

# Content Based Image Retrieval with Multi-Feature Classification by Back-propagation Neural Network

Suman Khokhar  
Department of Computer Science and Engineering  
MATS University  
Raipur, India

Satya Verma  
Head of Department  
Computer Science and Engineering  
MATS University  
Raipur, India

**Abstract:** Emergence of Internet, as well as digital image acquisition technology, has increased the usage of rich visual information such as images and videos. It has become the integral and essential part of everyone's life. Since image production has become easy and economical, images and videos are used extensively on the Internet so retrieving relevant images from large ever-growing image dataset has become a challenge. To combat this problem, one of the popular image retrieval approaches is Content-Based Image Retrieval (CBIR) which utilizes the visual features of the image i.e. color, texture, geometric (shape) and spatial information to retrieve visually similar images according to the given query image. This paper aims to exploit multiple features of an image i.e. color, geometric and texture with the Back-propagation Feedforward Neural Network (BFNN) for classification. Feature selection method is also exercised to focus on important features of an image and ignoring redundant and inappropriate information. The results have shown that CBIR technique with multi-feature classification by BFNN yields better precision and recall as compared to other state-of-the-art techniques of CBIR that uses single or other hybrid combination of features of an image.

**Keywords:** Back Propagation Feedforward Neural Network (BFNN); Content-Based Image Retrieval (CBIR); Gray Level Co-occurrence Matrix (GLCM); Text-Based Image Retrieval (TBIR); Zernike Moments (ZM);

## 1. INTRODUCTION

The Internet and World Wide Web (W3) has revolutionized everyone's life. Today it has become a preferred medium of daily communication. There is an English idiom which says "A picture/image is worth a thousand words". Any complicated idea can be easily remembered and conveyed via a single still image. An image exhibits some meaning and can be thought of a visual representation of some subject.

### 1.1 Motivation

Due to the scientific and technological advances in data storage and image acquisition methodologies, the volume of digital data i.e. images and videos have increased considerably [1]. To combat the problem of the ever-growing population of images, their organization, and management, image retrieval systems were developed. These systems retrieve images from the large image database on the basis of the user's query. There are three kinds of retrieval systems which are depicted in Figure 1.

Since 1979, image retrieval systems have been exploiting the techniques that are based on the textual description or annotation of images. This approach is known as *Text / Tag Based Image Retrieval (TBIR)* [2].

There are certain drawbacks of TBIR approach i.e. it is a time-consuming and slow process as it requires manual description of each image in database, costly and different human have a different understanding of the image making it subjective and context-sensitive.

An image can describe itself completely and efficiently as compared to any other descriptor, therefore an another approach called *Content Based Image Retrieval (CBIR)* was introduced and has been a major area of research in image

processing field. CBIR takes a query image (i.e. example image or sketch) as an input and retrieves visually similar images by comparing visual features of query image with the visual features of images in the image database. It uses the content of the images to represent and index them.

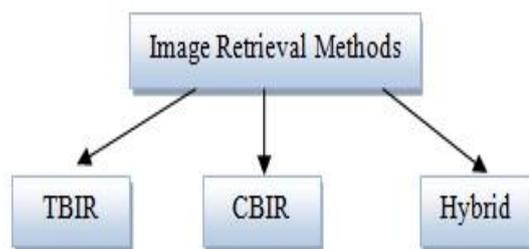


Figure 1: Types of Image Retrieval System

The visual features of an image include color, shape (geometric), texture and any other information that can be extracted from the image. These visual features are described by multi-dimensional feature vectors. This approach is also known by different names such as *Query-By-Image Content (QBIC)*, *Content-Based Visual Information Retrieval (CBVIR)* and *Reverse Image Search*. CBIR has found its application in many fields like medical diagnosis, satellite picturing, military, intellectual property, object recognition, fingerprint identification, face recognition, etc.

In *Mixed / Hybrid approach*, both textual query and visual features are combined (also known as fusion) to obtain the desired output image. This is a demanding research area in the field of multimedia information retrieval.

## 1.2 Related Work

In [3], Swapnil Saurav *et al.* extracted color features using various color models and different color descriptors such as color histogram, color moments and color coherence vectors with Support Vector Machines (SVM) classifier. In [4], Muhsina Kaipravan *et al.* proposed a system which integrates both color and texture features. To extract color and texture features, color moment and Gabor filter are employed respectively. This technique is implemented and verified on Wang image database.

In [5], Manpreet Kaur *et al.* presented a novel technique which combines image features like color, texture, and edges. Retrieval techniques like HSV, color moment are used for extracting color features, Gabor wavelet and wavelet transform for retrieving texture features with edge gradient technique for detecting shape features. In [6], Sakthivel Karthikeyan *et al.* proposed a system for Glaucoma diagnosis via Gray Level Co-occurrence Matrix (GLCM) for extracting the textual features and used Sequential Forward Floating Selection (SFSS) technique for selecting extracted features which are fed to Back Propagation Network (BPN).

In [7], Habib Mahi *et al.* proposed a system to recognize the buildings from Very High Spatial Resolution (VHSR) satellite images. The system uses Zernike Moments based shape descriptor after segmenting the image into homogeneous objects via MeanShift segmentation method and is fed to SVM for further classification purpose. In [8], Sudhir P. Vegad *et al.* proposed a hybrid approach which exploits color features with feed-forward BPN. The system uses HSV color model for obtaining color features.

The proposed CBIR system extracts color, shape and texture features of the image by applying *RGB Color Histogram*, *Zernike Moments* and *Gray Level Co-occurrence Matrix (GLCM)* respectively to attain the optimal retrieval accuracy and better performance in CBIR systems. Once all features are extracted from the image, *ReliefF* Filter Feature Selection technique is applied to choose appropriate features required to represent an image effectively. The System is trained and images are classified using *Back-propagation Feed-forward Neural Network (BFNN)*.

Since *LAB color space* model exceeds the gamut of the RGB and CMYK color models and is designed to approximate human vision, it is employed as a pre-processing step to extract texture and shape features. Images with the close resemblance to the query image are retrieved and indexed accordingly.

This paper is organized as follows: In Section II, CBIR architecture is explained. Proposed work is discussed in detail in Section III with feature extraction techniques. Section IV presents experimental results with the discussion on performance evaluation of different techniques. Finally, the paper is concluded in Section V with future research work recommendations.

## 2. CBIR ARCHITECTURE

The CBIR system analyzes image content and retrieves visually similar images from image database according to the query image. A query image can be any sketched figure or an example image. The visual content of an image is known as image features which are extracted by applying different techniques and are internally represented as feature vectors.

The image features are the properties of an image, for instance, color, texture, shape or spatial layout and these features are considered as low-level features of an image. CBIR system performs two main functions i.e. Feature Extraction and Similarity Measurement which is shown in Figure 2.

In the *Feature Extraction* phase, the features of images are extracted and these features are matched in *Similarity Measurement* phase.

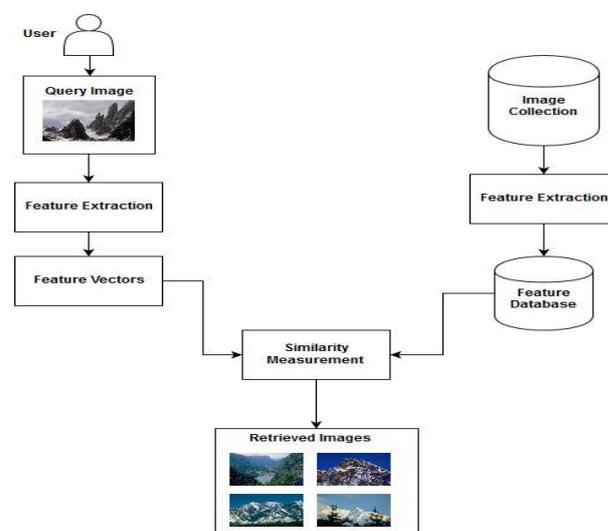


Figure 2: CBIR System

The feature vector of each image in the database is compared with the feature vector of the query image and the images with the small distance to the query image are considered as matched images and are retrieved from the image database.

## 3. PROPOSED APPROACH

The CBIR system extracts color, shape and texture features of the image. The performance of CBIR system depends on the method to obtain suitable features from the image. *The color* is the most widely used visual property of the image. It is a robust descriptor as it is uniform with respect to scaling, translation, and rotation of an image. *Texture or gray-level features* depicts visual patterns in an image and describes how they are spatially located. GLCM method is a very effective method for acquiring texture features and is applied in this paper.

*The shape* of an object or region in an image is represented by shape features. These features are commonly extracted and obtained after image segmentation. Region- growing method of region-based segmentation is employed from which Zernike moments are computed. The details of the methods used for obtaining image features are stated below in detail:

### 3.1 Gray-Level Co-occurrence Matrix

GLCM is a popular statistical method for obtaining texture features from images. It calculates co-occurring grayscale values at a specified offset over an image.

It produces a matrix with the direction (i.e. horizontal, vertical, left and right diagonals) and distances between the pixels. It is also known as Grey Tone Spatial Dependency Matrix and is used for calculating the "second order" texture features. GLCM is a square matrix with dimension  $N_g$ , where  $N_g$  refers to the number of gray levels in the image.

Each element  $[i, j]$  of the GLCM matrix is calculated by counting the number of times a pixel holding value  $i$  is adjacent to a pixel holding value  $j$  i.e.  $V[i, j]$ . The relation between the two pixels is computed at a time by considering two-pixel values i.e. reference pixel and neighbor pixel. The following Table 1 shows the structural arrangement of GLCM matrix elements.

**Table 1:** GLCM Matrix Structure

Neighbor pixel value / Reference Pixel value	0	1	2	...	$N_g$
0	0,0	0,1	0,2	...	0, $N_g$
1	1,0	1,1	1,2	...	1, $N_g$
2	2,0	2,1	2,2	...	2, $N_g$
.					
.	...	...	...	...	...
.					
$N_g$	$N_g,0$	$N_g,1$	$N_g,2$	...	$N_g, N_g$

Since the texture measure requires that each GLCM cell contains not a count, but rather a probability, therefore GLCM is normalized by dividing each element by the sum of values. The formula for computing probability matrix element is given below where matrix element  $P(i, j | \Delta x, \Delta y)$  depicts relative frequency with the given two neighborhood pixels, one with intensity 'i' and the other with intensity 'j' and pixels are separated by a pixel distance  $(\Delta x, \Delta y)$  [9].

$$P(i, j) = \frac{V(i, j)}{\sum_{i,j=0}^{N_g} V(i, j)} \quad (1)$$

The GLCM expressed as a probability matrix is given by

$$G = \begin{bmatrix} P(0,0) & P(0,1) & \dots & P(0, N_g) \\ P(1,0) & P(1,1) & \dots & P(1, N_g) \\ \vdots & \vdots & \ddots & \vdots \\ P(N_g, 0) & P(N_g, 1) & \dots & P(N_g, N_g) \end{bmatrix}$$

This paper extracts 22 features from GLCM matrices and their formula [6][7][8][9][10][14] is mentioned in Table2.

**Table 2:** Features extracted from GLCM

S. No.	Features	Formula
1	Entropy	$\sum_{i,j=0}^{N_g} P(i, j) \log(P(i, j))$
2	Energy	$\sum_i \sum_j P(i, j)^2$
3	Contrast	$\sum_i \sum_j (i - j)^2 P(i, j)$
4	Homogeneity	$\sum_i \sum_j \frac{P(i, j)}{1 + (i - j)^2}$
5	Correlation	$\frac{\sum_i \sum_j (i, j) P(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
6	Variance: Sum of Squares	$\sum_i \sum_j (i - \mu)^2 P(i, j)$
7	Autocorrelation	$\frac{\sum_i \sum_j (i, j) [P(i, j) - \mu_x][P(i, j) - \mu_y]}{\sigma_x \sigma_y}$ Where $\mu_x, \mu_y$ are means and $\sigma_x, \sigma_y$ are standard deviations.
8	Inverse Difference (INV)	$\sum \frac{P(i, j)}{1 +  i - j }$
9	Sum Average	$\sum_{i=2}^{2N_g} i P_{x+y}(i)$
10	Sum Entropy	$f_8 = - \sum_{i=2}^{2N_g} P_{x+y}(i) \log\{P_{x+y}(i)\}$
11	Sum Variance	$\sum_{i=2}^{2N_g} (i - f_8)^2 P_{x+y}(i)$
12	Difference Variance	$\sum_{i=0}^{N_g-1} i^2 P_{x-y}(i)$
13	Difference Entropy	$-\sum_{i=0}^{N_g-1} P_{x-y}(i) \log\{P_{x-y}(i)\}$
14	Information Measure of Correlation1	$\frac{HXY - HXY_1}{\max\{HX, HY\}}$ where $HXY = - \sum_i \sum_j P(i, j) \log(p(i, j))$ $HXY_1 = \sum_i \sum_j P(i, j) \log\{P_x(i)P_y(j)\}$ $HX, HY$ are the entropies of $P_x$ and $P_y$ .

15	Information Measure of Correlation2	$(1 - \exp[-2(HXY_2 - HXY)]^{\frac{1}{2}})$ where $HXY_2 = - \sum_i \sum_j P_x(i)P_y(j) \log\{ P_x(i)P_y(j)\}$
16	Maximum Correlation Coefficient	Square root of the second largest eigenvalue of Q where $Q(i, j) = \sum_k \frac{P(i, k)P(j, k)}{P_x(i)P_y(k)}$
17	Shade	$sgn(A) A ^{\frac{1}{3}}$ where $A = \sum_{i,j=0}^{N_g-1} \frac{P(i,j)(i+j-2\mu)^3}{\sigma^3(\sqrt{2(1+C)})^3}$ and C= Correlation feature
18	Prominence	$sgn(B) B ^{\frac{1}{4}}$ where $A = \sum_{i,j=0}^{N_g-1} \frac{P(i,j)(i+j-2\mu)^4}{4\sigma^4(1+C)^2}$
19	Inverse Difference Moment Normalized (IDN)	$\frac{\sum_i \sum_j \frac{1}{1 + (i - j)^2} P(i, j)}{\sum_i \sum_j P(i, j)}$
20	Inverse Difference Normalized (INN)	$\frac{INV}{\max(INV)}$
21	Maximum Probability	$\max\{P(i, j)\}$
22	Dissimilarity	$\sum_i P(i, j) i - j $

### 3.2 Zernike Moments

ZM is an excellent shape descriptor due to its description capability. It is a region-based moment and is very effective in image classification system. It is simply a projection of image functions onto polynomial orthogonal basis function. Rotation invariant and orthogonal property of ZM makes it suitable for shape feature extraction. The ZM can be computed on an image with the size N X N with the following expression [11].

$$Z_{n,m} = \frac{n+1}{\lambda_N} \sum_{c=0}^{N-1} \sum_{r=0}^{N-1} f(c, r) R_{n,m}(\rho_{cr}) e^{-jm\theta_{cr}} \quad (2)$$

Where f(c, r) is the image function in which c is column and r is row,  $\lambda_N$  represents a normalization factor,  $n \in N$  in which n is a positive integer and represents the order of the radial polynomial and  $R_{n,m}(\cdot)$  represents the real-valued 1-D radial polynomial, m represents repetition of the angle and satisfies the following constraints:

$$N - |m| = \text{even and } |m| \leq n$$

$\rho_{cr}$  is transformed distance and ranges from  $0 \leq \rho_{cr} \leq 1$  and  $\theta_{cr}$  is a phase which ranges from  $0 \leq \theta_{cr} \leq 2\pi$  at the pixel (c, r) and has the following formula:

$$\rho_{cr} = \frac{\sqrt{(2c-N+1)^2 + (2r-N+1)^2}}{N} \quad (3)$$

$$\theta_{cr} = \tan^{-1} \left( \frac{N-1-2r}{2c-N+1} \right) \quad (4)$$

This paper extract features using ZM where the order of the moment i.e. n is equal to 4 and repetition of the moment i.e. m is equal to 2.

### 3.3 Backpropagation Neural Network

BFNN is a network that is trained with a backpropagation training algorithm that calculates error signal by finding the difference between the training output from the target output. It backpropagates the errors, adjusts the weights and biases in the input and hidden layers to get close to the desired outcome. The architecture of BFNN consists of three layers i.e. input, hidden, and output layer [12] which is given in Figure 3.

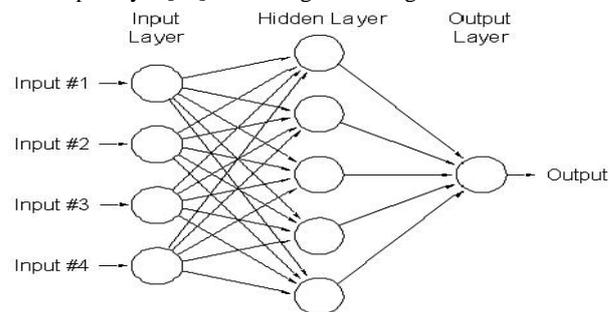


Figure 3: Basic BFNN Architecture

Images are divided into different classes in the image database and some images from each class are used for training the BFNN. A BFNN is repeated until a less training error is achieved. Once training sample is fed to the BFNN, the trained BFNN assigns one class to the query image and retrieve images of the same class as an output.

It sorts the output according to the following formula [12] which represents the distance between the query image and the retrieved images:

$$d_i = \left( \frac{1}{N} \sum_{j=1}^N W_j (f_j^i - f_j^q)^2 \right)^{\frac{1}{2}} \quad (5)$$

Where  $f_j^q$  represents the j<sup>th</sup> feature of the query image,  $f_j^i$  represents the j<sup>th</sup> feature of the retrieved image,  $W_j$  is the weight of j<sup>th</sup> feature and N is the dimension of the feature vector.

### 3.4 RGB Color Histogram

Color histogram identifies the proportion of pixels within an image carrying specific values. As it is independent of image size or orientation, the color histogram is obtained from RGB images with a number of bins equal to 3.

### 3.5 Relief Filter Feature Selection

Feature Selection is considered as a basic problem in pattern recognition system. As the selection of features in filter selection method is independent on the classifier used, therefore Relief filter method is implemented to calculate the quality of attributes or features by finding how their values differentiate between instances. It is very effective in solving

multi-class problems and is robust to noisy and incomplete data.

### 3.6 Proposed Algorithm

1. Load an RGB query image.
2. Implement LAB conversion which yields lightness (L) component and two color components a, and b.
3. Perform region-growing based segmentation on L component of the query image.
4. Apply Zernike Moments on the segmented image to obtain amplitude (A) and phase ( $\phi$ ) value of the moment.
5. Apply GLCM technique for extracting 22 texture features from query image.
6. Color features are obtained using RGB histogram with a number of bins equal to 3, therefore extracting 27 color features.
7. while (image in the training set)
8.     do
9.         LAB conversion
10.         calculate ZM and GLCM features
11.         obtain color features
12.         add features to feature database
13.     end;
14. end while;
15. Apply ReliefF feature selection technique which returns top 20 relevant features for similarity measurement.
16. BFNN is used as a classifier which takes 20 selected features for training network with 70% training data, 15% validation data and 15% for testing data with the number of hidden neurons as 20.
17. Finally, the proposed system returns class name of the query image and retrieves similar images.

### 3.7 Proposed System Flowchart

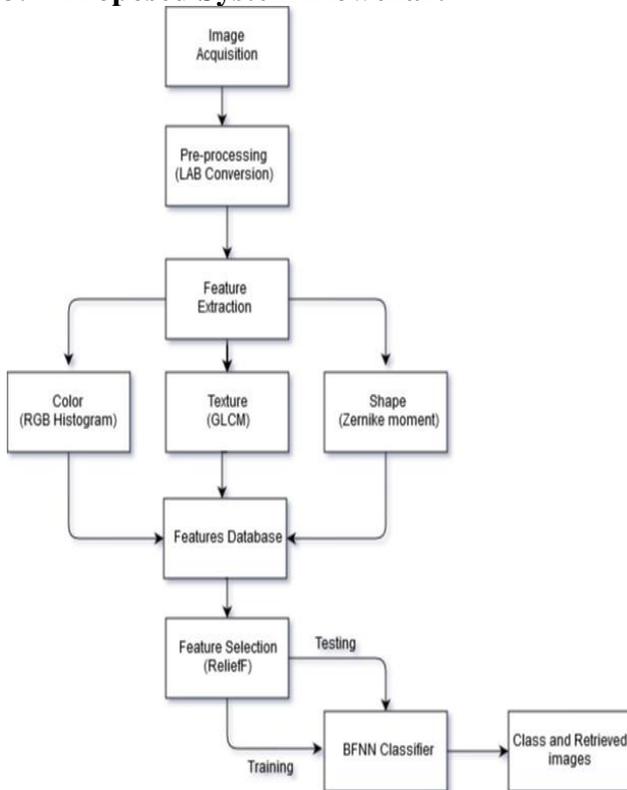


Figure 4: Proposed System Architecture

## 4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed system, MATLAB (R2016a) framework is used. The assessment of the performance of image retrieval methods is accomplished using two important measures known as *Precision* and *Recall*.

### 4.1 Implementation Environment and Dataset

The user interacts with the system via Graphical User Interface (GUI) which is created by a tool called GUIDE, the GUI Development Environment in MATLAB. To check the effectiveness of our system, the experiments are conducted on Wang image database. The Wang image dataset consists of 1000 images which are divided into 10 classes where each class includes 100 images. The categories of Wang dataset is presented in Table 3.

Table 3: Categories of Wang Image Dataset

S.No.	Categories	Images	S.No.	Categories	Images
1	African		6	Flower	
2	Beach		7	Food	
3	Bus		8	Hills	
4	Dinosaur		9	Horse	
5	Elephant		10	Monument	

### 4.2 Comparative Analysis and Performance Measurement

Precision depicts the “quality or exactness” of image retrieval whereas recall represents the “completeness.

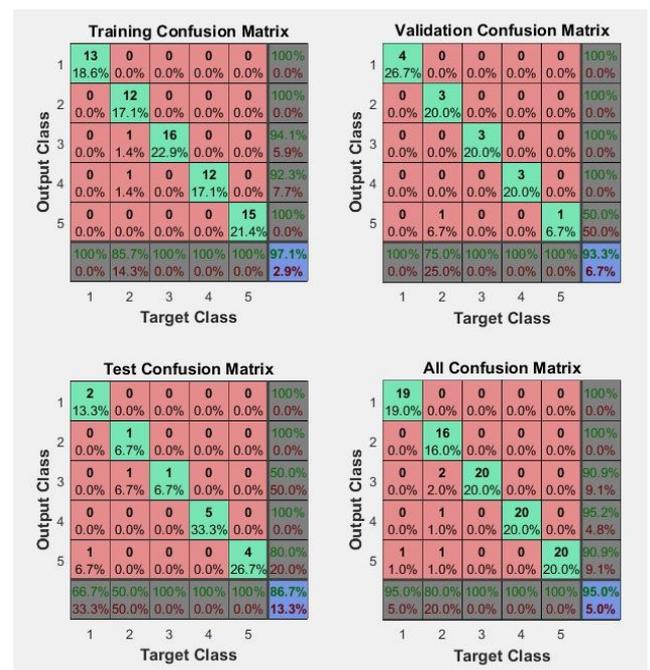


Figure 5: Plot of Confusion Matrix

Precision tells how many retrieved items are relevant and recall depicts how many relevant items are selected from the database. Precision and recall values are calculated from *confusion matrix* which is also known as error matrix and is a table that describes the performance of the classifier.

The confusion matrix for the first 5 image categories (i.e. African, Beach, Bus, Dinosaur, and Elephant) is displayed below in figure 5 providing African image as a query.

The formula for calculating the precision, recall, average precision and average recall is given below:

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$Precision_{avg}^{n,j} = \frac{1}{N_i} \sum_i^{N_i} precision_i(n) \quad (9)$$

$$Recall_{avg}^{n,j} = \frac{1}{N_i} \sum_i^{N_i} recall_i(n) \quad (10)$$

where TP is true positive, FP is false positive, and FN is false negative value in confusion matrix.

**Table 4:** Precision values of 5 categories

Class					
Precision	African	Beach	Bus	Dinosaur	Elephant
1	100	75	100	100	100
2	95	80	100	100	100
3	100	80	85	100	100
4	95	95	95	100	85
5	100	80	100	90	100
6	100	85	85	100	100
7	100	100	85	100	95
8	100	85	95	100	85
9	95	90	100	95	95
10	100	85	85	100	100

The Table 4 and 5 represent the precision values and recall values for evaluating the performance of BFNN classifier for 5 categories. The average precision and recall of proposed system are displayed in Figure 6 whereas comparative analysis of proposed system with the color moment and Gabor filter technique [4] is presented in Figure 7.

**Table 5:** Recall values of 5 classes

Class					
Recal l	Africa n	Beac h	Bu s	Dinosau r	Elephan t
1	90	100	87	100	100
2	100	100	90	95	90
3	100	84	94	100	86
4	100	90	86	95	100
5	90	100	95	100	87
6	95	89	94	100	90
7	90	90	100	100	100
8	90	93	81	100	90
9	95	100	90	100	100
10	95	89	94	100	90

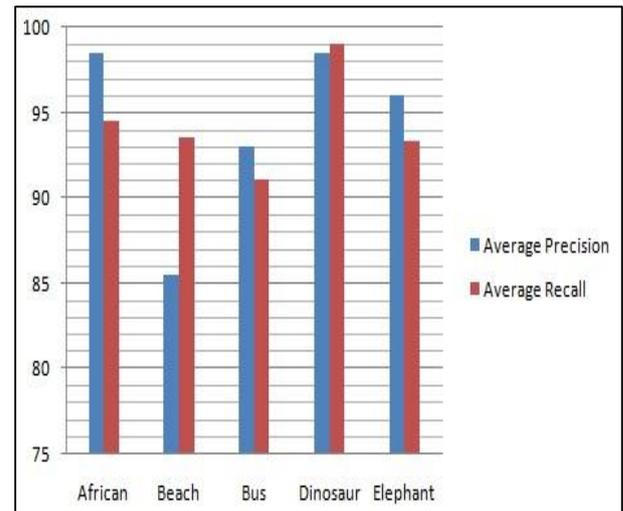


Figure 6: Average Precision and Recall values of 5 classes

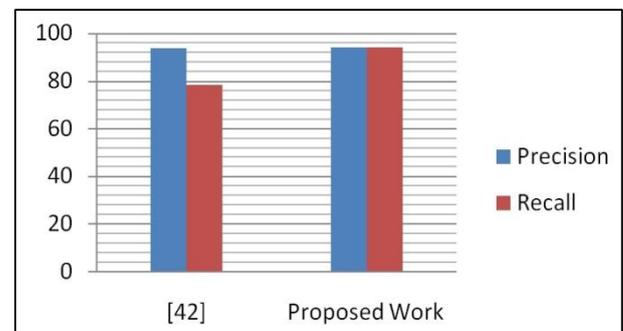


Figure 7: Comparison of proposed work with existing technique [4]

## 5. DISCUSSION

By the experimentation conducted on Wang image dataset, it is observed that the proposed system has attained the overall precision of 94.3% and recall of 94.28% in contrast to 94% precision and 78.4% recall of the system [4] with higher retrieval accuracy of 94.28%.

The proposed system has been able to achieve the optimal accuracy and better performance with significant improvement in recall measure.

## 6. CONCLUSION

This paper presented a CBIR approach that combines multiple visual features of the image to retrieve visually similar images. As only one or two features are not sufficient to describe an image completely, it is suggested to use all low-level features i.e. color, shape, and texture of the image to realize it effectively. Texture features are extracted using GLCM technique, geometric features are computed by applying Zernike Moments whereas, for color features, the RGB color histogram is employed. All extracted features are applied to ReliefF feature selection algorithm which selects top 20 features for BFNN classification. Experimental results show that the performance of the proposed work outperforms the other conventional image retrieval methods by attaining higher precision values, recall values, and accuracy of 94.28%.

In the future work, deep neural network (DNN) or more efficient color extraction technique can be availed to further improve image retrieval accuracy. The system can be customized for transfer learning.

## 7. REFERENCES

- [1] Oge Marquees, Florida Atlantic University. 2016. Visual Information Retrieval – The State of the Art, Published by the IEEE Computer Society.
- [2] Dr. Fuhui Long, Dr. Hongjiang Zhang and Prof. David Dagan Feng. Fundamentals of Content Based Image Retrieval.
- [3] Swapnil Saurav, Prajakta Belsare, and Siddhartha Sarkar. 2015. Holistic Correlation of Color Models, Color Features and Distance Metrics on Content-Based Image Retrieval, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 02 Issue: 07.
- [4] Muhsina Kaipravan, Rejiram R. 2016. A Novel CBIR System Based on Combination of Color Moment and Gabor Filter, Conference on Data Mining and Advanced Computing (SAPIENCE), INSPEC Accession Number: 16540589, IEEE.
- [5] Manpreet Kaur, NeelofarSohi. 2016. A Novel Technique For Content Based Image Retrieval Using Color, Texture And Edge Features, International Conference on Communication and Electronics Systems (ICCES), INSPEC Accession Number:16776306, IEEE.
- [6] Sakthivel Karthikeyan, N. Rengarajan. 2014 Performance analysis of gray level cooccurrence matrix texture features for Glaucoma diagnosis, American Journal of Applied Sciences 11 (2): 248-257, 2014 ISSN: 1546-9239, Science Publication.
- [7] Habib Mahi, Hadria Isabaten, and Chahira Serief. 2014. Zernike moments and SVM for Shape Classification in Very High-Resolution Satellite Images, The International Arab Journal of Information Technology, Vol. 11, No. 1.
- [8] Sudhir P. Vegad, Prashant and K. Italiya. 2015. Image Classification using Neural Network for Efficient Image Retrieval, International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO).
- [9] Seema Kolkur, D.R. Kalbande. 2016. Survey of Texture-Based Feature Extraction for Skin Disease Detection, IEEE.
- [10] Robert M. Haralick, K. Shanmugam, and It'shak Dinstein. 1973. Textural features for Image Classification, IEEE Transactions on systems, man, and cybernetics.
- [11] David A. Clausi. 1989. An analysis of co-occurrence texture statistics as a function of gray level quantization, Canadian Journal of Remote Sensing, Can. J. Remote Sensing, Vol. 28, No. 1, pp. 45–62, 2002M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science.
- [12] Amir Tahmasbi n, FatemehSaki, and Shahriar B. Shokouhi. 2011. Classification of benign and malignant masses based on Zernike moments, Computers in Biology and Medicine, 726–735, Elsevier.
- [13] Jun-Hua Han, De-Shuang Huang. 2005. A Novel BP-Based Image Retrieval System, IEEE.
- [14] Michael V. Boland.1999. Haralick texture features. [http://murphylab.web.cmu.edu/publications/boland/boland\\_node26.html](http://murphylab.web.cmu.edu/publications/boland/boland_node26.html), Available, [online] [Accessed]:12-Apr-2017.