IMAGE RETRIEVAL BASED ON FEATURE SELECTION METHOD

Lakshmi P.S[1], Thusnavis Bella Mary I[2], A.Vasuki[3], Ansa Saju[4], Kolcharam Yadhul[5]

[1]*PG Scholar, Department of ECE, Karunya University, Tamil Nadu, India*
[2]*Assistant Professor, Department of ECE, Karunya University, Tamil Nadu, India,*
[3]*Professor, Department of ECE, Kumaraguru College of Technology, Tamil Nadu, India*
[4]*PG Scholar, Department of ECE, Karunya University, Tamil Nadu, India*
[5]*PG Scholar, Department of ECE, Karunya University, Tamil Nadu, India*

Abstract

Content-Based Image Retrieval (CBIR) look at retrieval of analogous images from a large database for a given input query image. It has outstretched applications in image processing and pattern recognition. In this paper texture feature is extracted from image using Gray Level Co-occurrence Matrix (GLCM) technique. The technique is used for calculation of statistical measures such as energy, correlation, homogeneity and contrast features. A feature selection technique is brought to select optimal features. Feature selection method based on the Genetic algorithm approach is adapted to improve the accuracy of content- based image retrieval systems. Feature selection iron out the curse of dimensionality reduction. The major advantage of this approach is that little human intrusion is required for retrieving the required images from the database. The method is evaluated on Coral Database. Performance analysis is done by using precision and recall.

Keywords: *Content based image retrieval, Feature Selection, Genetic Algorithm, Gray Level Co-occurrence Matrix, Precision and Recall.*

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

1. INTRODUCTION

Content-Based Image Retrieval (CBIR) point out a method that retrieves images based on their content [1]. Content-based image retrieval, which is employed in miscellaneous areas such as entertainment, art, fashion design, advertising, history, medicine and industry, is a considerable challenge in the fields of pattern recognition and computer vision [2]. In a CBIR system, visual features are used to symbolize images for searching and indexing.

"Curse of dimensionality" is one of the major actuations for feature selection. Numerous features increase computational time without major performance transition in the testing phase. Feature Selection (FS) is a process of selecting only important features from large database to create subset of original dataset and it preserves the original features without altering the feature domains. FS is usually applied as a preprocessing step in data mining tasks by removing irrelevant or redundant features (dealing with the dimensionality curse), by leading to more efficient(reducing the computational cost and the amount of memory required) and accurate classification, clustering and similarity searching process. FS reduces the number of features by removing irrelevant, redundant, or noisy data.

In the course of retrieval or similarity searching process, the images that are most analogous to the query image based on some distance measure are reverberated. Feature extraction produces large number of features which may be irrelevant consequently in order to reduce the dimensionality reduction curse feature selection methods is employed. Feature selection method searches for most relevant feature subsets.

[1] Present a skeleton which employs a GA using an evaluation function contingent on the ranking concept to perform feature selection for CBIR. In [2] a hybrid approach is presented to reduce the semantic gap between low level visual features and high level semantics, over simultaneous feature adaptation and feature selection. In [4] feature selection technique also included to select optimal features not only to maximize the detection rate but also elucidate the computation of image retrieval.

The subsequent subdivisions of this paper are organized as follows: Section 2 describes the methodology. Section 3 shows the experimental results and section 4 concludes the paper.

2. METHODOLOGY

A set of coral natural images are taken and stored in the database. Features are extracted for each image in the database as well as for the query image. Feature selection method is used for selecting the optimal features. Relevant images are retrieved from the database using different similarity measures. The efficiency of the retrieval system is evaluated using the two performance measures, precision and recall and also comparison of different similarity metrics has been done.

The CBIR system consists of seven components namely, (i)Image Database (ii) Pre-processing (iii) Feature extraction (iv) Feature vector block (v) Feature selection (vi) Similarity Measure (vii) Retrieval Block (viii) Performance measures

2.1 Pre-Processing

The original color image is converted to gray scale image. As a pre-processing step histogram equalization is done. It enhances the contrast of values of an image by generating its flat histogram.

2.2 Texture Feature Extraction

Feature extraction is a method of transforming input data into set of features. The various features which can be extracted are color, texture and shape.Texture is an imperative feature of an image. Gray-level co-occurrence matrix is one of the extensively used statistical approaches to extract texture feature of an image, such as energy, contrast, correlation and homogeneity. Gray level co-occurrence matrix (GLCM) typifies the occurrence rate of pixel pair with different gray levels, given the distances between the pixels in different directions respectively.

1) Contrast-Measures the local variations in the gray-level co-occurrence matrix.

$$\sum_{i,j} \left| i - j \right|^2 p(i, j) \qquad (1)$$

2) Correlation- It measures the joint probability occurrence of the particularized pixel pairs

$$\sum_{i,j} \frac{(i - \mu i)(j - \mu j)\, p(i, j)}{\sigma_i \sigma_j} \qquad (2)$$

3) Energy- It gives the sum of squared elements in the GLCM. Energy can also be called as uniformity or the angular second moment.

$$\sum_{i} p(i, j)^2 \qquad (3)$$

4) Homogeneity-It is defined as the closeness of the dissemination of elements in the GLCM to the GLCM diagonal.

$$\sum_{i,j} \frac{p(i, j)}{1 + \left| i - j \right|} \qquad (4)$$

2.3 Texture Feature Selection

Feature selection algorithm is a computational solution to pinpoint relevant features. The feature selection method is to upsurge the precision of similarity searches and significantly decrease the data dimensionality. FS scale down the number of features by removing irrelevant, redundant, and noisy data. The optimal feature is selected using Genetic Algorithm that searches for the best feature subset corresponding to evaluation criteria positioned on the ranking quality. GA performs adaptive searching following the concepts from natural genetics and evaluation based on natural selection.

2.3.1 Genetic Algorithm

The genetic algorithm is based on the fundamental criterion of biological inheritance and evolution. GA works iteratively applying the genetic operations of selection, crossover and mutation to a population of individuals intended at creating better adapted individuals. The genetic algorithm is solved by Rastrigin's fitness function.

2.4 Similarity Measures

The similarity measure is associated with matching features to produce a result that is visually similar. Euclidean distance measure is calculated in this paper. Euclidean distance is used for texture feature analysis. Content-based image retrieval work out the visual similarities between a query image and images in a database. The retrieval result is not a single image but a list of images ranked by their similarities. Different similarity or distance measures will alter retrieval performances of an image retrieval system significantly.

$$\delta_1(x, y) = \left\| x - y \right\|_2 = \sqrt{\sum_{j=1}^{d} (x_j - y_j)^2} \qquad (5)$$

2.5 Algorithm

The retrieval of image for texture is done using following algorithm and the flow of the method is illustrated in the Fig-1 below.

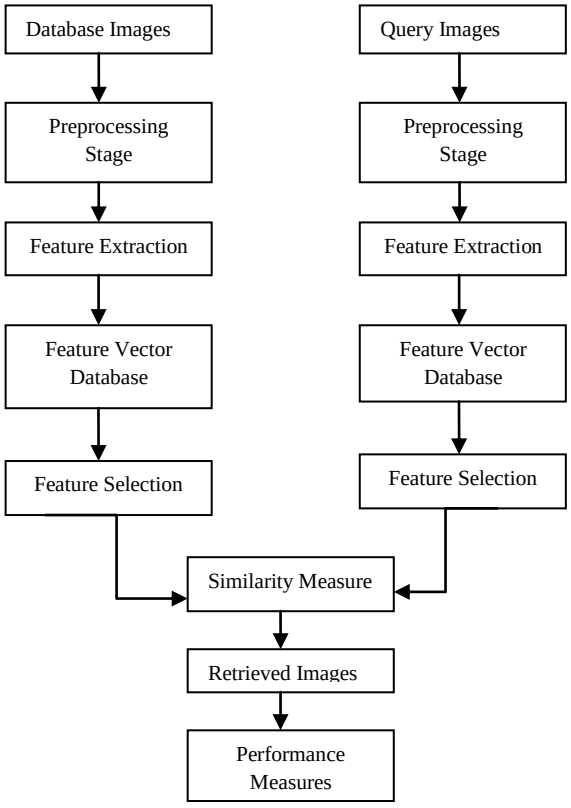

Fig-1: System Overview

Step1: Preprocess the database and target images.
Step2: Calculate the texture features by GLCM method for query image and database.
Step3: For each database images, repeat the above calculation.
Step4: Construct a combined feature vector.
Step5: Finding the best optimized feature value is done using feature selection method (Genetic Algorithm).
Step6: Similarity measurement is calculated.
Step7: Output image is retrieved.
Step8: Performance measure is computed.

3. EXPERIMENTAL RESULTS

The input color image converted into gray scale is preprocessed. The preprocessing is done using histogram equalization and it is illustrated in figure 2. The Feature extracted and feature selected based retrieved images are illustrated by figure 3 and figure 4 respectively.

3.1 Database

The image database chosen are Coral database .The dataset consists of Coral natural images. The images are of size 256x384 or 384x256. In this paper, 4 classes were taken and each class consists of 25 images. The classes are Bus, Dinosaur, Elephant and Food.

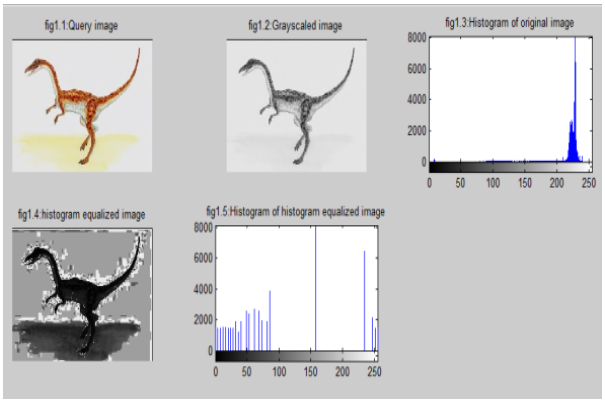

Fig -2: Input query image and preprocessing stage

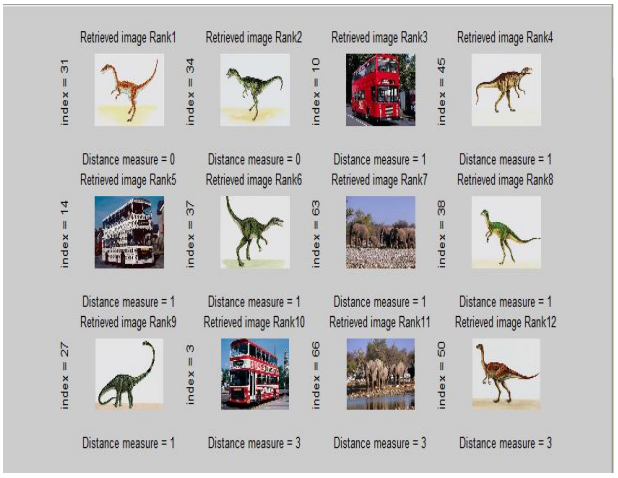

Fig -3: Texture Feature Extraction based Retrieved Images

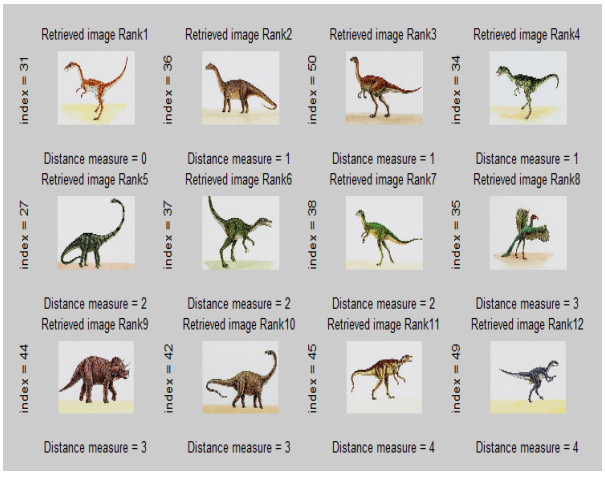

Fig -4: Feature Selection based Retrieved Images

Fig 3, represents the retrieved images based on texture feature extraction. From the retrieved results it can be found out that there are only 7 relevant images in this stage.

Fig 4, represents the retrieved images based on feature selection using Genetic algorithm. By this technique the number of relevant images in the retrieved set has increased. In this method the optimal features are selected. Hence the FS improves the retrieval rate.

3.2 Performance Measurements

For evaluation of divergent algorithms of image retrieval, an effectual performance measurement is necessary. Different performance measurements can be used for valid analysis. The performance of the CBIR system can be evaluated in terms of precision and recall.

3.2.1 Precision

Precision can be defined as the ratio of the number of correct retrieved images to the total number of images in the retrieved dataset. A precision value of 1.0 denotes that each result retrieved by a search was relevant.

$$\textbf{Precision} = \frac{\textbf{Number of relevant images retrieved}}{\textbf{Total number of images retrieved}} \quad (6)$$

The below Chart-1 explains that for every class after the feature selection the precision is improved. The Class II, Dinosaur shows high precision of 100% after the feature selection.

3.2.2 Recall

Recall is the fraction of the number of relevant images selected to the total number of relevant images in the database.

$$\text{Recall} = \frac{\text{Number of Relevant Images Retrieved}}{\text{Total Number of Relevant Images in the Database}} \quad (7)$$

For high retrieval accuracy the recall value should be low. In prominence to the Chart- 2, the recall value varies between 28%-48% during feature selection.

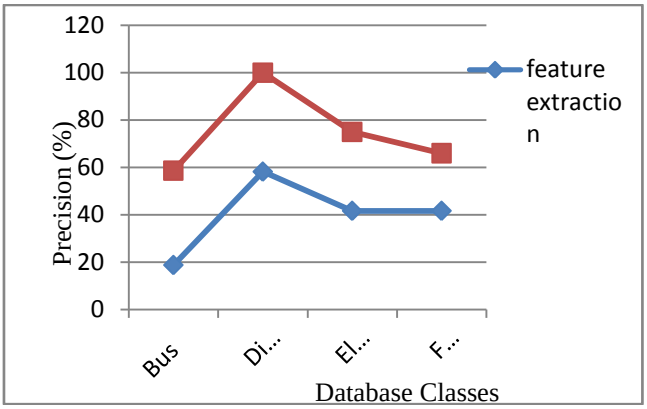

Chart -1: Precision vs. Database Images

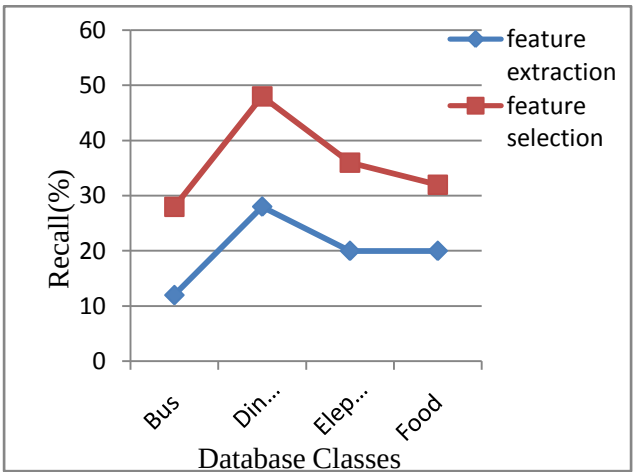

Chart -2: Recall vs. Database Images

Table -1: Comparative analysis of texture feature result based on feature extraction and feature selection

Image Class	Feature Extraction		Feature Selection	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Class I	18.75	12	58.33	28
Class II	58.22	28	100	48
Class III	41.66	20	75	36
Class IV	41.66	20	66	32

The above Table 1 illustrates the comparative analysis of texture feature result based on feature extraction and feature selection. The performance measure analysis is done for various image classes of coral database. From the table 1 the highest precision for feature extraction is 58.22% and after feature selection an improved precision value of 100% is obtained for class II. Recall values are between 12- 28% in feature extraction whereas it is improved to 28-48% by using feature selection method.

4. CONCLUSIONS

In this paper, evaluation of retrieval system is done based on different features for coral database. This paper performs retrieval of different input query images from the image database based on different feature values obtained from texture. Performance is evaluated using two performance measures like precision and recall. The performance of the retrieval system was observed to be better when the image retrieval was done based on texture features. The above technique is performed for Coral database. But retrieval time was observed to be longer. This problem can be eliminated by a process called feature selection. The results of the feature selection technique based on the performance measures showed that higher accuracy of the retrieval system can be obtained in lesser computation time. This work can be extended to various large medical image databases. The results are quite good for most of the query images by using feature selection method and it is possible to further improve by adding relevance feedback.

REFERENCES

[1] Sergio Francisco da Silva, Marcela Xavier Ribeiro,Joao do E.S. Batista Neto ,Caetano Traina -Jr ,Agma J.M. Traina , "Improving the ranking quality of medical image retrieval using a genetic feature selection method" , Decision Support Systems, 2011

[2] Esmat Rashedi, Hossein Nezamabadi –pour, Saeid Saryazdi , "A simultaneous feature adaptation and feature selection method for content- based image retrieval systems ", Knowledge –Based Systems, 2013

[3] Daniele Cerra, Mihai Datcu , "A Fast Compression-based similarity measure with applications to content – based image retrieval",J.Vis.Commun.Image R.,2012

[4] Apostolos Marakakis, Nikolaos Galatsanos, Aristidis Likas,Andreas Stafylopatis, "Relevance Feedback for Content-Based Image Retrieval Using Support Vector Machines and Feature Selection", Springer-Verlag Berlin Heidelberg 2009ICANN 2009, pp. 942–951, 2009.

[5] Chuen-Horng Lin , Rong-Tai Chen , Yung-Kuan Chan , "A smart content-based image retrieval system based on color and texture feature", Image and Vision Computing, 2009.

[6] Daniele Cerra, Mihai Datcu , "A Fast Compression-based similarity measure with applications to content – based image retrieval",J.Vis.Commun.Image R.,2012

[7] Esin Guldogan,Monceff Gabbouj ,"Feature selection for content based image retrieval", Springer,2008

[8] G.Quellec,M.Lamard,G.Cazuguel,B.Cochener,C.Roux, "Wavelet optimization for content –based image retrieval in medical databases",Medical Image Analysis,2010

[9] Heng Qi, Keqiu Li , Yanming Shen, Wenyu Qu , "An effective solution for trademark image retrieval by combining shape description and feature matching ", Pattern Recognition, 2010.

[10] Jennifer G. Dy,Carla E. Brodley,Avi Kak, Lynn S. Broderick, Alex M. Aisen, "Unsupervised Feature Selection applied to Content Based Retrieval of Lung Images", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 25, No. 3, March 2003

[11] 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

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