FEATURE SELECTION FOR IMAGE RETRIEVAL BASED ON EVOLUTIONARY COMPUTATION

Preeti Kushwaha[1], Rashmi R. Welekar[2]

[1]*M.Tech Scholar, Department of CSE, SRCOEM, Nagpur, Maharashtra, India,*
1991preeti24@gmail.com
[2]*Assistant Professor, Department of CSE, SRCOEM, Nagpur, Maharashtra, India,*
rashmi.welekar@gmail.com

Abstract

In this paper we implemented feature selection for content based image retrieval using evolutionary computation. In this system, we used feature extraction techniques for color, texture and shape. The three techniques are used for feature extraction such as color moment, Gabor filter, and Edge histogram descriptor. To reduce the dimensionality and find best optimal features from feature set using feature selection based on two evolutionary computations i.e. Genetic algorithm, and Binary Bat Algorithm. These subset features are divided into similar image classes using k-means clustering algorithm for fast retrieval and improve the computational time. We compared these two algorithms with different parameters i.e. precision, recall and computational time of image retrieval. The experimental result shows feature selection using BBA reduces the time for retrieval and also increases the retrieval precision, thus it gives better and faster results as compared to feature selection using GA. In this method selects different combinations of features which user retrieves more similar images. The CBIR system is more efficient and better performs using feature selection based on BBA.

Keywords: *CBIR, Feature Extraction, Feature Selection, Genetic Algorithm, Binary Bat Algorithm, K-Means Clustering, Precision and Recall*

----------***----------

1. INTRODUCTION

Most of the applications have large number of features set that can be used. These features causes high dimensional set that can affect negatively performance of pattern recognition, image processing, data mining, and other applications. Feature selection is a process of selecting best features from large features set to improve the quality of data. The main advantages of feature selection are removing noisy data and irrelevant the data.

In image retrieval, once the features are extracted from image database; the problem is that which features are relevant in retrieval process. It comprises high dimensionality of data that may cause 'Curse of dimensionality' problem. Our paper gives a solution to this problem by using evolutionary computation for feature selection. The two algorithms are use feature selection, namely Genetic Algorithm, and Binary Bat Algorithm; in image retrieval system.

The content based image retrieval is used in this paper which is an application of computer vision to image retrieval problem. It is allow to users to find similar images from image database according to query image on the basis of similarity in features (color, texture, and shape) that means features from query image compared with features from image database.

Many systems attempt to compare the query image with images in database to find most similar images, resulting in basically linear search, which is inefficient high computationally when the image database is large. To overcome this problem, image clustering has been treated as to speed up image retrieval in large databases. We select k-means algorithm because it is fit to cluster the large amount of data.

The rest of this paper surveys on related work, discusses methodology, our proposed work, discusses experimental results, and last is conclude this paper.

2. RELATED WORK

Here feature selection using different techniques for any applications have been used extensively mentioned in various research papers for better performance. A.Sylvia Rani et.al. [1] describes the feature selection and performed Binary Bat Algorithm for unsupervised feature selection with k-means clustering algorithm. They used different real world datasets and shows different combinations of features are selected which have more accuracy. Also compare with other optimization algorithm. 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.

3. METHODOLOGY

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

1. Three Feature extractions techniques are used for colour, shape, and texture.
2. Feature selection using Genetic Algorithm.
3. Feature selection using Binary Bat Algorithm
4. Clustering technique

The following new approaches describe below:

3.1 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 for color, texture and shape which are used in our proposed CBIR system.

3.1.1 Colour Moment

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:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij} \quad (1)$$

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

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

3.1.2 Gabor Filter

Texture feature extraction describes distribution of image intensities. For texture feature extraction a Gabor filter is simple and most extensively used approaches to extract texture feature from an original image. This filter is most popular technique for texture feature extraction. It contains filtering in spatial and frequency domain. By using bank of Gabor filter to analyze the texture, has different scales and orientations allows multichannel filtering of an image.
Mathematical formula for Gabor filter in spatial domain:-

$$g(x,y;\lambda,\theta,\varphi,\sigma,\gamma) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right),$$

Where,

$$x' = x\cos\theta + y\sin\theta \quad (4)$$

$$y' = -x\sin\theta + y\cos\theta$$

3.1.3 Edge histogram Descriptor:

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 edge densities by using 4 edge types: vertical, horizontal, 45° and 135°.

1	-1	1	1	√2	0	0	√2
1	-1	-1	-1	0	-√2	-√2	0

3.2 Feature Selection using Evolutionary Computation

For image retrieval, to reduce the dimensionality and find best features from large feature set using feature selection based on two evolutionary computations i.e. Genetic Algorithm and Binary Bat Algorithm that searches optimal features corresponding to evaluation match percentage on ranking quality.

3.2.1 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.

3.2.2 Binary Bat Algorithm

Binary bat Algorithm is inspired algorithm developed by Yang, 2010. It is based on echolocation micro bats. BBA develops a discrete version of bat algorithm to solve feature selection problems and classifications. In BBA each artificial bats have a position, velocity and frequency vector. The position in BBA is either 0 or 1. The movement of bats is updating their velocity, position and frequency using the following equations:

$$V_i(t+1) = V_i(t) + (X_i(t) - Gbest)F_i \quad (5)$$

$$F_i = F_{min} + (F_{max} - F_{min})\,\beta \quad (6)$$

Where V_i, X_i, and F_i are the velocity, position and frequency of i^{th} bat. β represents a random value between 0 and 1. The position of the can be updated by sigmoid transfer function, is defined as:

$$S\left(V_i^j\right) = \frac{1}{1+e^{-V_i^j}} \quad (7)$$

$$X_i^j = \{1 \;\; if\; sigmoid\; function > \sigma \quad (8)$$
$$\{0 \;\; if\; sigmoid\; function < \sigma$$

If the sigmoid function is greater than σ, then position of bat is 1; if the sigmoid function is less than σ, then position of bat is 0. σ is random value between 0 and 1.

To reduce the loudness and increase the pulse rate in BBA can be updated as follows:

$$A_i(t+1) = aA_i(t) \quad (9)$$

$$r_i(t+1) = r_i(0)[1 - \exp(-\gamma t)] \quad (10)$$

Where α and γ are constant.

3.3 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:

$$SSE = \sum_{i=1}^{k} \sum_{x \epsilon c_i} dist^2(m_i, x) \quad (11)$$

4. PROPOSED WORK

Here, we are preluding our proposed method for CBIR system. The new proposed CBIR system is shown in below:

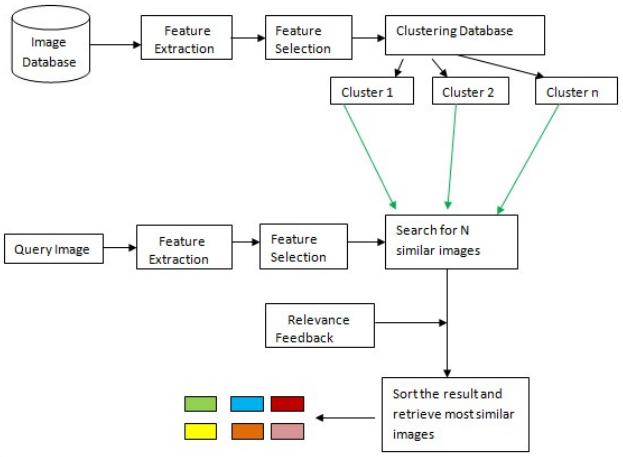

Fig.1 Proposed Architecture of CBIR System

The working of this system describes in step by step:

4.1 Preparing the features database

Step 1: Image database has color images in RGB color space.
Step 2: Calculate the color features (color moment), texture features (Gabor filter), and shape features (Edge histogram descriptor).
Step 3: These features are store in feature vector database.
Step 4: Select best feature**s** from feature database using two Evolutionary computations i.e. Genetic Algorithm, and Binary Bat Algorithm.
Step 5: Cluster the images in database using k-means Algorithm**.**

4.2 Querying

Step 1: User enter the query image in RGB color space
Step 2: Extract the features for the query image by using same techniques (color, texture and shape)
Step 3: Select best features from feature database using two evolutionary computations i.e. Genetic Algorithm, and Binary Bat Algorithm.
Step 4: Calculate the distance between the query image and the images in cluster database to find smallest distance
Step 5: the most matching images will be retrieved and then sorted in descending order.
Step 6: the first N similar images are retrieved; where N is number of the retrieve images required by user.

Here is Algorithm of two Evolutionary Computations:

4.3 Genetic Algorithm

- Randomly generate an initial population
  P(t) = (f1, f2,…..fn) of chromosome
  Where (f1,f2,….fn) represents initial feature set
- Chromosomes creation

Select random features from initial feature set of each chromosome
Note: - Length of the chromosome is equal size for each chromosome.

- Evaluate fitness function using standard deviation of each chromosome in the current population P(t).
- Generate new population by repeating steps until new population is complete
  Encode each chromosome
  Crossover: - with the crossover probability Pc, if rand < Pc, select chromosomes for crossover from current population. After that, mate the selected chromosomes randomly using single point crossover to form new o ffspring.
  Mutation: - Mutate the new offspring with mutation probability Pm (flip '0' or '1')
  Then, Final population pool is filled up with n chromosomes.
- Repeat the process until it reaches maximum iteration

4.4 Binary Bat Algorithm

- Initialize the bat population
- Calculate fitness value of initial bats using sum of square error
- The initial population bat which has minimum fitness is the global best (gbest)
- In all iteration, adjust velocity, frequency and the position as given in Eq. (5) (6) (8)
- Calculate sigmoid transfer function Eq.(7)
- If the randomly generated is greater than the initial pulse rate (0.9) , then the gbest is updated with the new bats.
- Calculate the fitness value of new bats
- If the initial bat's fitness is less than the new bat's fitness and the random number is greater than the initial loudness (0.5) then the initial bat is updated
- If the new bat's fitness is less than the gbest then update the gbest
- Repeat step 4, until maximum iterations have been reached

5. EXPERIMENTAL RESULT AND DISCUSSION

The proposed system is implemented using MATLAB program of version 8.1.0.604 (R2013a) which engrafted with image processing tools and statistical tools, and also made in GUI. We are considering the new approaches which the enhancement is done. We introduce database of image that we choose to test our system.

5.1 Image Database

For the experiment, we use image database in our evaluation is WANG database. We were taken 5 classes and each class contains 20 images. The classes are Flowers, Bus, Food, Elephant, and Dinosaurs.

5.2 Feature Extraction

The features extracted from images include color (color moment), texture (Gabor filter), and shape (Edge Histogram Descriptor). An image represented using 18 features using color moment for RGB and HSI. Mean and standard deviation for Gabor filters for 16 scales and 4 orientations are calculated, so the Gabor feature set consists of 128 features. Edge histogram descriptor is calculated along with four edge densities vertical, horizontal, 45 degree, and 135 degree; 150 features are obtained. So totally 296 features are obtained, and stored in feature database. Then, perform feature selection using Binary Bat Algorithm. Selected features are stored in cluster database using k-means algorithm.

5.3 Performance Measurements

In CBIR, the most commonly used performance measures are Precision and Recall.

Precision:

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

Recall:

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

5.4 Expected Outcome

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

Feature Selection using Genetic Algorithm

```
Genetic Algorithm
Selected features  =
124  71  145  195  200  170  182  174  212  128  233  270  133  206  219

Number of features selected  = 85
```

After calculating the 296 features for all images in database, Only 85 features are selected from 296 features using Genetic algorithm.

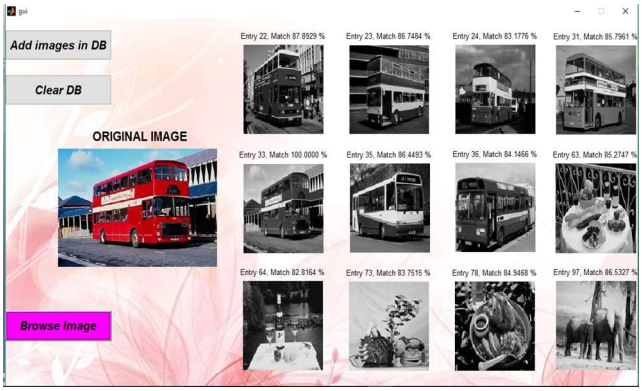

Fig- 2. Feature Selection using GA based Retrieved Images

In figure 2, perform distance between query image and images in cluster database are calculated using Manhattan

distance and the images which have minimum distance which is similar to query image are retrieved. In the figure, seven images are match to query image out of 12 images.

1) Feature Selection using Binary Bat Algorithm:

```
Binary Bat Algorithm
Selected features  =
3  67  135  73  109  25  178  216  107  126  227  110  249  266  149  93

Number of features selected  = 35
```

After calculating the 296 features for all images in database, Only 35 features are selected from 296 features using Binary Bat algorithm.

Fig.3 Feature Selection using BBA based Retrieved Images

After test the image in system, all images are matched to query image. So, feature selection using binary bat algorithm have less features subset and perform better than feature selection using genetic algorithm.

Table-1: Comparative Analysis of Feature Selection using GA and Feature Selection using BBA

	Genetic Algorithm			Binary Bat Algorithm		
Images	Time (s)	Precision (%)	Recall (%)	Time (s)	Precision (%)	Recall (%)
Flowers	10.309	83.4	25	8.979	96.8	30.9
Dinosaurs	8.6770	58.3	35.5	8.008	100	60
Food	9.9859	20	10.8	7.007	75	45
Elephant	8.8531	33.4	20.6	7.925	83.4	50.2
Buses	8.3678	70	35	7.985	100	50

Table 1 shows feature selection using GA is compared with Feature Selection using BBA in terms of precision, recall, and computational time. The computational time of feature selection using BBA takes less than feature selection using GA. The Precision and Recall of BBA have more and perform better than GA.

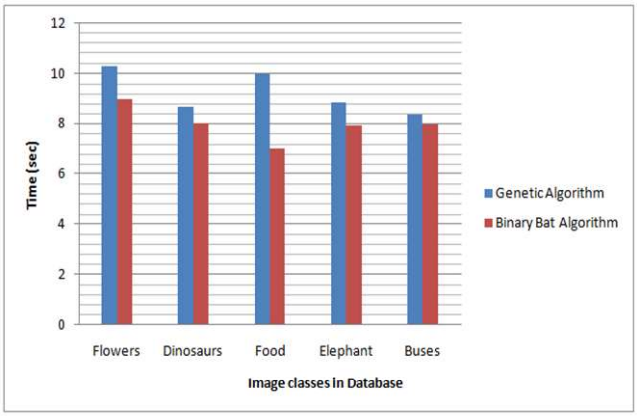

Chart -4: Time (s) vs. Image classes in Database

Figure 4 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 by feature selection using BBA. The bar chart shows, by using feature selection based on BBA takes less time and fast retrieval. It is more efficient and better performance than feature selection using GA.

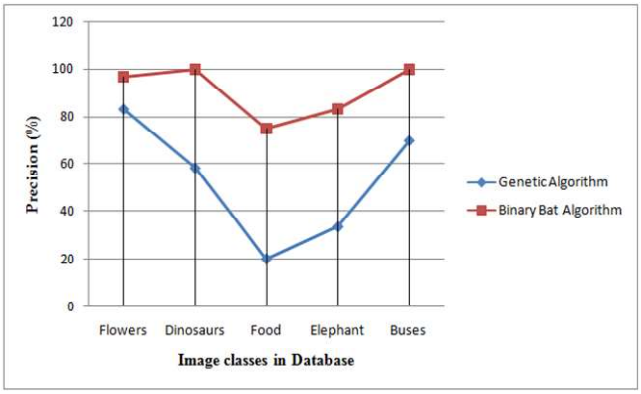

Chart -5: Precision (%) vs. Image classes in Database

Chart – 6: Recall (%) vs. Image classes in Database

Figure 5 and Figure 6 shows, precision and recall of each image class in database. Blue indicates precision and recall of each image class feature selection using GA. Red indicates precision and recall of each image class using feature selection based on BBA. The chart shows highest precision and recall of each image class using feature selection based on Binary Bat Algorithm is perform better than feature selection using Genetic algorithm.

6. CONCLUSION

In this paper, we introduce and implement feature selection for CBIR system based on evolutionary computation. 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 evolutionary computation for best optimal features from feature set, and also use clustering for reduces elapsed time of system. The experimental result shows feature selection using Binary Bat Algorithm reduces the time for retrieval and also increases the retrieval precision and recall, thus it gives better and faster results as compared to feature selection using Genetic Algorithm. The elapsed time of the system is reduced. From the result, it is clear that feature selection using BBA is more optimize the searching time in seconds and also shows highest precision and recall of each image class as compared with feature selection using GA.

In future work, we try to use other optimization algorithm for scale down computational time and provide better accuracy result of CBIR system.

ACKNOWLEDGEMENT

I express my sincere gratitude to my guide Mrs**.** R.R. Welekar, Professor (CSE Department, SRCOEM, Nagpur, India) for her constant help, worthful guidance, and encouragement during the work. I would like thanks to her for showing me some examples related to the topic of my research.

REFERENCES

[1]. Dr. R.R.Rajalaxmi, A.Sylvia Rani, "Unsupervised feature selection using Binary bat algorithm", 2'nd International conference of electronics and communication system (ICECS), *IEEE* sponsored, 2015, pages: 451-456

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

[3]. 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.

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

[5]. 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.

[6]. 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.

[7]. 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.

[8]. 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.

[9]. 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.

[10]. 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.

[11]. 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.

[12]. 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.

[13]. 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.

[14]. 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.

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

[16]. 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.

[17]. 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

[18]. 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

[19]. 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.

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