based on mechanically extracted primitive features such as color, shape, texture, and even the spatial relationships amid objects, has been employed since the 1990's [3]. In the last ten years, a huge contract of research work on image retrieval has concentrated on CBIR technology. This technology still lack development, and is not yet being used on a significant scale. The concepts which are presently used for CBIR system are all under research [4]. The design of CBIR systems involves research on data bases and image processing, handling problems that vary from storage issues to friendly user interfaces. Images are particularly complex to manage –besides the volume they occupy; retrieval is an application-and-context-dependent mission [5].

A similarity measure is to compare two images and it is a matching function. Shape is an important characteristic to identify and distinguish objects [8, 9]. Shape descriptors are classified in to boundary-based (or contour-based) and region-based methods [9]. A lot of research work has been carried out on image retrieval by many researchers, expanding in both depth and breath [11, 12, 7, 10, 13, 14]. The term content based image retrieval seems to have originated with the work of kato [15] for the automatic retrieval of the images from a data base, based on the color and shape present. Since then, the term has widely been used to describe the process of retrieving desired images from a collection database, on the basis of syntactical image features (color, texture and shape). The extraction of color features from digital images depends on an understanding of the theory of color and the representation of color in digital image. Digital imaging has resulted in an exponential increase in the volume of digital images. The need to find a desired image from a collection of databases has wide applications, such as in crime prevention by automatic face detection, finger print, medical diagnosis, to name a few shape and texture using elastic energy-based approach to measure image similarity has been presented in [16].

In this piece of writing, we widely survey, examine, and compute current progress and expectations prospects

of image retrieval. The proposed image retrieval method is purely on the basis of shape and color. The distances between the characteristic vectors of the query image and those of the images in the database are then calculated and retrieval is performed with the help of a logic scheme. The logic scheme provides a well-organized way to search for the image database. Modern retrieval systems have included user's significance response to transform the retrieval process in order to produce perceptually and semantically more consequential retrieval results. In this paper, we introduce these elementary techniques for content-based image retrieval. The rest of this article is arranged as follows: System overview of the proposed method is discussed in Section 2. The techniques of the current decade are presented in detail, in Section 3. The experiment results are discussed in section 4. We conclude in Section 5.

II. SYSTEM OVERVIEW

2.1 Gradient Edge Detection & Color Histogram

Shape may be defined as characteristic surface configuration of an object and a contour (outline). It permits an object to be distinguished from its surrounding by its outline. The proposed method based on Boundary shape representation, it only uses the outer boundary of the shape by using Gradient edge detection. This is done by describing the considered region using its external characteristics, i.e., and the pixels along the object boundary.

- The main idea of Gradient is to use the boundary as the shape features.
- The main idea of moment invariants is to use region based moments which are invariant to transformation as the shape feature.

Color is one of the most widely used features in image retrieval. It is robust to background complication and invariant of image size and orientation. In color based image retrieval, histogram is the most commonly used representation for color features. Statistically, it utilizes a property that images having similar contents should have similar color distributions.

III. PROPOSED METHOD

The proposed method involves the research of Gradient and Color Histogram is used for implementation of CBIR using color and shape descriptor method. The main proposed method of CBIR arises from the following two points.

- Extraction of shape and color features
- Retrieval of similar image from image database

The data flow diagram of the proposed method is shown in Figure 1.

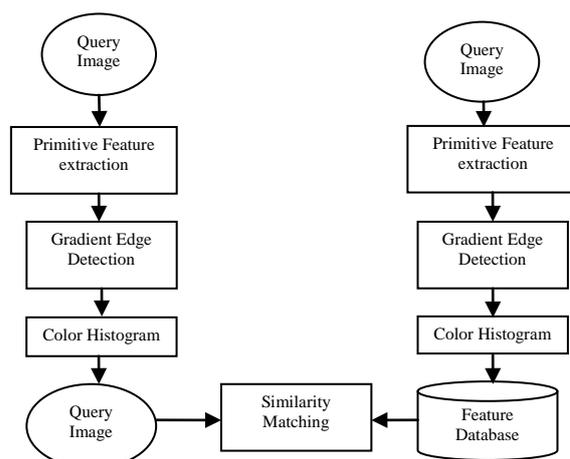


Fig 1: Data flow diagram for the proposed method

In the query RGB image is converted to get R, G, B components. The R component is used to compute sharp edges of the image using gradient approach. Figure 2 represents the conversion of the edge image from the query image.



Fig 2: Conversion of Edge Detection image

A color histogram H_y is a vector $(y_1, y_2 \dots y_n)$, in which each bucket y_j contains the number of pixels of color j in the image. Typically images are represented in the RGB color space, and a few of the most significant bits are used from each color channel. We use the 2 most significant bits of each color channel, for a total of $n = 64$ buckets in the histogram.

For a given image I , the color histogram H_{y1} is a compact summary of the image. A database of images can be queried to find the most similar image to I , and can return the image I' with the most similar color histogram H_{y1} . Typically color histograms are compared using the sum of squared differences (L_2 -distance) or the sum of absolute value of differences (L_1 -distance). So the most similar image to I would be the image I' minimizing the L_2 -distance or L_1 -distance. Note that we are assuming that weighted evenly across different color buckets for simplicity.

The histogram array looks like:

$$H_y (h_1, h_2 \dots h_n)$$

$$h_j = \frac{\text{the number of pixels of color } j \text{ in the image}}{\text{total number of pixels in the image}}$$

We divide the histogram array by the total number of pixels in the image in order to normalize the values of the histogram regardless of the image sizes.

Distance formula:

$$|H_I - H_{I'}| = \sum_{j=1}^n |H_I[j] - H_{I'}[j]|$$

- $H(h_1, h_2, \dots, h_n)$ - a vector, in which each component h_j is the number of pixels of color j in the image
- n - number of distinct(discretized) color
- I - an image
- H_I - the color histogram of image I
- I^1 - histogram of second image

Algorithm

Step1: Consider the query image and get R, G, B components of the image.

Step2: Compute gradient map (R_y) of the image using R component.

Step3: Use 1 vector to describe the query image.

Step4: First two most significant bits of R_y , G and B values from each pixels of image is taken. Combine these values bitwise and we can form a 6-bit number from this.

Step5: Create an array called histogram of size 64. Increment the array at the index of the number that was calculated from step 4.

Step6: Do 4, 5 for all the pixels in the image.

Step7: Extract the histogram for the database images and store them into vector.

Step8: Compare the feature vector of query image with the feature vectors of the database of images by using the distance formula.

Step9: If the images match exactly then the value for the similarity measure will be 0.

Step10: If the images vary slightly then the value may differ like 0.01 or 0.04.

Step11: If the images don't match means the value for the similarity measure will be 0.05.

Step12: The output will be displayed according to the threshold value that has been set by the administrator.

IV. EXPERIMENT RESULTS AND DISCUSSION

The proposed method is implemented using Mat lab and tested on MUT 1000 data base containing 1000 images of the different photos in JPEG format of size 256×256 and 512×512 as shown in Figure 3. The proposed method layout is simple and user-friendly. Figure 4 shows the selecting query image from proposed the layout.

The proposed method of assessing the effectiveness of the retrieval algorithm is highly dependent upon the determination of the 'expected relevant results'. The experiment results of the proposed method are as shown in Figure 5 for the query image 200.jpg from MUL 1000 DB. The process of defining these expected results is highly subjective and difficult as one must determine the visual similarity between almost 50 images in the database with the query. Figure 6 and Figure 7 shows the retrieved images from DB for query image 400.jpg and 600.jpg of the proposed method.

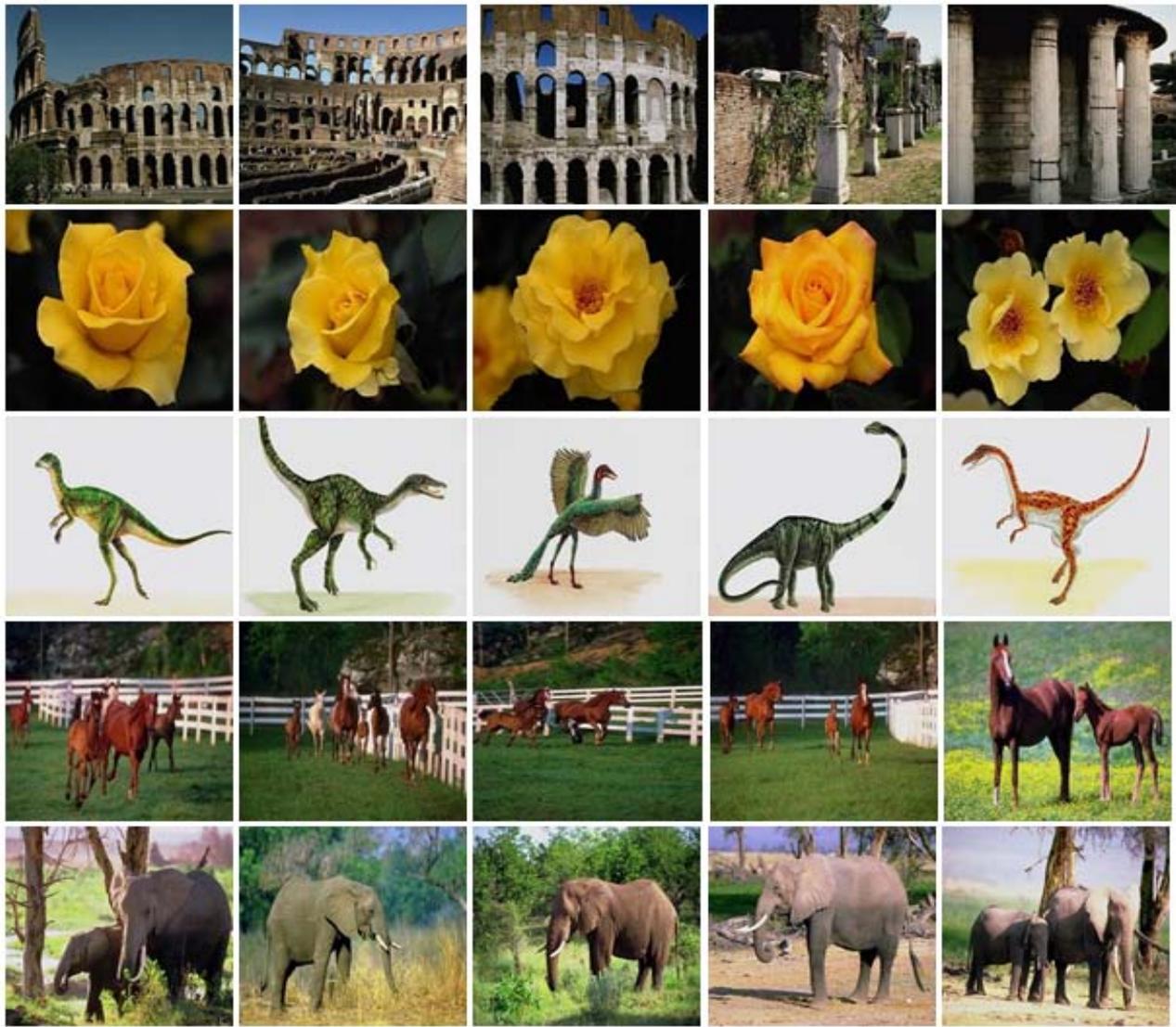


Fig 3: Sample database of 25 images

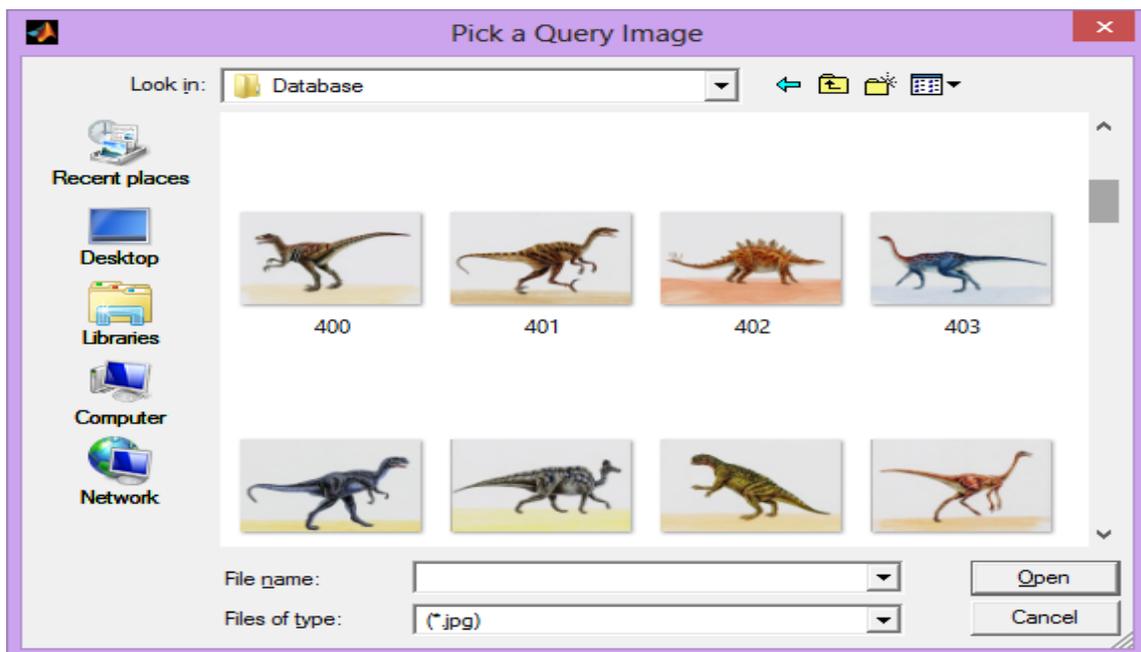


Fig 4: Selection of query image

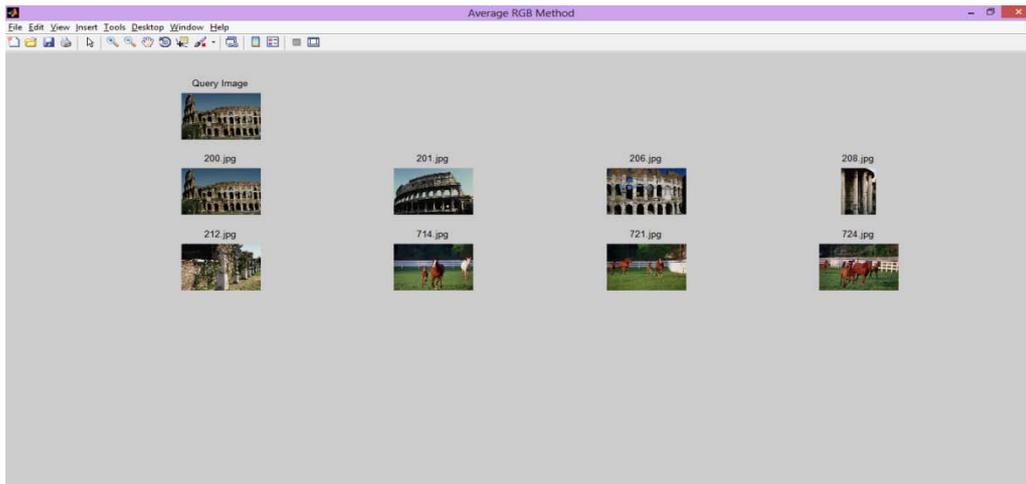


Fig 5: Retrieved images from MUL 1000 database for the query image 200.jpg

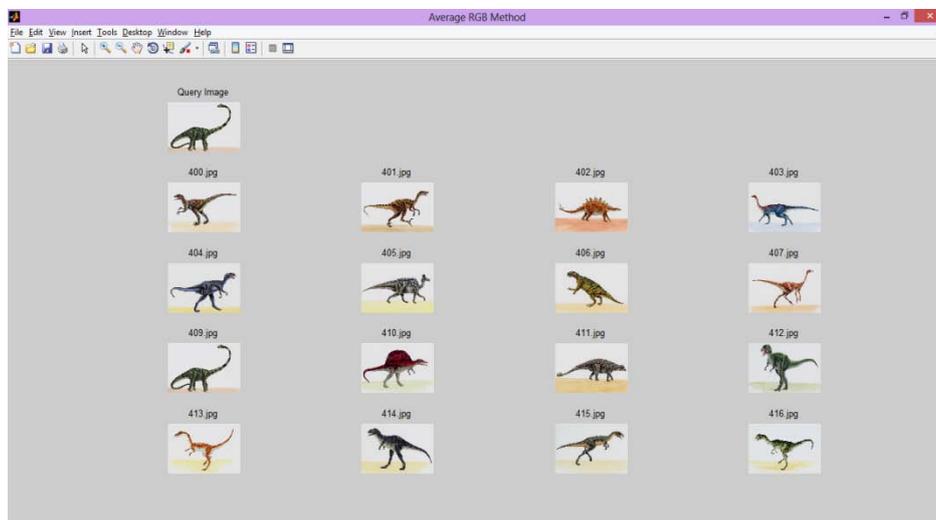


Fig 6: Retrieved images from MUL 1000 database for the query image 400.jpg

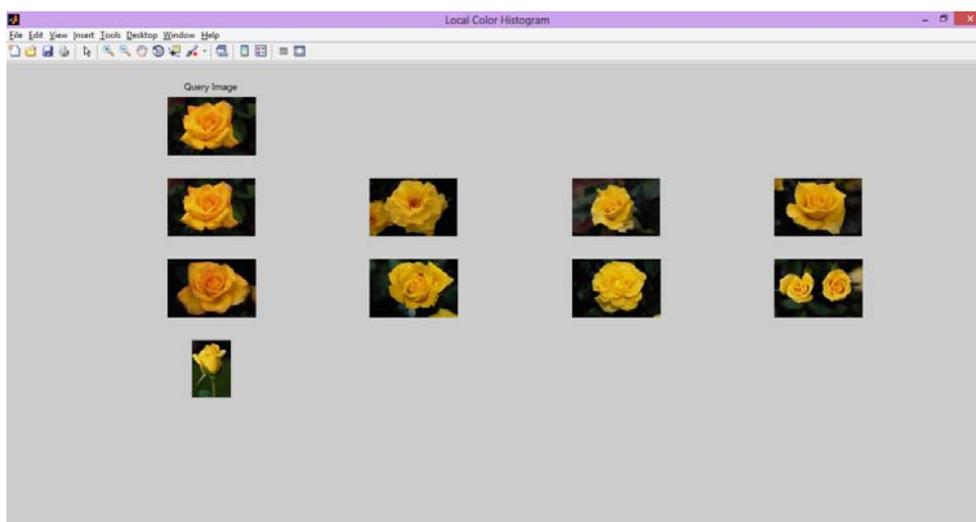


Fig 7: Retrieved images from MUL 1000 database for the query image 600.jpg

Table 1: Number of Relevant image in MUL 1000 DB

S.I No	Query Image	No. of Relevant image in MUL 1000 DB
1	200.jp g	51
2	400.jp g	59
3	600.jp g	29
4	800.jp g	35
5	100.jp g	41
6	300.jp g	42
7	500.jp g	23
8	700.jp g	34

V. CONCLUSION

The study's implementation of a shape and color histogram based image retrieval system identified numerous strengths in the algorithm performance as an image retrieval system. The proposed method integrates various features perfectly in Content based image retrieval system by using Gradient edge detection and color histogram and reflects the user subjective requirements. The experiments results achieve good performance and demonstrate the efficiency and robustness of system. During testing, the basic algorithm produced some good results in that it is able to retrieve many relevant images. The proposed method is particularly after other computerized online services are used to identify which images in the database are relevant. As a result, there is a substation increase in the retrieval speed. The whole indexing time for the 1000 image database takes 4-5 minutes.

REFERENCES

- [1] John Eakins and Margaret Graham, "Content-based Image Retrieval", JISC Technology Applications Programme, University of Northumbria at Newcastle, January 1999 .
- [2] Rui Y. & Huang T. S., Chang S. F, "Image retrieval: current techniques, directions, and open issues", Journal of Visual Communication and Image Representation, 10, 39-62.
- [3] Karin Kailing, Hans-Peter Kriegel and Stefan Schonauer, " Content- based Image Retrieval Using Multiple Representations". Proc. 8th Int. Conf. On Knowledge-Based Intelligent Information and Engineering Systems (KES'2004), Wellington, New Zealand, 2004, pp. 982-988. I
- [4] Thomas Seidl and Hans-Peter Kriegel, "Efficient User-Adaptable Similarity Search in Large Multimedia Databases," in Proceedings of the

- 23rd International Conference on Very Large Data Bases VLDB'97, Athens, Greece, August 1997, Found at: <http://www.vldb.org/conf/1997/P506.PDF>
- [5] Yuri, T.S. Huarq and S.F. Chang, "Image retrieval: current techniques, promising directions and open issues", Journal of Visual Communication and Image Representation, 10(4):39-62, 1999.
- [6] F.Mahmoudi , J.Shanbehzadeh,A..M.Eftekhari Moghadam (2003), "Image retrieval based on shape similarity by edge orientation autocorrelogram", Pattern Recognition 36 1725-1736.
- [7] Y.Ruri, T.S. Huang and S.F.Chang, "Image Retrieval: current Techniques, promising directions, and open issues", Journal of Communications and Image Representation, 10(1):39-62.March 1999.
- [8] S.Loncaric, "A Survey of Shape Analysis Techniques", Pattern Recognition, 31(8):983-1190 Aug 1998.
- [9] D.Zhing and G.Lu, " Review of Shape Representation and Description", Pattern Recognition 37(1):1-19,Jan 2004.
- [10] H.Nishida, "Structural feature indexing for retrieval of partially visible shapes", pattern Recognitions 35(2002)55-67.
- [11] R.Datta, D.Joshi, J.Li and J.Z.Wang, "Image retrieval: Ideas, influence, and trends of the new age", ACM computing Survey, 40, no.2, pp.1-60, 2008.
- [12] J.Eakins and M. Graham, "Content -Based Image Retrieval", Techniques report, JISC Technology Application Program, 1999.
- [13] A.M.Smeulders, M.Worrying and S.Santini, A.Gupta and R.Jain, "Content base image retrieval at the early year", IEEE Transaction on Pattern Analysis and Machine Intelligence, 22(12):pp.1349-1380,2000.
- [14] Y.Liu, D.Zang, G.Lu and W.Y.Ma, "A survey of content base image retrieval with high-level semantics", Pattern Recognition, Vol-40, pp-262-282, 2007.
- [15] T.Kato, "Database architecture for content-base image retrieval", In Proceeding of the SPIE-The International Society for Optical Engineering,vol.1662,pp.112-113,1992.
- [16] Huang, P., and Jean.Y, "Using 2d c+-strings are spatial knowledge representation for image data base system", 27, 1249-1257(1994).

Dr.S.MaruthuPerumal received his M.E. Computer Science and Engineering from Sathyabama University Chennai. He completed his Ph.D in the area of digital watermarking. He is having Fourteen years of teaching experience. At present he is working as an Professor and Head of the Department CSE Godavari Institute of Engineering and Technology, Rajahmundry. He published Twenty Eight research publications in various International, National Conferences and Journal. His

research interest includes Image processing, Digital Watermarking, Steganography and Security. He is a life member of ISCA, IAENG and Institutional Member of CSI.

G RoslineNesaKumari received her M.E. Degree from Sathyabama University Chennai in 2005. She is Pursuing her Ph.D degree in Computer Science and Engineering at Dr MGR University Chennai. She is having Twelve years of teaching experience. At present she is working as an Associate Professor, IT in Godavari Institute of Engineering and Technology, Rajahmundry. She published thirty two research publications in various International, National Conferences and Journal. She is a life member of Indian Science Congress Association (ISCA), IAENG, Red Cross. Her research interest includes Image processing, Digital Watermarking and Security.