Efficient Content Based Image Retrieval System in Visual Words, Color and Edge Directive Descriptors and Fuzzy Color and Texture Histogram

K. Haridas, Dr. Antony SelvadossThanamani

Ph.D Scholar, Department of Computer Science, NGM College Pollachi**,** India[1]

HOD, Department of Computer Science (Aided), NGM College Pollachi**,** India[2]

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 precision, recall and accuracy value for Bag of Visual words, Color and Edge Directive Descriptors, Fuzzy Color and Texture Histogram. The methods are implemented and tested based on three parameters like precision value, recall value and Accuracy rate. The   Experimental results show that FCTH (Fuzzy Color and Texture Histogram) method is more efficient when comparing with other methods. The FCTH shows 93.21 %   accuracy in Content Based Image Retrieval system.

KEYWORDS: Content Based Image Retrieval, Visual Words, Edge Directive Descriptors, Fuzzy Color and Texture Histogram.

I.   INTRODUCTION

Content Based Image Retrieval (CBIR) has discussed early 1990s by the idea to find and retrieve images independently 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 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 stores them efficiently.

The earliest use of the term C*ontent-Based Image Retrieval* in the literature seems to have been by H.B.Kekre [1] **et.al.** 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 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
➢ 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 [3]et.al. has proposed a novel approach based on calculating the distance between the images. **Chee Sun Won[5] et.al**. 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´enol´eQuellec[8] et.al.** The method is evaluated on two classified databases: one for diabetic retinopathy follow-up (DRD) and one for screening mammography (DDSM). **H.B.Kekre[9] et.al.** Proposed a Performance Comparison of Image Retrieval Techniques using Wavelet Pyramids of Walsh, Haar and Kekre Transforms.**H.B.Kekre[10] et.al.** Proposed a Content Based Image Retrieval is the application of computer vision techniques to the image retrieval problem of searching for digital images in large databases. **Hemalatha[13]et.al.** 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[14] et.al.** 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-Hwung Su[15] et.al.** Image Retrieval by mining user navigation pattern can be divided into two major operations namely offline knowledge discovery and online image retrieval. **LatikaPinjarkar[17] et.al**. Proposed a Comparative Evaluation of Image Retrieval Algorithms using Relevance Feedback and its Applications. The Feedback (RF) techniques were incorporated into CBIR. **Liu Yang[18] et.al.,** 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[19] et.al. K.Haridas[20]et.al.** proposed image retrieval post-processing step by finding image similarity clustering to reduce the images retrieving space by introducing a graph-theoretic approach using Color Feature Extraction and Texture Feature Extraction. **N S T Sai [21]et.al.** Proposed an Image Retrieval using DWT with Row andColumn Pixel Distributions of BMP Image.**Neetu Sharma. S[22] et.al.** Proposed an Efficient CBIR Using Color Histogram Processing. **Ramamurthy, B.[24] et.al.** Proposed a Content Based Medical Image Retrieval with Texture Content Using Gray Level Co-occurrence Matrix and K-Means Clustering Algorithms.**Rajshree S[25] et.al.** 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 is discussed using the edge histogram **Rajendran [27]et.al.** 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 [28]et.al.** Proposed a Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm.

III. CONTENT BASED IMAGE RETRIVAL

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. One advantage of a signature over the original pixel values is the significant content of image representation. 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

IV.   K- MEANS ALGORITHM

In this bag of visual words approach for content based image retrieval, local features are extracted from an image collection and clustered typically using k-means. The computed cluster centers are called visual words and form codebook, which serves as the basis for indexing newly added images. If a new image is added to the existing image collection, local features will have to be extracted and assigned to the best fitting visual word(s). The resulting local feature histogram representing the distribution of local features over the former computed clusters serves as a fingerprint for the new image. Fig 1. shows the similar images retrieved using Bag of Visual Words.
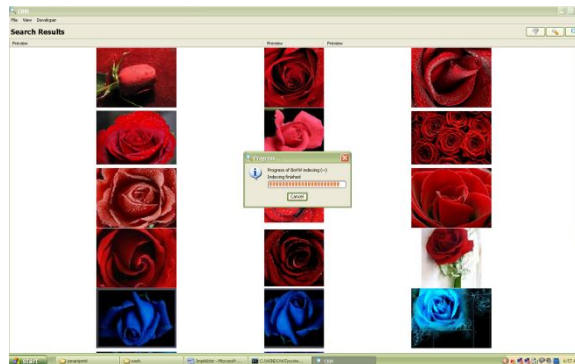

Fig.1 Similar Images Retrieved Using Bag of Visual Words

There are typically three steps are involved in features for image analysis. First, is a detection step which identifies interesting locations in the image usually according to some measure of saliency. These are termed interest points. Second, is to calculate a descriptor for each of the image patches centered at the detected locations. Third steps is codebook creation, all local features (one for each key point) are clustered. For each cluster a visual word is found. Typically a mean vector (in case of k-means) or a medoid is used as visual word. In this work K –means approach is employed.

V.   IMPLEMENTATION OF INDEXING BY LATENT SEMANTIC ANALYSIS

Latent Semantic Analysis is used for textual indexing and visual indexing. In this research work LSA is used for visual indexing.
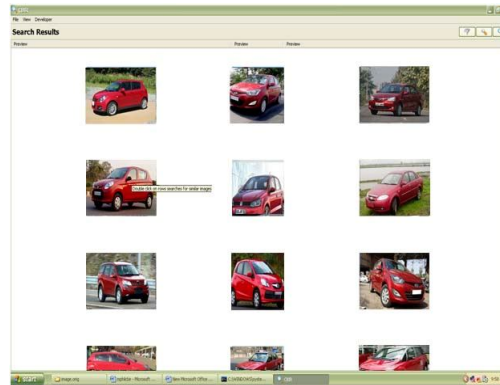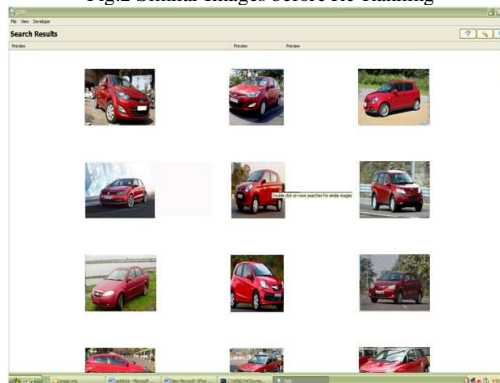
Fig.2 Similar Images before Re-Ranking

Fig.3 Results of Similar Images after Re-Ranking Using Latent Semantic Analysis

Fig.2 represents the similar images before Re-Ranking and Fig.3represents the results using LSA. Based on multimodal clues extracted from the initial text search results and any available auxiliary knowledge. The image Re-Ranking process improves search accuracy by reordering the visual documents

VI.   IMPLEMENTATION OF FEATURE EXTRACTION TECHNIQUES

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 auto color, Edge Directive Descriptor, Fuzzy Color and Texture Histogram. Fig.4 represents the process involved in Color and Edge Directivity Descriptor.
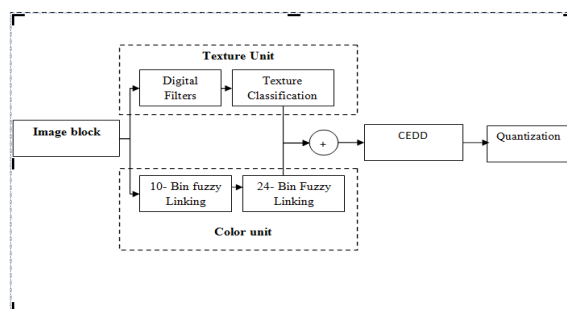

Fig. 4 Process of Color and Edge Directivity Descriptor.

VII.   IMPLEMENTATION OF COLOR AND EDGE DIRECTIVITY DESCRIPTOR (CEDD)

In this work image is given as an input, Fig.5 shows the input image (query image)for CEDD and Fig.6 shows the similar images of query image which are matched using CEDD.

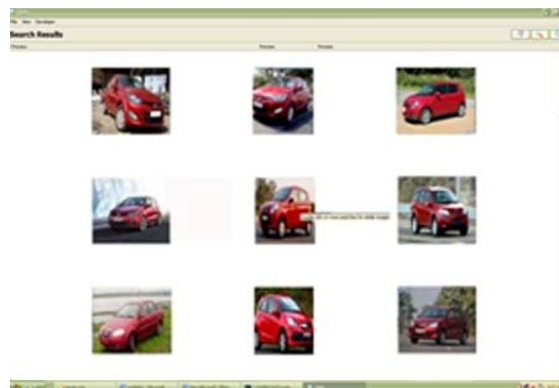Fig. 5 Query Image for Color and Edge Directivity Descriptor

Fig.6 Matched Images retrieved Using Color and Edge Directivity Descriptor

VIII.IMPLEMENTATION OF FUZZY COLOR AND TEXTURE HISTOGRAM (FCTH)

The histogram is constituted by 8 regions, as these are determined by the fuzzy system that takes decision with regards to the texture of the image. Each region is constituted by 24 individual regions, as these results from the second fuzzy system. Overall, the output that results includes $8 \times 24 = 192$ bins. Based on the content of the bins the respecting final histogram is produced. In order to shape the histogram, image is dispatched in 1600 blocks. This number of blocks was chosen as a compromise between the image detail and the computational demand. Each block passes successively from all the fuzzy systems. If the bin is defined that results from the fuzzy system of texture detection as N and as M the bin that results from fuzzy system that shapes fuzzy 24-bins color linking histogram, then each block is placed in the bin position: $N \times 24 + M$. Fig 7 illustrates the whole process of FCTH and Fig. 8 shows the Query Image of a car for Fuzzy Color and Texture Histogram.
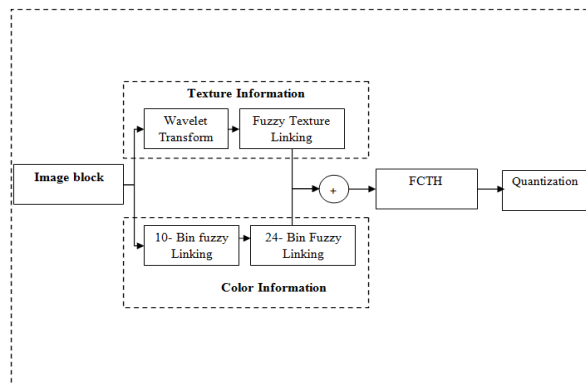
Fig. 7 Process of Fuzzy Color and Texture Histogram

Fig. 8 Query Image for Fuzzy Color and Texture Histogram

IX.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. Fig.9 shows the collection of sample images.
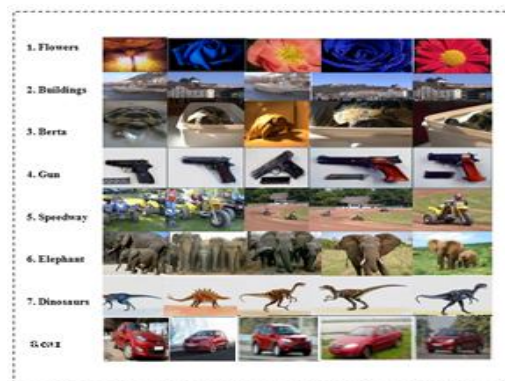

Fig.9 Some Sample Images from the Database Used for Experimentation

X. PERFORMANCE EVALUATION METRICS OF IMAGE MINING USING CBIR

To assess the retrieval effectiveness, the precision and recall are used 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.

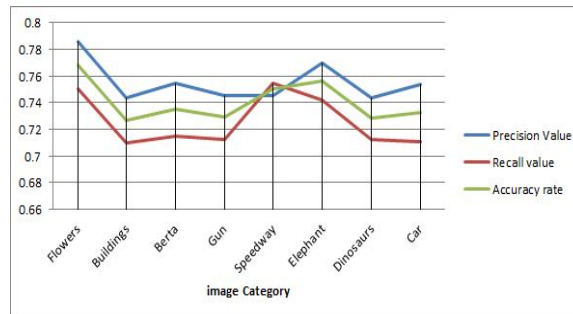$$\text{Accuracy Rate} = \frac{\text{Precision} + \text{Recall}}{2}$$

Fig.10 Precision, Recall and Accuracy Rate for Bag of Visual Words

Fig.10 represents Evaluation of precision, recall and accuracy value for Bag of Visual Words. An average accuracy rate for a Bag of Visual Words is 0.636 or 63.6%. Precision, recall and accuracy value for images in coral database is calculated for Bag of Visual Words.
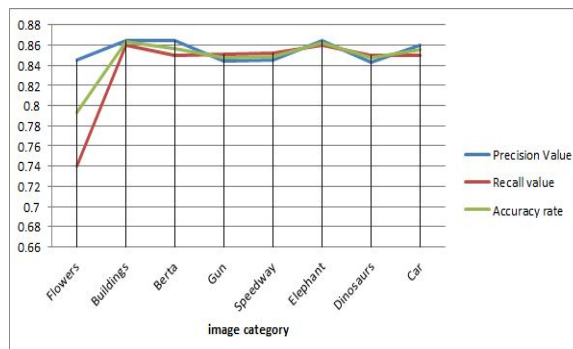

Fig.11 Precision, Recall and Accuracy Rate for CEDD

Fig.11 represents Evaluation of precision, recall and accuracy value for Color and Edge Directivity Descriptor. An average accuracy rate for Color and Edge Directivity Descriptor is 0.846 or 84.6%. Precision, recall and accuracy value for images in corel database is calculated for Color and Edge Directivity Descriptor.
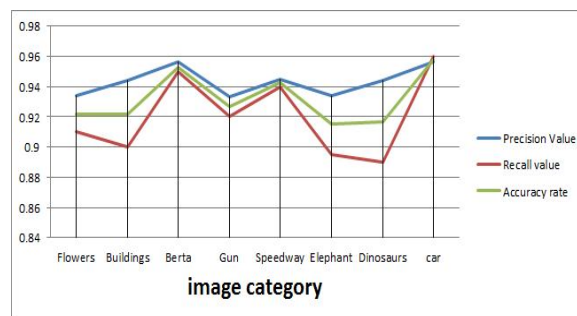

Fig.12 Precision, Recall and Accuracy Rate for FCTH

Fig.12 represents Evaluation of precision, recall and accuracy value for Fuzzy Color and Texture Histogram. An average accuracy rate for FCTH is 0.93 or 93%. Precision, recall and accuracy value for images in coral database is calculated for FCTH.
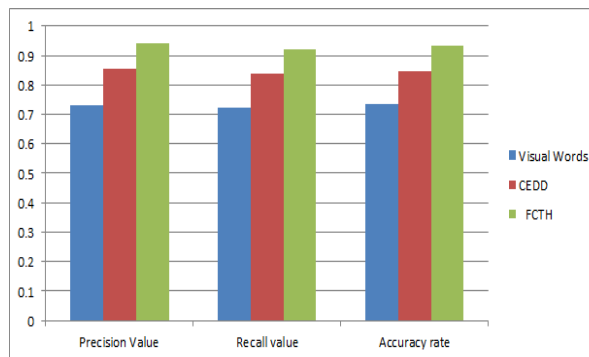

Fig.13 Average Precision, Recall and Accuracy Rate for various approaches

Fig.13 represents the average evaluation of precision, recall and accuracy value for various methods. From this experimentation the FCTH gives better performance when comparing to the other methods experimented.

X1. 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 value for Bag of Visual words, Color and Edge Directive Descriptors, Fuzzy Color with Texture Histogram. From the experimentation FCTH (Fuzzy Color and Texture Histogram) method gives better when comparing with other results than the other methods used for experimentation. The FCTH Shows 93.21 % Accuracy in Content Based Image Retrieval system. The FCTH 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 retrievaland 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. H.B.Kekre, Dhirendra 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.
2. Santini, S and Jain, R C *"The graphical specification of similarity queries"* Journal of Visual Languages and Computing, pp 403-421. 1997.
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. Chiang C, Hung Y, Yang H., and Lee G., *"Region-based image retrieval using color-size features of watershed regions,"* Journal of Visual Communication and Image Representation , vol. 20, no. 3, pp. 167-177, April 2009.
5. 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.
6. 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.
7. 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.
8. Gw´enol´eQuellec, 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 inserm-00331184, version 1 – 15,Oct 2008.
9. 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.
10. H.B.Kekre, Sudeep D. Thepade, AkshayMaloo, *"Query by Image Content using Color-Texture Features Extracted from Haar Wavelet Pyramid",*IJCA Special Issue on Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications,CASCT, pp 1-8, 2010.
11. H.B.Kekre, Dhirendra 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. Jeffrey W. Seifert, *"Data Mining: An Overview"* Analyst in Information Science and Technology Policy Resources, Science, and Industry Division, CRS 1-16, 2004.
17. LatikaPinjarkar, Manisha Sharma and Kamal Mehta, *"CompartiveEvalution of Image retrieval Algorithnsuaing Relevance Feedback and its Applications "*International Journal of Computer Applications (0975 – 888), Volume 48– No.18, pp. 12-16, June 2012.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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
23. Osmar R. Zaiane, *"Principles of Knowledge Discovery in Database"* Department of Computer Science, University of Alberta, pp 1-15, 1999.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.

BIOGRAPHY

Mr. K.Haridas is presently working as Asssistant 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.