Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 5, May 2015, pg.148 – 157

RESEARCH ARTICLE

AN EFFICIENT CONTENT BASED IMAGE RETRIEVAL SYSTEM USING SOFT COMPUTING TECHNIQUE

Manpreet Singh[1], Sumit Chopra[2], Jagdeep Kaur[3]

[1]Department of computer science, KCCEIT SBSNAGAR, India
[2]Department of computer science, KCCEIT SBSNAGAR, India
[3]Department of computer science, KCCEIT SBSNAGAR, India
[1] msmahal86@gmail.com; [2] Sumitonhunt@gmail.com; [3] jagdeep.kaur@kcinstitutes.com

Abstract— As the use of searching digital images has increased a lot in recent years, so a system is required which search images from large database on the bases of query image and return results in form of images which matched with query image. The present paper introduces a content based image retrieval system using artificial neural network (ANN) approach as soft computing technique. The proposed system is composed of three major phases: feature extraction, ANN classifier and matching strategy. In feature extraction phase, local and global features are extracted and then Gabor filter is applied to enhance images so that better results can be calculated. The artificial neural network (ANN) in our proposed system serves as a classifier so that selected features of a query image are used as an input to find multi classes as an output which have largest similarity to the query image. At last feature matching strategy is used to retrieve images from database as result. Finally the results presented in this paper by using proposed system are accurate and takes less computation time.

Keywords— ANN, CBIR, Color and Texture features, Similarity measures, Classification

I. INTRODUCTION

In today's world the use and processing of digital images has increased a lot, researchers keep working on developing an improved methods of retrieving images from a large database. Generally, there are two main approaches of image retrieval procedures:-

1. Annotation-based image retrieval (ABIR)
2. Content-based image retrieval (CBIR)

In ABIR, images are often annotated by keywords. ABIR approach provides accurate and best results when images are well-named or annotated, but it still has some drawbacks such as: to annotate image manually is time-consuming, annotation done by human is subjective, and there are many images which could not be annotated because it is difficult to describe their content with words. The second approach is CBIR. It works on visual content of image to index and retrieve images from large database. The goal of CBIR is to retrieve images from a database that are similar to an image placed as a query. CBIR avoids and removes many problems which occurred in ABIR approach. Thus, in recent years researchers focussing on this area to find better methods so that excellent results can be achieved. The CBIR system performance mainly depends on the image representation methods and functions that use similarity matching techniques [1].

In the last few years, a lot of research is going on content based image retrieval. The main approach used is querying by example (QBE). In this approach low-level image features like color histograms, textures and shapes are computed from query image and that features are matched with database images features to find relevant images. In CBIR, each image that is stored in the database has its features extracted and compared to the features of the query image.

The organization of this paper is as follows: Literature Review is discussed in Section II. Feature extraction is performed in Section III. Artificial neural network classifier is introduced in Section IV. Matching strategy is discussed in Section V. The performance evaluation is described in Section VI. Methodology is proposed in section VII. The experimental results are reported in Section VIII. Finally, the conclusion and future scope are presented in Section IX.

II. LITERATURE REVIEW

A.Ramesh Kumar et.al. (2013) presented a paper on content based image retrieval using color histogram. They used color histogram of an image as feature vector for retrieving images from database [19]. **Vivek kapur et.al. (2013)** used hybrid approach to retrieve images from database. They used combination of both text-based and content based feature vectors to retrieve images from database. They used pyramid-structured wavelet transform as feature matching approach [14]. **Arvind Nagathan et. al. (2014)** has used color histogram as color descriptor, GLCM (gray level co-occurrence matrix) as texture descriptor and edge histogram as edge descriptor. Then they classify the images using neural network to retrieve images from database [2]. **Shereena et. al. (2014)** has combined color and texture features as extraction features of image and then they have used neural network for classification purpose to retrieve images from database [4]. **Gunjan Khosla et.al. (2014)** has calculated the color histogram of images as feature vectors. They have used distance metrics Manhattan distance and Euclidean distance to determine similarities between a pair of images [5]. **M.E. ElAlami (2014)** has presented a new matching strategy for content based image retrieval system. He described model into three major phases namely: features extraction, ANN classifier and matching strategy. In feature extraction phase, it extracts a color and texture features. He used combination of multiple features to get better results. The artificial neural network (ANN) in this model serves as a classifier so that the selected features of query image are the input and its output is one of the multi classes that have the largest similarity to the query image. This model presents an effective feature matching strategy that depends on the idea of the minimum area between two vectors to compute the similarity value between a query image and the images in the determined class [1].

III.FEATURE EXTRACTION

CBIR results are very much dependent on feature types. Generally there are two categories of features:

- Global features
- Local features

Global features include color and texture histograms and color layout of the whole image. **Local features** include color, texture, and shape features for sub-images, segmented regions, and interest points. The present paper integrated local color and texture features to improve the system performance. Various color features are color correlogram, color moments, color co-occurrence matrix etc. Texture features are gabor texture and wavelet transform etc.

IV.ANN CLASSIFIER

Artificial neural network (ANN) is well known as powerful tool in the area of pattern classification. Various types of neural networks techniques have been employed for image classification problems because of their generalization ability. The process of feed-forward, back-propagation is repeated until the output reaches a desired accuracy, or until a given number of training cycles has been completed. Assume that, a fully connected feed-forward network is chosen for our implementation and the images database is divided into two different samples which are training, and testing samples. The training samples are used to train the network, and the network is adjusted according to its error. Testing samples are then used to provide an independent measure of the network performance during and after training. The performance of the designed classifier is measured in terms of accuracy. This term refers to the ability of the model to correctly predict the class label of new unseen data. Classification accuracy is calculated by determining the percentage of cases in which the test sets are correctly classified. A good classification test always results from high values of accuracy. The accuracy value can be calculated as follows:

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ No.of\ \mathrm{Im}ages}$$

V. MATCHING STRATEGY

A major goal of content-based retrieval is finding the best matched (most similar) images from the multimedia database with respect to a query image. This section describes the three similarity measures which we have used as our similarity measure.

1. **Spearman Similarity Measure: -** If image intensities do not contain ties when they are ordered from the smallest to the largest, then by replacing the intensities with their ranks and calculating the Pearson correlation coefficient between the ranks in two images, Spearman rank correlation will be obtained. Spearman Rank Correlation measures the correlation between two sequences of values. The two sequences are ranked separately and the differences in rank are calculated at each position, i. The distance between sequences X = (X1, X2, etc.) and Y = (Y1, Y2, etc.) is computed using the following formula:

$$1 - \frac{6\sum_{i=1}^{n}(rank(X_i) - rank(Y_i)}{n(n^2-1)}$$

where Xi and Yi are the ith values of sequences X and Y respectively.

The range of Spearman Correlation is from -1 to 1. Spearman Correlation can detect certain linear and non-linear correlations. However, Pearson Correlation may be more appropriate for finding linear correlations.

2. **Correlation Similarity Measure:-**

In this case, similarity between two items i and j is measured by computing the *Pearson r* correlation corri,j. To make the correlation computation accurate we must first isolate the co-rated cases (i.e., cases where the users rated both i and j). Let the set of users who both rated i and j are denoted by U then the correlation similarity is given by

$$Sim(i,j) = corr_{i,j} = \frac{\sum_{u\in U}(R_{u,i} - \overline{R_i})(R_{u,j} - \overline{R_j})}{\sqrt{\sum_{u\in U}(R_{u,i} - \overline{R_i})^2}\sqrt{\sum_{u\in U}(R_{u,j} - \overline{R_j})^2}}$$

Here Ru,i denotes the rating of user u on item i, R_i is the average rating of the i-th item.

3. **Relative Deviation Similarity Measure:-**

The relative standard deviation (RSD) is useful for comparing the uncertainty between different measurements of varying absolute magnitude. The RSD is calculated from the standard deviation, s, and is commonly expressed as parts per thousand (ppt) or percentage (%):

$$RSD = (\frac{S}{\bar{x}}) * 1000\ ppt\ \%$$

The %-RSD is also called the "coefficient of variance" or CV.

VI. PERFORMANCE EVALUATION

Evaluation of retrieval performance is a crucial problem in content-based image retrieval. Many different methods for measuring the performance of a system have been created and used by researchers [1]. We have evaluated performance of CBIR system on the bases of three methods –

- Confusion Matrix
- Precision
- Recall

Confusion matrix is used to compare the performance of the CBIR system using different distance metrics. To evaluate the overall performance of the CBIR system and compare the different distance metrics for retrieval accuracy, confusion matrix is calculated. A confusion matrix represents the actual classifications compared with the number of correct and incorrect prediction. The confusion matrix is n-by-n matrix, where n is the number of classes from the dataset. Each row represents the number of instances in actual class. Each column represents the number of instances in predicted class [9].

The other two are common evaluation methods namely recall (or sensitivity) and precision (or specificity). Recall measures the ability of the system to retrieve all models that are relevant, while precision measures the ability of the system to retrieve only models that are relevant. The precision and recall rates are computed by the following equations:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Where TP denotes the number of retrieved images which are similar to the query, FP indicates the number of retrieved images dissimilar to the query, and FN is the number of images in the database which are similar to the query but not retrieved. A perfect system response would result in a straight line at precision 1 up to recall 1. Curves that are higher up are better than curves that are further lower down [1].

VII. PROPOSED METHODOLOGY

We are proposing here an algorithm which provides best results as we have tested it on simple database. The database contains images from beaches, buses, Africa etc. In our algorithm, first query image is loaded to the cbir system then its local and global features are calculated. When query is successfully loads, a dataset of images is loaded which contain extracted image features. Artificial neural network back propagation learning algorithm applied for classification of image features. System is trained to find out feature weight. Once all the features are calculated from database images and from query image, then their features are matched with each other by similarity measure criteria. We have used three similarity measures Spearman, Correlation and Relative Deviation distances.

The algorithm for the CBIR system is described below:-

Step 1: Browse or input the query image to the system.
Step 2: Calculate local and global features such as color correlogram, color moments, Wavelet moments etc.
Step 3: Create the training dataset to store the extracted features of the images.
Step 4: Select the different distances to process the query image extracted features. Distances used are Correlation, Relative Deviation and spearman etc.
Step 5: Apply the gabor Wavelet to process the gray scale image.
Step 6: Apply the colour Auto Correlogram to process the colour of the images
Step 7: Apply the HSV histogram to process the image according to the Hue, saturation and intensity of the image.
Step 8: Apply artificial neural network back propagation learning algorithm to extract the contents of the images that is matched with the query image.
Step 9: Generate the confusion matrix
Step 10: Calculate the precision and recall graph and values of the extracted images.
Step 11: Analyse the results and repeat the step 2 to step 8 on multiple images.
Step 12: Stop

VIII. EXPERIMENTAL RESULTS

1. In first case we have used Spearman distance as our Similarity metrics, following results are shown:-

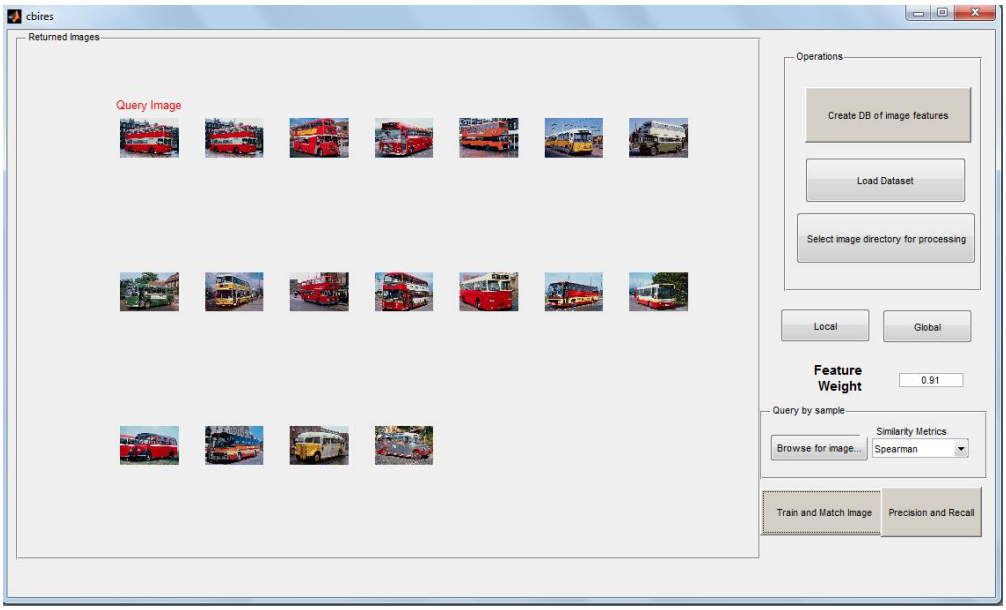

Fig 1 Retrieved images from database using Spearman Similarity measure

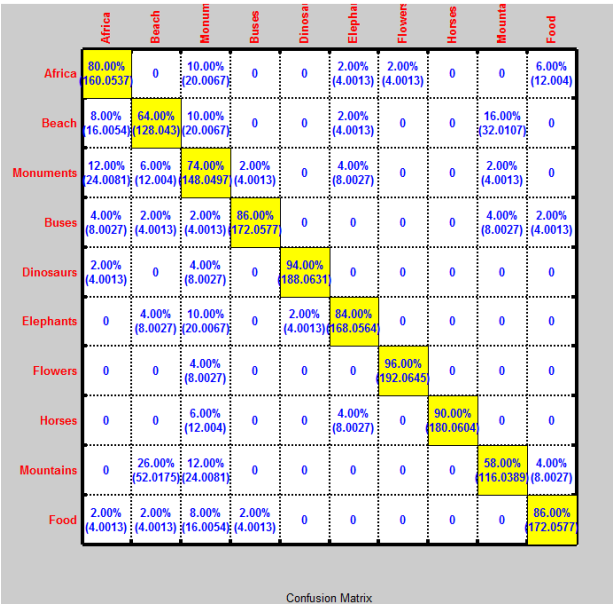

Fig 2 Confusion Matrix of Spearman Similarity Measure

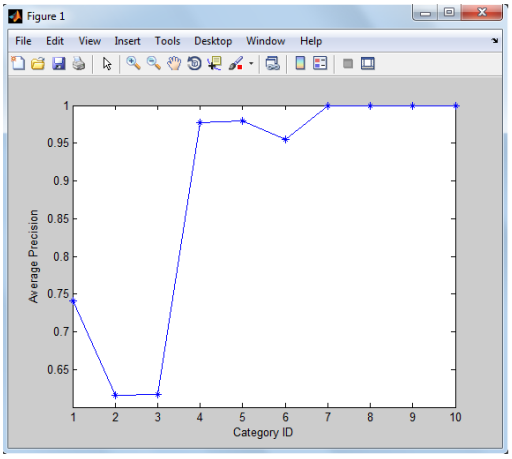

Fig 3 Precision Graph of Spearman Similarity Measure

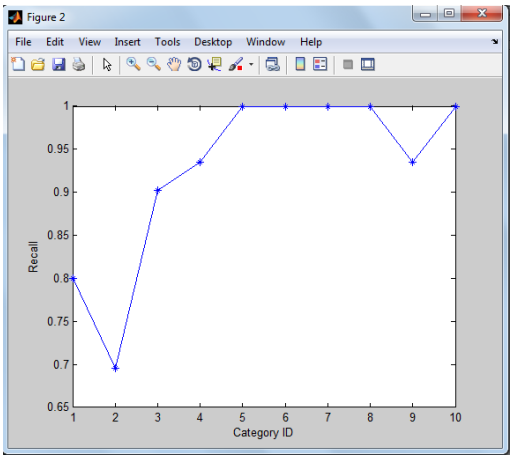

Fig 4 Recall Graph of Spearman Similarity Measure

2. In second case we have used Relative Deviation distance as our Similarity metrics, results are shown below:-

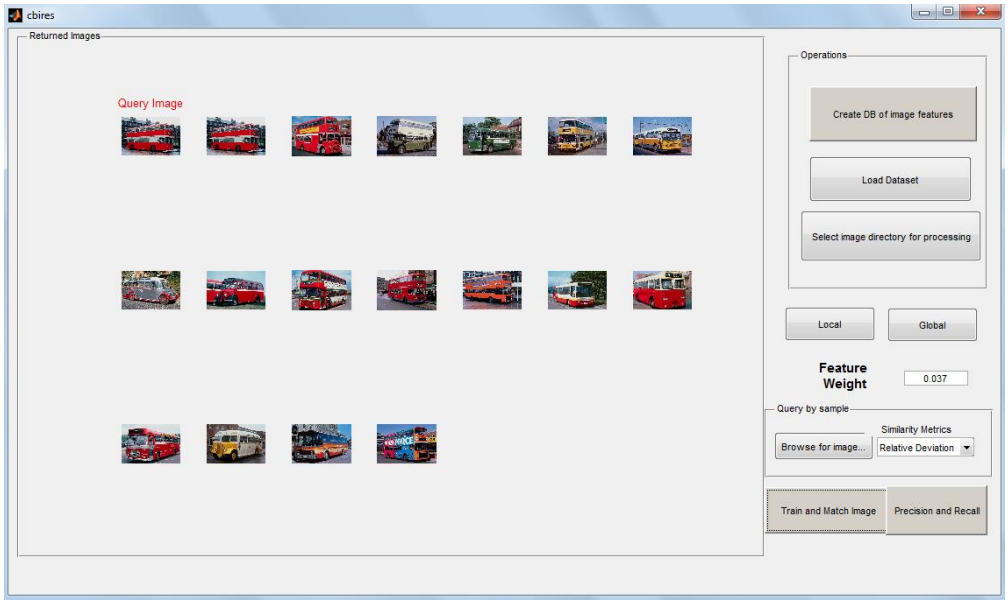

Fig 5 Retrieved images from database using Relative Deviation Similarity measure

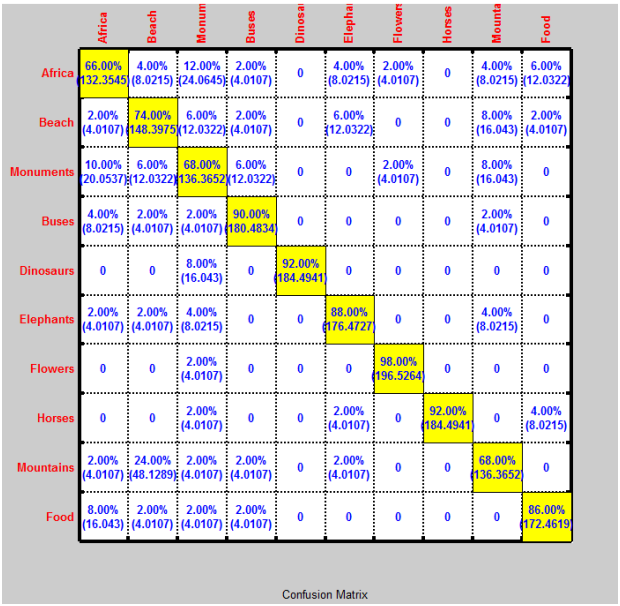

Fig 6 Confusion Matrix of Relative Deviation Similarity Measure

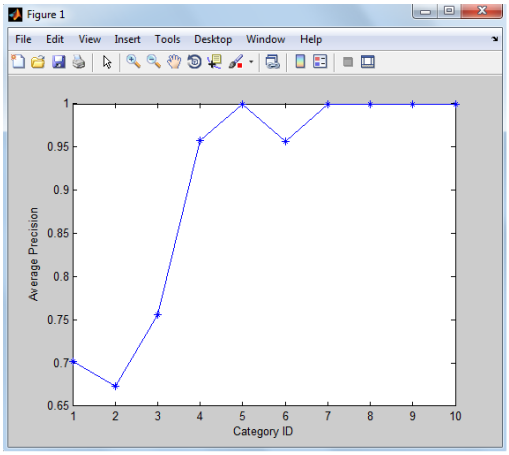

Fig 7 Precision Graph of Relative Deviation Similarity Measure

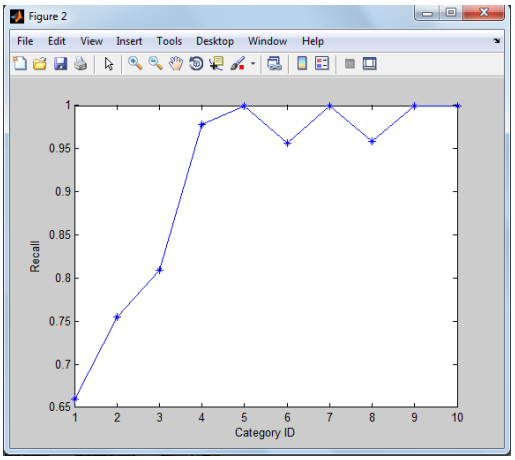

Fig 8 Recall Graph of Relative Deviation Similarity Measure

3. In third case we have used Correlation distance as our Similarity metrics, results are shown below:-

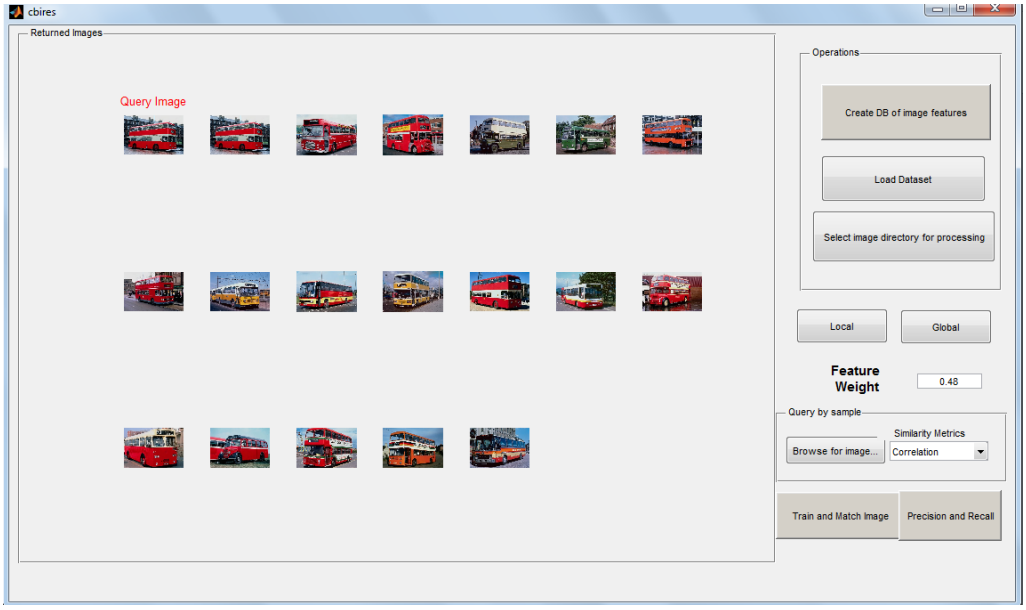

Fig 9 Retrieved images from database using Correlation Similarity measure

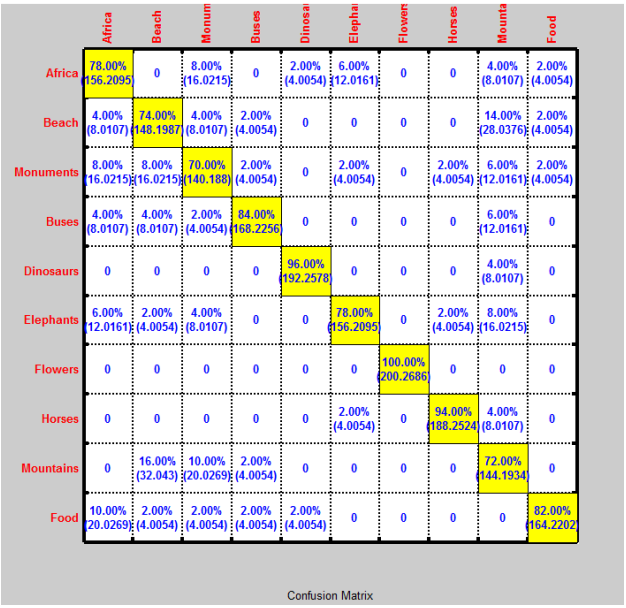

Fig 10 Confusion Matrix of Correlation Similarity Measure

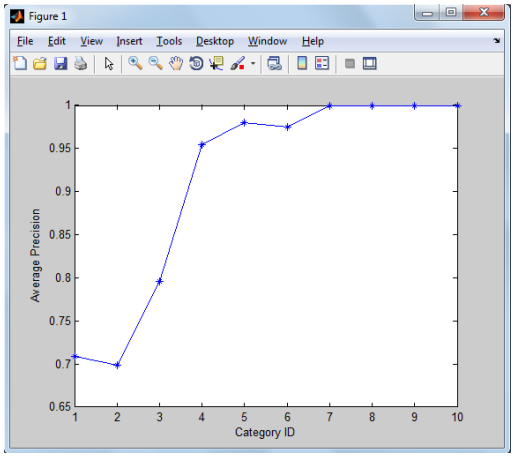

Fig 11 Precision Graph of Correlation Similarity Measure

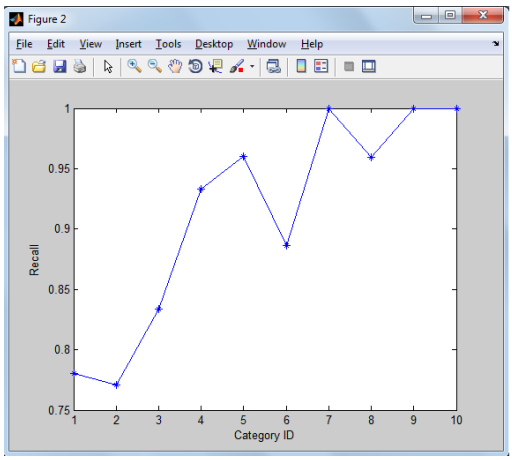

Fig 12 Recall Graph of Correlation Similarity Measure

IX. Conclusion And Future Scope

In this paper, an algorithm has been proposed to retrieve image from database which are matched with query image. Local and global features of images are calculated. Classification is done by using artificial neural network. Feed Forward back propagation learning technique is used to train and test images of database. We have used three distances spearman, Correlation and Relative Deviation as our similarity metrics. The performance of the system is shown in the form of Confusion matrix, precision and recall graphs. In future, user can use other soft computing technique like fuzzy logic, Neuro-Fuzzy techniques and use other similarity measure to make improvement in the system.

References

[1] M.E. ElAlami, "A new matching strategy for content based image retrieval system" Elsevier, Applied soft computing 14(2014).

[2] Arvind Nagathan, Manimozhi, jitendranath Mungara, “Content based image retrieval using feed forward back-propagation neural network”, IJCSNS International Journal of Computer Science and Network Security, Vol.14 No.6, June 2014.

[3] Avneet kaur chhabra and Vijay Birchha, “An efficient content based image retrieval system”, International journal of computer applications, Vol 105 - No.9, November 2014.

[4] Shereena V.B. and Julie M. David, “Content based Image Retrieval: Classification using neural networks”, The International Journal of Multimedia & Its Applications (IJMA) Vol.6, No.5, October 2014.

[5] Gunjan Khosla, Dr. Navin Rajpal and Jasvinder Singh, “Evaluation of Euclidean and Manhattan Metrics in Content Based Image Retrieval System”, International Journal of Engineering Research and Applications, ISSN: 2248-9622, Vol. 4, Issue 9(Version 1), September 2014.

[6] Deepu Rani and Monica Goyal, "A research paper on Content based image retrieval system using improved SVM Technique", International journal of Engineering and computer science, Vol 3, Issue 12, 2014.
[7] Chandrashekhar G.Patil, Dr. Mahesh T.Kotle, "Improvement in performance of cbir by using fusion and evolutionary computation", International journal of advanced research in computer science and software engineering, Vol 4, June 2014.
[8] K.Sakthivel, R.Nallusamy, C.Kavitha, "Image retrieval using fused features", International journal of computer, Information Systems and control engineering Vol:8, No:9, 2014.
[9] Dr.Meenakshi Sharma & Anjali Batra, "Analysis of distance measures in content based image retrieval", global journal of computer science and technology, Vol 14, Issue 2, Year 2014.
[10] Ms.Sonali, B.Maind and Ms.Priyanka Wankar, "Research paper on basic of artificial neural network", International journal on recent and innovation trends in computing and communication, Vol 2 Issue 1 IJRITCC January 2014.
[11] Monika Pathak, Sukhdev Singh, "Implications and emerging trends in digital image processing", International journal of computer science and information technologies, Vol 5, 2014.
[12] Prashant Nair, Dhileep Kumar and Ramesh Shahabadkar, "Content based image retrieval based on Color and Texture features using Hog Descriptor", Global journal of research analysis, Volume-3, Isuue-8, August 2014. ISSN No. 2277-8160.
[13] Pranoti Mane and N.G.Bawane, "Comparative performance evaluation of edge histogram descriptors and color structure descriptors in content based image retrieval", International journal of computer applications, NCIPET-2013.
[14] Vivek Kapur, Dr.P.T.Karule, and Dr. M.M. Raguwanshi, "Content based image retrieval using soft computing technique", 2nd international conference on computational Techniques and artificial Intelligence, March 17-18, 2013 Dubai (UAE).
[15] Mrs.Vrushali Yashwant Erande, Prof. P.R.Badadapure, "Content based image retrieval using neural network", International journal of Scientific & engineering Research, Volume 4, Issue 8, August 2013.Volume 83-No.12, December 2013.
[16] Ahmed Talib, Massudi Mahmuddin, Husniza Husni, Loay E.George, "A weighted dominant color descriptor for content based image retrieval", ELSEVIER, J.VIS.Commun.Image R.24(2013).
[17] Guang Hai-Liu and jing-Yu Yang, "Content based image retrieval using color difference histogram", ELSEVIER, Pattern Recognition 2013.
[18] Arvind Nagathan and Mrs I. Manimozhi, "Content base image retrieval system using feed-forward Back-propagation neural network", International Journal of Computer Science Engineering (IJCSE), Vol 2, july 2013.
[19] A.Ramesh Kumar, D.Saravanan, "Content based image retrieval using color histogram", International journal of computer science and information technologies, Vol 4, 2013.
[20] Dr.Prashant Chatur and pushpanjali chouragade, "A soft computing approach for image searching using visual re-ranking", International journal of computer engineering and Technology, Vol 4, Issue 2, March-April 2013.
[21] Fuhar Joshi, "Analysis of existing CBIR systems: improvements and validation using color features", International journal of Emerging technology and advanced engineering, Vol 3, Issue 5, May 2013.
[22] Manimala Singha and K. Hemachandran, "Content based image retrieval using color and Texture", Signal and Image Processing: An international journal (SIPIJ), Vol 3, No.1, February 2012.
[23] Swapna Borde, Udhav bhosle, "Content based image retrieval using clustering", International journal of computer applications, Volume 60-No.19, December 2012.
[24] S.Manoharan and S.Sathappan, "A comparison and analysis of soft computing techniques for content based image retrieval", International journal of computer applications, Vol 59,December 2012.
[25] Jerome Da Rugna, Gael Chareyron and Hubert Konik, "About Segmentation step in Content based image retrieval", Proceedings of the world congress on engineering and computer science 2011, Vol 1, October 19-21, 2011.
[26] Nidhi Singhai and Prof. Shishir K. Shandilya, "A survey on Content based Image Retrieval System", International journal of computer applications(0975-8887) Vol 4, No.2, July 2010.
[27] Thomas M. Deserno, Sameer Antani and Rodney Long, "Ontology of gaps in Content based image retrieval", Journal of Digital imaging, Vol 22, No.2, April 2009 pp 202-215.
[28] Young Deok Chun, Nam Chul Kim, Ick Hoon Jang, "Content based image retrieval using multi-resolution color and texture features", IEEE Transactions on Multimedia, Vol 10, No.6, October 2008.
Ying Liu, Dengsheng Zhang, Guojun lu, Wei-ying Ma, "A survey of content based image retrieval with high level semantics", ELSEVIER Pattern recognition 2007.

