

Well-Organized Content based Image Retrieval System in RGB Color Histogram, Tamura Texture and Gabor Feature

K. Haridas¹, Dr. Antony Selvadoss Thanamani²

Ph.D Scholar, Department of Computer Science , NGM College Pollachi¹

HOD, Department of Computer Science (Aided), NGM College Pollachi²

Abstract: Content Based Image Retrieval is a challenging method of capturing relevant images from a large storage space. A new low level feature contains histogram, color and texture information. This element is intended for use in image retrieval and image indexing systems. This paper experiments various methods available for Content based image retrieval System, they are RGB Color Histogram, Tamura Texture and Gabor Feature. The methods are implemented and tested based on three parameters like Precision value, Recall value and Accuracy rate. The Experimental results show that Gabor Feature method is more efficient when comparing with other methods. The Gabor Feature 81.7% Accuracy in Content Based Image Retrieval system.

Keywords: Histogram, Content Based Image Retrieval , Color, Tamura Texture, Gabor Feature and RGB.

I. INTRODUCTION

Content Based Image Retrieval (CBIR) has discussed early 1990s by the idea to find and retrieve images independent from metadata other than extracted from the image itself. However a satisfactory solution has not been found yet, but a problem has been isolated: Researchers defined the semantic gap, which refers to the inability of a machine to fully understand and interpret images based on automatically extracted data. In current research efforts in visual information retrieval especially global features, which denote features capturing characteristics of the whole image instead of focusing for instance on segments, regions or patches, have lost part of their significance. In applied research however Content based Image Retrieval – for instance as part of a complex system – is often relying on fast, global features at least as a foundation for further research. CBIR is an efficient library allowing researchers to integrate CBIR based on global features in an easy way. Content based Image Retrieval (CBIR) is an automatic character strings) has already been logically structured by the author Santini and Jain, [2]. CBIR draws many of its methods from the field of image processing and computer vision, and is regarded by some as a subset of that field. It differs from these fields principally through its emphasis on the retrieval of images with desired characteristics from a collection of significant size Research and development issues in CBIR cover a range of topics, many shared with mainstream image mining and information retrieval.

Some of the most important are:

- understanding image users' needs and information-seeking behavior
- identification of suitable ways of describing image content

process to search relevant images based on user input. The input could be parameters, sketches or example images. A typical CBIR process first extracts the image features and store them efficiently.

The earliest use of the term *content-based image retrieval* in the literature seems to have been by Kato [1] to describe his experiments into automatic retrieval of images from a database by color and shape feature. CBIR differs from classical information retrieval in that image databases are essentially unstructured, since digitized images consist purely of arrays of pixel intensities, with no inherent meaning. One of the key issues with any kind of image mining is the need to extract useful information from the raw data (such as recognizing the presence of particular shapes or textures) before any kind of reasoning about the image's contents is possible. Image databases thus differ fundamentally from text databases, where the raw material (words stored as ASCII

- extracting such features from raw images
- providing compact storage for large image databases
- matching query and stored images in a way that reflects human similarity judgments efficiently accessing stored images by content providing usable human interfaces to CBIR systems

II. LITERATURE REVIEW

Aura Conci and Castro [3] has proposed a novel approach based on calculating the distance between the images. **Chee Sun Won** [4] proposed an efficient use of MPEG-7 Color Layout and Edge Histogram Descriptors in

CBIR Systems. Another method for medical case retrieval has proposed by **Gw'anol'eQuellecet.al.** [7]. The method is evaluated on two classified databases: one for diabetic retinopathy follow-up (DRD) and one for screening mammography (DDSM). **H.B.Kekre et.al.**[8]proposed a Performance Comparison of Image Retrieval Techniques using Wavelet Pyramids of Walsh, Haar and Kekre Transforms. **Hemalatha and Devasena**[9]proposed a research to find out the accurate images while mining an image (multimedia) database and developed an innovative technique for mining images by means of LIM dependent image matching method with neural networks. **Hiremath P. S and JagadeeshPujari** [10] proposed Content based Image Retrieval based on Color, Texture and Shape features using Image and its complement. Color, texture and shape information have been the primitive image descriptors in Content based Image Retrieval systems. Efficient relevance feedback for Content based Image Retrieval by mining user navigation pattern is proposed by **ja-huangsu et al.** [11]. Image Retrieval by mining user navigation pattern can be divided into two major operations namely offline knowledge discovery and online image retrieval. **K.Haridas[13]et.al.** proposed image retrieval post-processing step by finding image similarity clustering to reduce the images retrieving using Color Feature Extraction and Texture Feature Extraction **LatikaPinjarkar et.al.** [14] Proposed a Comparative Evaluation of Image Retrieval Algorithms using Relevance Feedback and its Applications. The Feedback (RF) techniques were incorporated into CBIR. **Liu Yang et al.,** [15] have proposed research on Content based Image Retrieval in medical images like X-ray images collected from plain radiography. Another methodology is proposed using hierarchical and K-Means clustering technique by **Murthy et al.** [16]. **N S T Sai and R C Patil** [17] proposed an Image Retrieval using DWT with Row and Column Pixel Distributions of BMP Image. **Neetu Sharma. S et al.** [18] proposed an Efficient CBIR Using Color Histogram Processing. **Ramamurthy, B. and K.R. Chandran** [20] proposed a Content Based Medical Image Retrieval with Texture Content Using Gray Level Co-occurrence Matrix and K-Means Clustering Algorithms. **Rajshree S and Dubey** [21] illustrated about an Image mining methods which is dependent on the Color Histogram. They have examined a histogram-based search techniques and color texture techniques in two different color spaces, RGB and HSV. Histogram search distinguish an image through its color distribution. Another method [22] is discussed using the edge histogram **Rajendran and Madheswaran** [23] discussed an improved image mining technique. An enhanced image mining technique for brain tumor classification is proposed using pruned association rule with MARI algorithm. **Rajendran P and Madheswaran M** [24] proposed a Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm.

CONTENT BASED IMAGE RETRIEVAL

This paper experiments methodology Content-based retrieval, it uses the contents of images to represent and access the images. A typical content-based retrieval

system is divided into off-line feature extraction and online image retrieval. In off-line stage, the system automatically extracts visual attributes (color, shape, texture, and spatial information) of each image in the database based on its pixel values and stores them in a different database within the system called a feature database. The feature data (also known as image signature) for each of the visual attributes of each image is very much smaller in size compared to the image data, thus the feature database contains an abstraction (compact form) of the images in the image database. However, a more important reason for using the signature is to gain an improved correlation between image representation and visual semantics. In on-line image retrieval, the user can submit a query example to the retrieval system in search of desired images. The system represents this example with a feature vector. The distances (i.e., similarities) between the feature vectors of the query example and those of the media in the feature database are then computed and ranked. Retrieval is conducted by applying an indexing scheme to provide an efficient way of searching the image database. Finally, the system ranks the search results and then returns the results that are most similar to the query examples. If the user is not satisfied with the search results, he can provide relevance feedback to the retrieval system, which contains a mechanism to learn the user's information needs.

The proposed system implemented based on the following techniques.

K- Means algorithm Indexing by Latent semantic analysis Feature Extraction techniques

III. IMPLEMENTATION OF FEATURE EXTRACTION TECHNIQUES

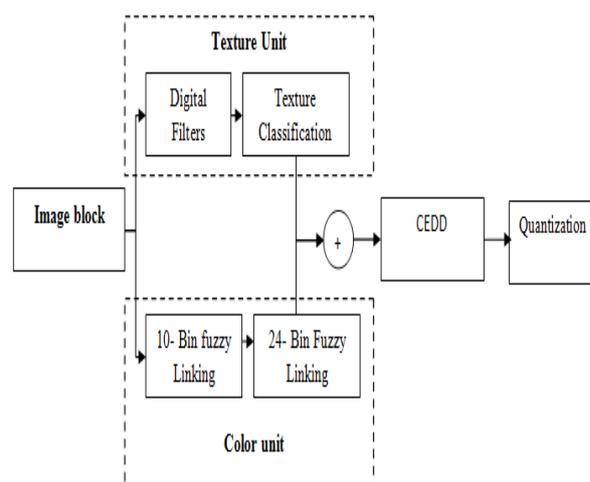


Figure- 1 Process of Color and Edge Directivity Descriptor

Several features are used in the Image Retrieval system. The popular amongst them are color features and texture features. The features which are used in this work are RGB color histogram, Tamura Feature, Gabor feature. The \Process of Color and Edge Directivity Descriptor is shown in the figure -1

IV. RGB COLOR HISTOGRAM

A color histogram represents the distribution of colors in an image, through a set of bins, where each histogram bin corresponds to a color in the quantized color space. A color histogram for a given image is represented by a vector:

$$H = \{H[0], H[1], H[2], \dots, H[n]\}$$

Where i is the color bin in the color histogram and $H[i]$ represents the number of pixels of color i in the image, and n is the total number of bins used in color histogram. Typically, each pixel in an image will be assigned to a bin of a color histogram. Accordingly in the color histogram of an image, the value of each bin gives the number of pixels that has the same corresponding color. In order to compare images of different sizes, color histograms should be normalized. The normalized color histogram H' is given as

$$H' = \left\{ \frac{H'[0]}{p}, \frac{H'[1]}{p}, \frac{H'[2]}{p}, \dots, \frac{H'[n]}{p} \right\}$$

Where $H'[i] = \frac{H[i]}{p}$, p is the total number of pixels of an image.

In this work image is given as an input, Figure 2 shows the input image (query image) Figure 3 shows the similar images of query image which are matched using RGB color histogram.



Figure 2 Query Image for RGB color histogram

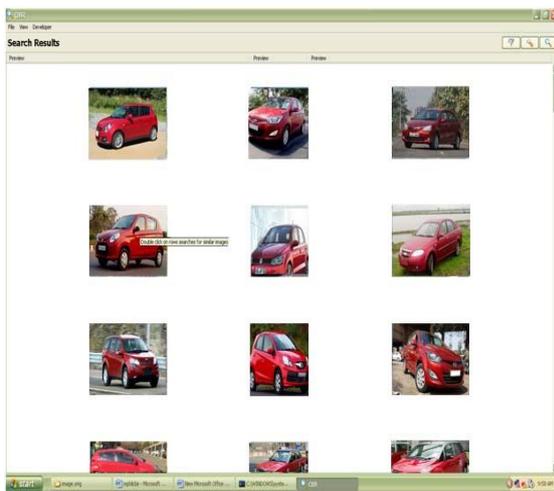


Fig 3 Matched Images retrieved Using RGB Color Histogram

TAMURA TEXTURE FEATURES

CBIR systems use in most cases the set of six visual features, namely, coarseness, contrast, directionality, line likeness, regularity, roughness. Figure - 4 shows the input image (query image) Figure-5 shows the similar images of query image which are matched using Tamura Texture Feature.

COARSENESS

Coarseness relates to distances of notable spatial variations of grey levels, that is, implicitly, to the size of the primitive elements forming the texture. The proposed computational procedure accounts for differences between the average signals for the non-overlapping windows of different size:

- At each pixel (x,y) , compute six averages for the windows of size $2^k \times 2^k$, $k=0,1,\dots,5$, around the pixel.
- At each pixel, compute absolute differences $E_k(x,y)$ between the pairs of non-overlapping averages in the horizontal and vertical directions.
- At each pixel, find the value of k that maximizes the difference $E_k(x,y)$ in either direction and set the best size $S_{best}(x,y)=2^k$.

Compute the coarseness feature (crs) F_{crs} by averaging $S_{best}(x,y)$ over the entire image. Instead of the average of $S_{best}(x,y)$, an improved coarseness feature to deal with textures having multiple coarseness properties is a histogram characterizing the whole distribution of the best sizes over the image.

A. Contrast

Contrast measures how grey levels q ; $q = 0, 1, \dots, q_{max}$, vary in the image g and to what extent their distribution is biased to black or white. The second-order and normalized fourth-order central moments of the grey level histogram (empirical probability distribution), that is, the variance, σ^2 , and kurtosis, μ_4 , are used to define the contrast:

$$F_{con} = \frac{\sigma}{\mu_4} w \text{ here } \sigma = \frac{\mu_4}{\sigma^4}$$

$$\sigma^2 = \sum_{q=0}^{q_{max}} (q - m)^2 \Pr(q|g)$$

$$\mu_4 = \sum_{q=0}^{q_{max}} (q - m)^4 \Pr(q|g)$$

m is the mean grey level, i.e. the first order moment of the grey level probability distribution. The value $n=0.25$ is recommended as the best for discriminating the textures.

B. Directionality

Degree of directionality is measured using the frequency distribution of oriented local edges against their directional angles. The edge strength $e(x,y)$ and the directional angle $a(x,y)$ are computed using the Sobel edge detector approximating the pixel-wise x - and y -derivatives of the image:



Figure -6 Query Image Gabor Texture Features

The deterministic periodicity of the image is analyzed using the autocorrelation function. The corresponding Wold feature set consists of the frequencies and the magnitudes of the harmonic spectral peaks (e.g., the K largest peaks). The indeterministic (random) components of the image are modeled with the multiresolution simultaneous autoregressive (MR-SAR) process. The retrieval uses matching of the harmonic peaks and the distances between the MRSAR parameters. The similarity measure involves a weighted ordering based on the confidence in the query pattern regularity. Experiments with some natural texture databases had shown that the Wold model provides perceptually better quality retrieval than the MR-SAR model or the Tamura's. An alternative to the spatial domain for computing the texture features is to use domains of specific transforms, such as the discrete Fourier transform (DFT), the discrete cosine transform (DCT), or the discrete wavelet transforms (DWT). Global power spectra computed from the DFT have not been effective in texture classification and retrieval, comparing to local features in small windows. At present, most promising for texture retrieval are multiresolution features obtained with orthogonal wavelet transforms or with Gabor filtering. The features describe spatial distributions of oriented edges in the image at multiple scales.

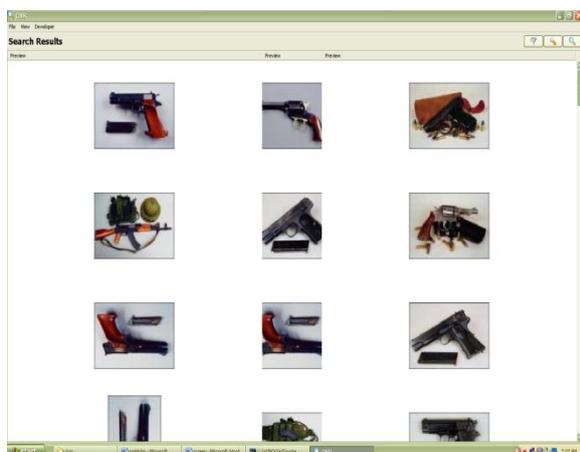


Figure 7 Matched images retrieved using Gabor Texture Features

A 2D Gabor function $\gamma(x,y)$ and its Fourier transform $\Gamma(u,v)$ are as follows:

$$\gamma(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\sqrt{-1}Wx \right]$$

$$\Gamma(u,v) = \exp \left[-\frac{1}{2} \left(\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right) \right]$$

where $\sigma_u = 1/2\pi\sigma_x$ and $\sigma_v = 1/2\pi\sigma_y$. The Gabor function is a product of an elliptical Gaussian and a complex-plane wave and it minimizes joint 2D uncertainty in both spatial and frequency domain. Appropriate dilations and rotations of this function yield a class of self-similar Gabor filters for orientation- and scale-tunable edge and line detection. The filters form a complete but non-orthogonal basis set for expanding an image and getting its localized spatial frequency description. The total number of Gabor filters is equal to the product of the numbers of scales and orientations. Figure 6 shows the input image (query image) Figure 7 shows the similar images of query image which are matched using Gabor Texture Feature.

VI. EXPERIMENTAL RESULTS

The image retrieval system is implemented using JAVA. The image mining techniques are tested on the corel image database of 800 images spread across 10 categories of Flowers, Buildings, Betas, Guns, Speedway, Elephants, Dinosaurs and Cars. Each class contains 100 images; Figure 8 shows the collection of sample images.

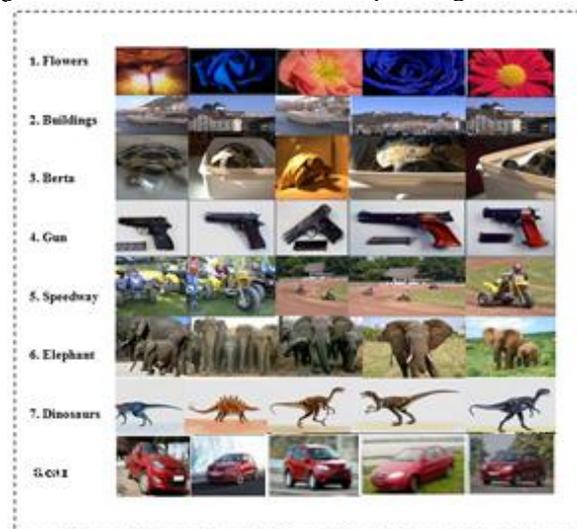


Figure 8 Some Sample Images from the Database Used for Experimentation

VII. PERFORMANCE EVALUATION METRICS OF IMAGE MINING USING CBIR

To assess the retrieval effectiveness, I have used the precision and recall as statistical comparison parameters for the proposed image mining techniques. The standard definitions of these two measures are given by following equations. The level of retrieval accuracy achieved by a system is important to establish its performance. If the outcome is satisfactory and promising, it can be used as a

standard in future research works. In image mining using CBIR, precision-recall is the most widely used measurement method to evaluate the retrieval accuracy. Precision, P, is defined as the ratio of the number of retrieved relevant images to the total number of retrieved images. Precision P measures the accuracy of the retrieval.

$$\text{Precision (P)} = \frac{\text{(Number of relevant images retrieved)}}{\text{(No of images retrieved)}}$$

Recall, R, is defined as the ratio of the number of retrieved relevant images to the total number of relevant images in the whole database.

$$\text{Recall (R)} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in database}}$$

The retrieval efficiency, namely recall precision and accuracy were calculated for color images from image database. Standard formulas have been used to compute accuracy rate.

$$\text{Accuracy Rate} = \frac{\text{Precision} + \text{Recall}}{2}$$

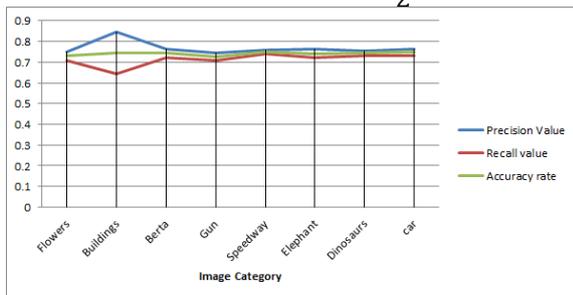


Figure -9 Precision, Recall and accuracy rate for RGB Color Histogram

Figure 9 represents Evaluation of precision, recall and accuracy value for RGB Color Histogram. An average accuracy rate for RGB Color Histogram is 0.741 or 74.1%. Precision, recall and accuracy value for images in corel database is calculated for RGB Color Histogram.

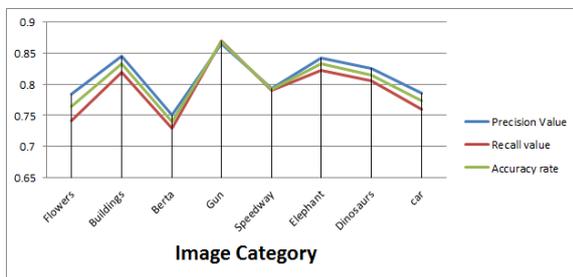


Figure - 10 Precision, Recall and accuracy rate for Tamura Texture Features

Figure 10 represents Evaluation of precision, recall and accuracy value for Tamura Texture Features. An average accuracy rate for Tamura Texture Features is 0.802 or 80%. Precision, recall and accuracy value for images in corel database is calculated for Tamura Texture Features.

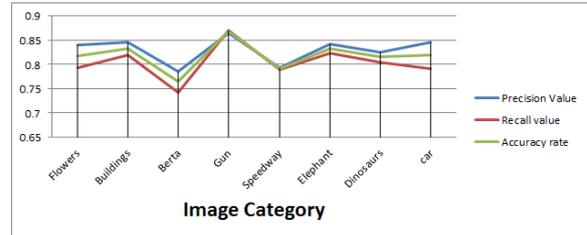


Figure 11 Precision, Recall and accuracy rate for Gabor Texture Features

Figure 11 represents Evaluation of precision, recall and accuracy value for Gabor Texture Features. An average accuracy rate for Gabor Texture Features is 0.817 or 81.7%. Precision, recall and accuracy value for images in corel database is calculated for Gabor Texture Features.

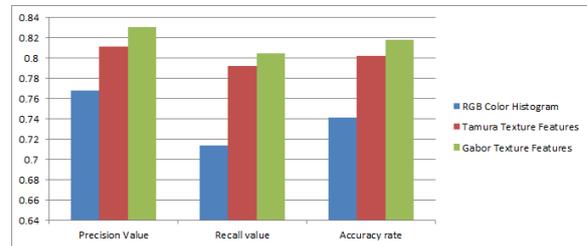


Figure 12 Average Precision, Recall and accuracy rate for various approaches

Figure 12 represents Evaluation of precision, recall and accuracy value for various methods. From this experimentation the Gabor Texture Features gives better Performance when comparing to the other methods experimented.

VIII. CONCLUSION AND FUTURE ENHANCEMENT

Content Based Image Retrieval is a challenging method of capturing relevant images from a large storage space. A new low level feature contains histogram, color and texture information. This element is intended for use in image retrieval and image indexing systems. This paper experiments various Content based image retrieval methods which is widely used. Three methods used for experimentation is RGB Color Histogram, Tamura Texture and Gabor Feature. From the experimentation Gabor Texture Feature method gives better when comparing with other methods used for experimentation. The Gabor Texture Features 81.7% Accuracy in content based image retrieval system. The Gabor Texture Features can be enhanced with the needs. This research work is useful for image searching, in future it is planned to connect semantic web-based image retrieval and facial recognition. It can be further developed to include more operations and analysis, as changes are required in the system to adapt to the external developments. Future enhancement can be made to the system at any later points. The codes are efficiently written to make it reusable and chargeable.

REFERENCES

- [1]. Hirata K and Kato T, "Query by visual example – content-based image retrieval" in EDBT'92, Third International Conference on Extending Database Technology, 56-71, 1992.
- [2]. H.B.Kekre, Dharendra Mishra, StutiNarula, Vidhi Shah , " Color Feature Extraction For Cbir", International Journal of Engineering Science and Technology (IJEST), ISSN : 0975-5462 Vol. 3 No.12, pp. 8357-8365, December 2011.
- [3]. Aura Conci and Everest Mathias M. M. Castro, "Image mining by content" ExpertSystems with Applications, vol. 23, no. 4, pp. 377-383, 2002.
- [4]. Chee Sun Won, , Dong Kwon Park, and Soo-Jun Park, "Efficient Use of MPEG-7 Edge Histogram Descriptor" , ETRI Journal, Vol. 24, No 1, pp. 23-30, February 2002.
- [5]. Datta R, Li J, and Wang J, "Content-based image retrieval - approaches and trends of the new age," ACM Computing Surveys, vol. 40, no. 2, Article 5, pp. 1-60 April 2008.
- [6]. Datcu M and Seidel K, "Image information mining: exploration of image content in large archives".IEEE Conference on Aerospace, Vol.3, pp. 253- 264, 2000.
- [7]. Gw'eno'le'Quellec, Mathieu Lamard, Guy Cazuguel, Christian Roux, B'eatriceCochener, "Multimodal medical case retrieval using the Dezert-SmarandacheTheory", Annual International Conference of the IEEE Engineering in Medicine andBiology Society inerm-00331184, version 1 – 15,Oct 2008.
- [8]. H.B.Kekre,Sudeep D. Thepade, AkshayMaloo, "Performance Comparison of Image Retrieval Techniques using Wavelet Pyramids of Walsh, Haar and Kekre Transforms" International Journal of Computer Applications (0975 – 8887)Volume 4, No.10, pp.1-8, August 2010.
- [9]. Hemalatha M. and Lakshmi Devasena C, "A Hybrid Image Mining Technique using LIMbased Data Mining Algorithm," International Journal of Computer Applications, vol. 25, no 2, pp. 1-5, 2011.
- [10]. HiremathP. S and JagadeeshPujari, "Content based Image Retrieval based on Color, Texture and Shape features using Image and its complement", International Journal of Computer Science and Security, Vol.1 : no 4,pp 25-35.
- [11]. Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu and Vincent S. Tseng, "Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns", IEEE Transactions on Knowledge and Data Engineering, Vol. 23, no. 3, pp. 360-372, March 2011.
- [12]. Jeffrey W. Seifert, "Data Mining: An Overview" Analyst in Information Science and Technology Policy Resources, Science, and Industry Division, CRS 1-16, 2004.
- [13]. Haridas, K., and Antony Selvadoss Thanamani. "An Efficient Image Clustering and Content Based Image Retrieval Using Fuzzy K Means Clustering Algorithm." International Review on Computers and Software (IRECOS) 9.1 (2014): 147-153.
- [14]. LatikaPinjarkar, Manisha Sharma and Kamal Mehta, "ComparativeEvaluation of Image retrieval Algorithmsuaing Relevance Feedback and its Applications "International Journal of Computer Applications (0975 – 888), Volume 48– No.18, pp. 12-16, June 2012.
- [15]. Liu Yang,Rahul Sukthankar, and Steven C.H. Hoi, "A Boosting Framework for Visuality-Preserving Distance Metric Learning and Its Application to Medical Image Retrieval", IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 32, no. 1, pp. 30-44, January 2010.
- [16]. Murthy V.S.V.S, Vamsidhar E, swarup Kumar J.N.V.R, Sankara Rao P, "Content based Image Retrieval using Hierarchical and K-Means Clustering Techniques", International Journal of Engineering Science and Technology Vol. 2(3), pp. 209-212, 2010.
- [17]. N S T Sai and R C Patil, "Image Retrieval using DWT with Row andColumn Pixel Distributions of BMP Image" International Journal on Computer Science and Engineering,Vol. 02, No. 08,pp. 2559-2566, 2010.
- [18]. Neetu Sharma. S1, PareshRawat . S2 and jaikaranSingh.S, "Efficient CBIR Using Color Histogram Processing", An International Journal(SIPIJ), Vol.2, No.1, pp. 94-112, March 2011
- [19]. Osmar R. Zaiane, "Principles of Knowledge Discovery in Database" Department of Computer Science, University of Alberta, pp 1-15, 1999.
- [20]. Ramamurthy, B. and K.R. Chandran, "Content Based Medical Image Retrieval with Texture ContentUsing Gray Level Co-occurrence Matrix and K-Means Clustering Algorithms"Journal of Computer Science 8 (7), ISSN 1549-3636, pp. 1070-1076, 2012.
- [21]. Rajshree S. Dubey , "Image Mining using Content based Image Retrieval System", IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, pp. 2353-2356, 2010.
- [22]. Rajshree S. Dubey, RajnishChoubey and Joy Bhattacharjee, "Multi Feature Content Based Image Retrieval" (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 06, pp. 2145-2149, 2010.
- [23]. Rajendran P and Madheswaran M, "An Improved Image Mining Technique For Brain Tumour Classification Using Efficient classifier", International Journal of Computer Science and Information Security, vol. 6, no. 3, pp. 107-116, 2009.
- [24]. Rajendran P and Madheswaran M, "Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm", Journal Of Computing, Vol. 2, Issue 1, Issn 2151-9617,pp. 127-136, January 2010.

BIOGRAPHIES



Mr K.Haridas is presently working as Assistant professor in the Department of Computer Applications, NGM College, Pollachi, Coimbatore (DT), India (affiliated to Bharathiar University, Coimbatore). He has presented papers and participated in Conferences, Seminars. He has published many papers

in international/national journals . His area of interest is Data Mining and Software Engineering. He has 14 years of Teaching Experience. He is a member of Computer Society of India.



Dr Antony SelvadossThanamani is presently working as professor and Head, Dept of Computer Science, NGM College, Coimbatore, India (affiliated to Bharathiar University, Coimbatore). He has published many papers in international/national journals and written many books. His areas of interest include

E-Learning, Software Engineering, Data Mining, Networking, Parallel and Distributed Computing. He has to his credit 26 years of teaching and research experience. His current research interests include Grid Computing, Cloud Computing, Semantic Web. He is a life member of Computer Society of India, Life member of Indian Society for Technical Education, Life member of Indian Science Congress, Life member of Computer Science, Teachers Association, New York and Member of Computer Science, Teachers Association, India.