Feature Selection for Image Retrieval based on Genetic Algorithm

Ms. Preeti Kushwaha[1], Mrs. R.R.Welekar[2]

[1] M. Tech scholar, Department of CSE, SRCOEM, Nagpur, Maharashtra, India
[2] Assistant Professor, Department of CSE, SRCOEM, Nagpur, Maharashtra, India

Abstract — **This paper describes the development and implementation of feature selection for content based image retrieval. We are working on CBIR system with new efficient technique. In this system, we use multi feature extraction such as colour, texture and shape. The three techniques are used for feature extraction such as colour moment, gray level co-occurrence matrix and edge histogram descriptor. To reduce curse of dimensionality and find best optimal features from feature set using feature selection based on genetic algorithm. These features are divided into similar image classes using clustering for fast retrieval and improve the execution time. Clustering technique is done by k-means algorithm. The experimental result shows feature selection using GA reduces the time for retrieval and also increases the retrieval precision, thus it gives better and faster results as compared to normal image retrieval system. The result also shows precision and recall of proposed approach compared to previous approach for each image class. The CBIR system is more efficient and better performs using feature selection based on Genetic Algorithm.**

Keywords — **CBIR, Feature Extraction, Feature Selection, Genetic Algorithm, K-means Clustering, Precision and Recall.**

I. Introduction

In recent years, continuous development of multimedia technologies, storing high quality of images and digital image collection data are rapidly increasing with social networking (such as facebook, yahoo, Google, etc), storage technologies, academic website, research and development etc to upload images, videos, many possible means. With the use of internet, image capturing devices (scanner, camera, etc) and handheld devices for tremendous collection of digital images is generated every day. The intent for development of many general use of image retrieval system requires efficient searching, browsing and retrieving tools.

In current technologies the acquirement, transmission, stack away, and handling are allow in large collection of database. With the increasing in use of network and multimedia development, users are not gratified with traditional image retrieval. So in recent year, CBIR has become an area of wide interest and source of fast retrieval and exact.

Content based image retrieval is image search technique; automatically extract features (colour, texture, shape, etc) to allow searching relevant images from large image database to given input query image based on similarity in features from query image compared with feature from database. The feature extraction techniques is commonly use in CBIR. There are three feature extractions such as colour feature extraction, texture feature extraction and shape feature extraction. Each feature extraction has several techniques. Colour feature extraction is

low level visual feature. The respective techniques are used in colour extraction such as colour moment, vector quantization, co-occurrence matrix, etc. Texture feature extraction is also low level visual feature. It measures look for visual patterns in images. It describes distribution of image intensity. In texture, many techniques are used like Gabor filter, Tamura features, etc. Shape feature extraction is high level visual feature. It describes surface of an object within images or particular region. It contains two methods, first is external boundary of shape; another is shape of whole region. The respective techniques are edge density (sobel, prewitt, canny,etc), moment invariant, etc.

In CBIR, the problem is which features are relevant in retrieval process that means large number of features is irrelevant. To avoid this problem, feature selection is used to reduce extraneous, excess and noisy data. The main objective of feature selection is find best feature from large feature set. To improve the performance of content based image retrieval system, feature selection which includes optimal features.

Genetic algorithm is an approach of feature selection for best optimal feature subset from large feature set. Genetic Algorithm is found on primal of evolution, natural selection and biology inheritance. Genetic algorithm is used to find optimal or best solutions to computational problem that minimizes or maximizes a particular function. Genetic algorithm works iteratively by using Genetic operators such as Selection, recombination and Mutation.The optimal feature is selected using Genetic algorithm that searches for the best feature subset corresponding to better image retrieval result.

Clustering is collection of articles which are similar between same clusters while dissimilar articles belong to other cluster. By using clustering in CBIR system is reducing the elapsed time of system. Different clustering techniques are used in CBIR such as k-means clustering, hierarchical clustering, SIFT,etc.

The rest of paper is organized as follow: in section 2 describes related work to content based image retrieval using various techniques. Section 3 presents methodology. In section 4 introduced on framework of proposed approach. The experimental result shows in section 5; and last section 6 concludes the paper.

II. Related Work

In content based image retrieval system, many techniques have been used extensively mentioned in various research papers for better performance. Lakshmi p.s et.al. [4] Performed retrieval of different input query images from the image database based on texture feature. Texture feature is extracted from image using gray level co-occurrence matrix (GLCM). They approached feature selection using genetic algorithm (GA) to improve the accuracy of content based image retrieval. The results of feature selection based on the performance measures (precision and recall) showed higher accuracy of the retrieval system can obtained in lesser computation time. P.K.Bhargavi et.al. [9] Contributes that content based image retrieval system based on the

relevant feature. They used color coherence vector and Gabor wavelets feature extraction technique. For Feature Discrimination, it used maximum entropy method for transforming numerical features with nominal using Class Attribute Interdependence Maximization (CAIM) algorithm. They also analyzed proposed approach by optimizing it with the feature selection using Particle Swarm optimization (PSO) algorithm for extracting the near relevant features. The result showed effectiveness and efficiency of the proposed model is compared with other models using precision and recall. C.V. Rashmi et.al. [7] Views that novel image retrieval using Ant Colony Optimization and Relevance Feedback. The proposed system, feature vector of the image is extracted by calculation of color correlogram, Gabor filter and edge histogram descriptors. In their model, feature selection using ACO technique to optimize the features for speed up retrieval and similarity computation. They used support vector machine (SVM) to improve efficiency of the system by using Relevance Feedback.

Clustering is used in CBIR system, Mit Patel et.al. [7] Describes collection of features or a dataset is divided into similar image classes using clustering and classification. The clustering is done with k-means clustering, and classification is done with fuzzy rule based classification. These algorithms are based on texture and color information. In their proposed model, the result showed accuracy is increases and retrieval time is decreases. They compared with proposed model and normal model.

III. METHODOLOGY

In this section, we are introducing methods for new proposed system. As described further the new proposed system is done with three efficient techniques such as

- Three Feature extractions techniques are used for colour, shape, and texture.
- Feature selection using Genetic Algorithm. In this method, we used new fitness function
- Clustering technique

The following new approaches describe below:

A. Feature Extraction

Feature extraction is most valuable operation of CBIR system. It translates the input data into set of features. In this section, we describe three feature extraction techniques which are used in our proposed CBIR system.

1) Color Feature Extraction

Color feature extraction represents different color model such as RGB (Red, Green, and Blue), HSV (Hue, Saturation, and value), CMYK (Cyan, Magenta, yellow), etc.

Color moment represents characterized a color image. There are 3 different color moments: first order is mean, second order is standard deviation, and third order is skewness of color; are extracted from RGB and HSI color spaces to form an 18-dimensional, using the following mathematical formulation:

$$\mu_i = \frac{1}{N} \sum_{j=1}^{N} p_{ij} \tag{1}$$

$$\sigma_i = \sqrt{\left(\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - \mu_i)^2\right)} \tag{2}$$

$$\theta_i = \sqrt[3]{\left(\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - \mu_i)^3\right)} \tag{3}$$

where, p_{ij} is an ith color channel at the jth image pixel.

2) Texture Feature Extraction

Texture feature extraction describes distribution of image intensities. For texture feature extraction a gray level co-occurrence matrix is simple and most extensively used approaches to extract texture feature from an original image. There are four components of GLCM that are used to characterize the texture such as entropy, contrast, energy and homogeneity. The following formulas are given below:

$$Energy = \sum_m \sum_n x^2(m,n) \tag{4}$$

$$Entropy = \sum_m \sum_n x(m,n) \log x(m,n) \tag{5}$$

$$Contrast = \sum_m (m-n)^2 x(m,n) \tag{6}$$

$$Homogenity = \sum_m \sum_n \frac{x(m,n)}{1+|m-n|} \tag{7}$$

where, (m,n) represents number of rows and number of columns of the image x.

3) Shape Feature Extraction

Shape describes surface of an object within images or particular region. Edge histogram represents 4 directional edges. The image is subdivided into 4 x 4 sub images i.e. 16 sub blocks. For each of the sub images, compute the histogram by using 4 edge types: vertical, horizontal, 45_o and 135_o.

1	-1	1	1	$\sqrt{2}$	0	0	$\sqrt{2}$
1	-1	-1	-1	0	$-\sqrt{2}$	$-\sqrt{2}$	0

B. Feature Selection using Genetic Algorithm

Genetic algorithm is compute to find solutions to search and optimization problems. Genetic algorithm is used to find optimal or best solutions to computational problem that minimizes or maximizes a particular function. They simulate biological process of natural selection and reproduction to solve for 'fittest' solutions. This is called 'survival of fittest' used for optimization problems. The basic components to Genetic algorithms are:

1. **Initial population of chromosomes:** Let m be the number of features. The size of population is N. To create random population P of N number of chromosomes is given below:

$$P = [C_1, C_2, \dots C_N]$$

2. **Fitness function:** Standard Deviation is used to evaluate the fitness of each individual population.

3. **Selection:** Select two parents from population according to their best fitness, which can generate new offspring. It assures that only the best fittest solutions made to generate offspring.

4. **Recombination or Crossover:** Recombinant the parents to form new offspring from two parents string, by copying selected bit of each parents.

5. **Mutation:** after the performance of crossover, mutate the new offspring from single parent. It reduces local optimum.

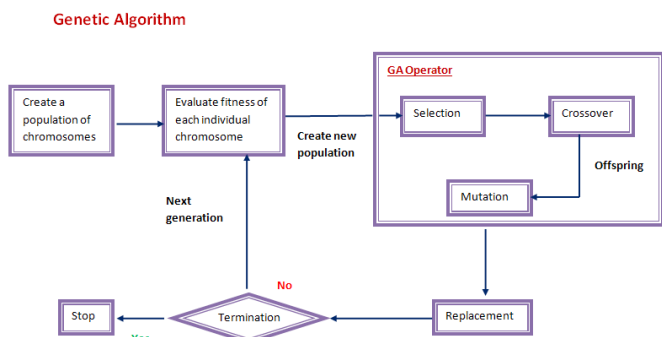The block diagram of genetic algorithm is shown in below:

Genetic Algorithm

Fig. 1. Block Diagram of Genetic Algorithm.

C. Clustering

Clustering is collection of articles which are similar between same clusters while dissimilar articles belong to other cluster. In proposed system, we use k-means algorithm. We select the k-means algorithm because it manages the large number of image in cluster. In CBIR system is reducing the elapsed time of system and fast retrieval. Using k-means algorithm, the results are measure by sum of among cluster between every vector and its centroid cluster. To calculate centroid of each cluster using sum of squared error, the given formula below:

$$SSE = \sum_{i=1}^{k} \sum_{x \in c_i} dist^2(m_i, x) \tag{8}$$

IV. Proposed System

Here, we propose new CBIR system using three approaches which are described in previous section. The proposed architecture of CBIR system is shown in below:
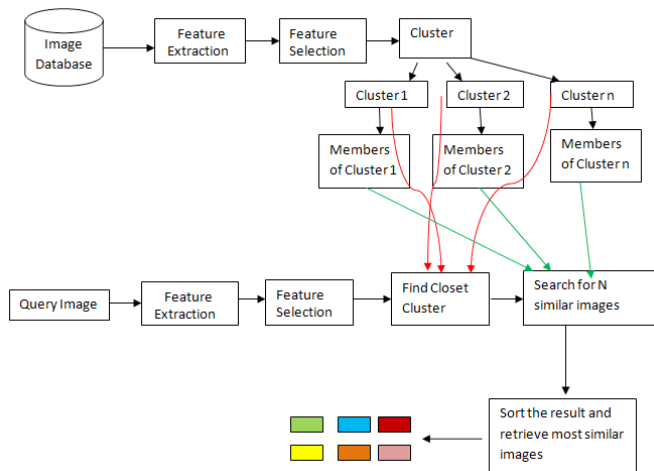

Fig. 2. Proposed Architecture of CBIR System.

The working of this system describes in step by step:
- The system will extract feature from images by using color, texture and shape automatically.
- These feature images are store in feature vector database.
- Then, we use feature selection using Genetic Algorithm that searches for the best feature subset from feature vector database.
- We use clustering which contains similar image classes. After that, system will compute distance between query image and centroid cluster to find smaller distance.
- The most similar images will retrieve and shows to the user.

V. Experimental Result and Discussion

In this section, we present a proposed CBIR system which is introduced in previous section. We are considering the new approaches which the enhancement is done. Here is first part of the expected result is present. We introduce database of image that we choose to test our system.

A. Image Database

We use image database in our evaluation is WANG database. It is subset of COREL database. It contains 1000 images in JPEG format. In this paper, 6 classes were taken and each class contains 15 images. The classes are Flowers, Bus, Architecture, Food, Elephant, and Dinosaurs.

B. Performance Measurements

In this section, we perform of the CBIR system can be evaluate in terms of precision and recall. And also compute computational time of each class.

1. **Precision:** It is defined as ratio of number of relevant images retrieved and total number of images retrieved

$$Precision = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved}$$

If precision value is 1.0 that means result images is retrieved by search was retrieved.

2. **Recall**: It is defined as ratio of number of relevant images retrieved and total number of images retrieved

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

C. Expected outcome

Here is expected outcome of each class in terms of precision, recall and computational time (s).
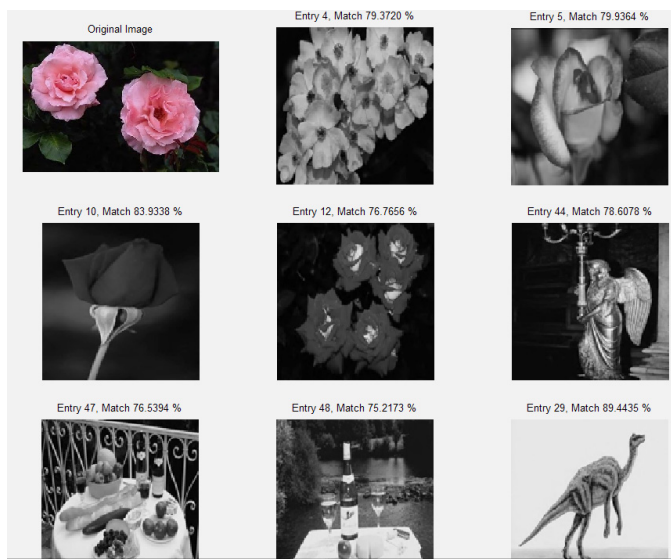

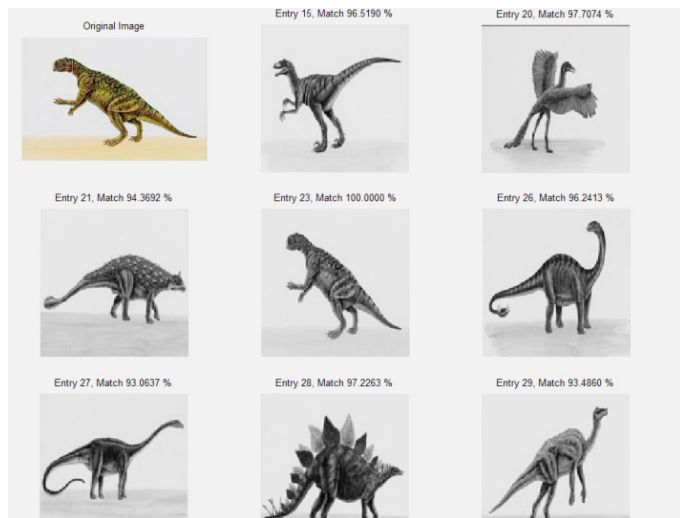Fig. 3. Without Feature Selection Retrieved Images.

Fig. 4. Feature Selection using Genetic Algorithm based Retrieved images.

In Fig.3 shows, relevant features and irrelevant features based retrieved images without using feature selection based on Genetic algorithm in CBIR system. In Fig.4 shows, most optimal relevant images using feature selection based on Genetic algorithm.

TABLE I

COMPARATIVE ANALYSIS OF CBIR ALGORITHMS BASED ON PARAMETERS

Images	Without Feature Selection			Feature selection using GA		
	Time(s)	Precision (%)	Recall (%)	Time(s)	Precision (%)	Recall (%)
Architecture	1.94	62	40	1.89	75.8	58
Flowers	1.94	48.2	22.4	1.89	54.12	25.9
Elephant	1.90	43.2	16.8	1.77	58.9	23
Buses	1.96	46.2	27.6	1.74	59.4	35.8
Food	1.90	50.6	30.2	1.85	57.2	50.89
Dinosaurs	1.99	85	82.4	1.94	100	93.36

Table 1 shows feature selection using Genetic Algorithm is compared with without Feature Selection in terms of precision, recall, and computational time. The computational time of feature selection using GA takes less than without feature selection. The performance measure analysis is done for each image class. The highest precision value in dinosaurs contain 85% without feature selection, and after using feature selection based on GA improve precision value contain 100% in dinosaurs image. Recall values contains in the range 22-82% without using feature selection and 23-94% using feature selection based on GA.
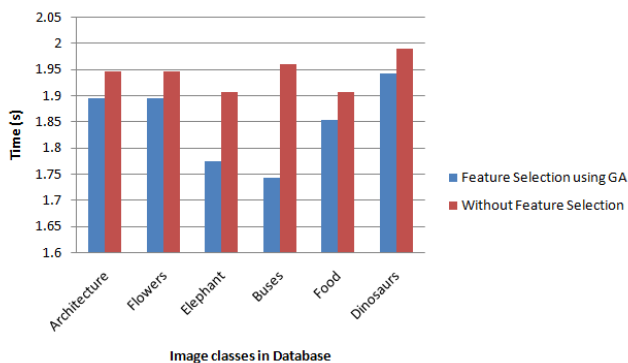

Fig. 5.a. Time (s) vs. Image classes in Database.

Figure 5.a shows computational time of each class in database. Blue indicates computational time of system by using feature selection based on GA. Red indicates computational time of system without feature selection. The bar chart shows, by using feature selection based on GA takes less time and fast retrieval. It is more efficient and better performance.
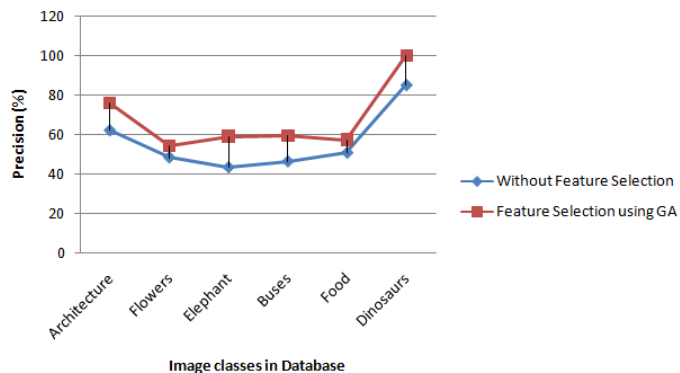

Fig. 5.b. Precision (%) vs. Image classes in Database.

Fig. 5.c. Recall (%) vs. Image classes in Database.

Figure 5.b and Figure 5.c shows, precision and recall of each image class in database. Blue indicates precision and recall of each image class without using feature selection. Red indicates precision and recall of each image class using feature selection based on Genetic Algorithm. The chart shows highest precision and recall of each image class using feature selection based on Genetic Algorithm is better than without feature selection.

TABLE II

PRECISION FOR PREVIOUS APPROACH AND PROPOSED APPROACH

Images	Previous Approach		Proposed Approach	
	Feature Extraction	Feature Selection	Feature Extraction	Feature Selection
	Precision (%)	Precision(%)	Precision (%)	Precision (%)
Bus	18.75	58.33	46.2	59.4
Dinosaurs	58.22	100	85	100
Elephant	41.66	51.66	43.2	58.9
Food	41.66	50	50.6	57.2
Average	**40.01**	**64.99**	**56.25**	**68.87**

TABLE III

Recall for Previous Approach and Proposed Approach

	Previous Approach		Proposed Approach	
	Feature Extraction	Feature Selection	Feature Extraction	Feature Selection
Images	Recall (%)	Recall(%)	Recall (%)	Recall (%)
Bus	12	28	27.6	35.8
Dinosaurs	28	48	82.4	93.36
Elephant	15.3	20	16.8	23
Food	20	23	30.2	50.89
Average	**18.825**	**32**	**39.25**	**50.76**

Table 2 and Table 3 are showing the precision and recall of Previous Approach and Proposed Approach for each Image class. The Previous approach contains Texture feature result based on feature extraction and feature selection using Genetic Algorithm; clustering is not used in this approach. The Proposed approach contains Multi Feature (Color, Texture, and Shape) result based on feature extraction and feature selection using Genetic Algorithm; clustering is used in this approach.
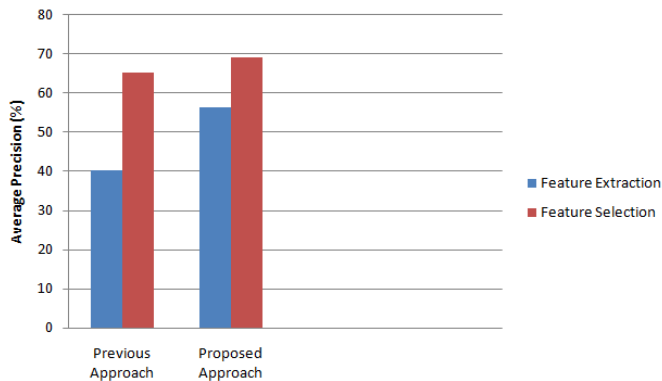

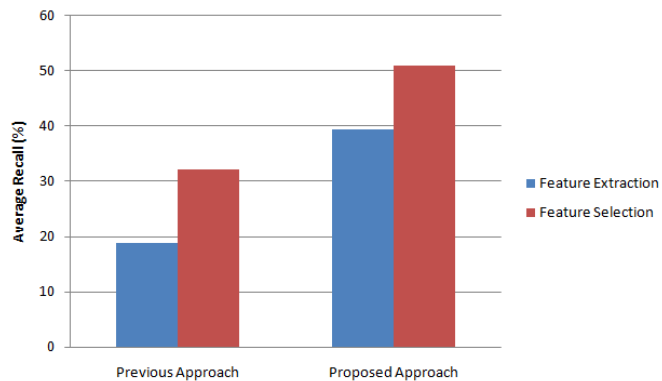Fig. 6. Average Precision (%) for Previous and Proposed Approach.

Fig. 7. Average Recall (%) for Previous and Proposed Approach.

In figure 6 and figure7 are showing Average precision and Average recall for previous and proposed approaches separately. The blue bar shows feature extraction of previous and proposed approaches; red bar shows feature extraction of previous and proposed approaches.

VI. Conclusion

In this paper, we introduce about the CBIR system with different techniques. Different Feature extraction techniques for colour, texture, and shape are used in CBIR system for better image retrieval. The

proposed work is CBIR system based on Feature selection using genetic algorithm for best optimal features from feature set, and also use clustering for reduce elapsed time of system. The experimental result shows feature selection using Genetic Algorithm reduces the time for retrieval and also increases the retrieval precision and recall, thus it gives better and faster results as compared to normal image retrieval system. The CBIR system is more efficient and better performs using feature selection based on Genetic Algorithm. The computational time is reduced. From the result, it is clear that feature selection using Genetic Algorithm is more optimize the searching time in seconds and also shows highest precision and recall of each image class as compared with normal image retrieval system i.e. without feature selection.

In future work, we try to use feature selection based on other optimization algorithm for scale down computational time and provide better accuracy result of CBIR system. It can give better performance of CBIR system. It is also possible to improve performance of retrieval system by relevance feedback.

References

[1] T. Hilda and Dr. R.R. Rajalaxmi, "Effective feature selection for supervised learning using Genetic Algorithm", *IEEE* sponsored 2'nd International conference of electronics and communication, 2015 ISSN 4788-7225.

[2] Keyuri M. Zinzuvadia, Prof. Keyur N.Brahmbhatt, Prof. Bhavesh A. Tanawala, "A survey on Feature Based Image Retrieval using classification and Relevance Feedback technique", International Journal of Innovative Research in Computer and Communication engineering (IJIRCCE), volume: 03, Issues: 01, 2015, pages: 508-513.

[3] Slimane Benloucif and Bachir Bouchehan,"Impact of feature selection on the performance of content- based image retrieval", *IEEE* transaction 2014, ISSN 4799-7508.

[4] Lakshmi P.S, Thusnavis Bella Mary, A.Vasuki, Kolcharam Yadhul, Ansa Saju,"Image retrieval based on feature selection method" International Journal of Research in Engineering and Technology (IJRET), volume: 03, Issues: 07, 2014, pages: 34-38.

[5] Kommineni Jenni, Satria Mandala, "Pre-processing Image Database for Efficient content Based Image Retrieval", International conference on Advances in Computing, Communications and Informatics, IEEE, Sept-2014, pages: 968-972.

[6] C.V. Rashmi, Sk.Sjida Praveen and G. Prathibha, "A Novel Image Retrieval System Using ACO and Relevance Feedback" International Journal of Advanced Research in Computer science and Software Engineering (IJARCSSE), volume: 4, Issues: 7, 2014, pages: 422-427.

[7] Mit Patel, Kanu Patel, Keyur Bhrahmbhatt, "Feature based Image retrieval based on clustering, classification techniques using low level image features", International Journal of Advance Engineering and Reasearch Development (IJAERD), volume: 1, issue:5, 2014, pages: 1-9.

[8] Ramandeep Kaur and Ashok Kumar Bathla, "Enhanced Content-Based Image Retrieval Using Cuckoo Search Algorithm", International Journal of Advanced Research in Computer science and Software Engineering (IJARCSSE),volume; 4, Issue: 9, 2014, pages: 233-241.

[9] P.K.Bhargavi, S.Bhuvana, and Dr.R.Radhakrishnan, "A Novel Content Based Image Retrieval Model Based on the Most Relevant Features Using Particle Swarm Optimization", Journal of Global Research in Computer Science (JGRCS), vol.: 4, 2013, pages: 25-30.

[10] Annesha Malakar, Joydeep Mukherjee, "Image Clustering using Color Moments, Histogram, Edge and K-means Clustering", International Journal of Science and Research(IJSR), vol.1, 2013, pages: 532-537.

[11] Afifi A.J., Ashour W.M., *"Content-Based Image Retrieval Using Invariant Color and Texture Features"*, IEEE Int. Conference on Digital Image Computing techniques and Applications, 2012, pp. 1-6.

[12]   Manimala Singha and K.Hemachandran, "Content Based Image Retrieval using Color and Texture", Signal & Image Processing: An International Journal (SIPIJ),   Vol.3, No.1, February 2012, pp. 234-237.

[13]   Jayant Mishra, Anubhav Sharma and Kapil Chaturvedi, "An Unsupervised Cluster-based Image Retrieval Algorithm using Relevance Feedback", International Journal of Managing Information Technology (IJMIT) Vol.3, No.2, May 2011, pages: 10-16.

[14] L.Ladha, T.Deepa, "Feature Selection Methods And Algorithms", International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 5 May 2011.

[15] Rajshree S. Dubey, Rajnish Choubey, Jor Bhattacharjee, "Multi feature Content Based Image Retrieval", International Journal on Computer Science and Engineering (IJCSE), volume: 2, No. 06, 2010, pages: 2145-2149.

[16]   Jianhua Wu, Zhaorong Wei, Youli Chang ,"Color and Texture Feature For Content based image retrieval", International Journal of Digital Content Technology and its Applications,2010.

[17]   D. Joshi, R. Datta, J. Li, and J.Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age", ACM Comput. Surv., April-2008.

[18] Tianzhong Zhao, Jianjiang Lu, Yafei Zhang1, Qi Xiao, "Feature Selection based on Genetic Algorithm for CBIR", Congress on Image and Signal processing, IEEE, volume: 02, May-2008, pages: 495-499.

[19]   Jonathan E. Rowe, "Genetic Algorithm theory", ACM GECCO '07: Proceedings of the 9th annual conference companion on Genetic and evolutionary computation, July-2007.

[20]   Ke-jia Chen, Zhi-Hua Zhou, Hong-Bin Dai, "Enhancing relevance feedback in image retrieval using unlabeled data", ACM trans. on Information Systems, April-2006.

Preeti Kushwaha, she is student of Master of Technology in Computer science engineering at Ramdeobaba college of Engineering and Management, Nagpur. She has received her Bachelor of Technology degree in Information Technology from the same college. Her area of interest is Image Processing and she published one survey paper in International conference named RICE 2016.

Rashmi R. Welekar, she is working as Assistant Professor at Ramdeobaba college of Engineering and Management, Nagpur. She has been teaching from last 10 years in the field of Computer science & engineering. She did M.Tech from Nagpur university. Her areas of interest are Image Processing, Networking, and Pattern recognition. She has about 10 research papers published in International journals, 6 International conferences, and 1 National conference.