Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (P) Vol.6 No.1, 2017, pp.1-5

DOI: https://doi.org/10.51983/ajcst-2017.6.1.1779

A New Architecture for Image Retrieval Optimization with HARP Algorithm

S.Selvam[1] and S.Thabasu Kannan[2]
[1] Research Scholar, Dept. of Computer Science,
Bharathiar University, Coimbatore, Tamilnadu, India.
[2]Principal, Pannai College of Engg. & Tech, Sivagangai – 630 561, Tamil Nadu, India
E-mail: s.selvammscmphil@gmail.com
(Received 18 December 2016; Revised 30 December 2016; Accepted 23 January 2017; Available online 30 January 2017)

Abstract **- Image retrieval plays a vital role in image processing.The main aim of this paper is to build more generalized CBIR system, which increase the searching ability and to improve theretrieval accuracy. The proposed method is experimented and analyzed with large database. The result show that the architecture of new CBIR system is good performance in speed and decreasing retrieval time**.

Keyword: **CBIR, Genetic Algorithm, HARP Algorithm, Precision, Recall.**

I. INTRODUCTION

In contemporary world the rapid development in personal Computer, institutional computer and internet necessitates need to find a particular image or group of image. Retrieval images using their visual features such as color, shape and texture.

The Content Based Image similarity measurement algorithms, if chosen correctly for a particular image representation techniques will definitely increase the efficiency and effectiveness retrieval of data. In a CBIR system, the retrieval of images has been done by similarity comparison between the query image and all candidate images in the database. To find the similarity between two images, the method is to measure the distance between the feature vectors representing in the two images. Also find more similarity or relative images, the heuristic approach based Genetic Algorithm has been used in the CBIR system.

Image retrieval techniques are color, texture and shape. This method is applied to get an image from the image database. The main aim of the paper is to reduce the calculating time and user interaction. The conventional CBIR system display the huge amount of result at the end of the process .This will drove the user to spend large amount of time to analyze the output images.

In the proposed system evaluates textual features and color features for find the similarity between query and database images. This approach will reduce the output result to some levels based on the user thresholds value.

The step of this study is to reduce semantic gap between high level concept and low level features. CBIR calculates the similarity between the input query image and database images. Since there might be chances for unwanted result at the end of the retrieval process. By using HARP cluster, the output images and selects one representative image.

The subsequent process is to evaluate their performance with regard to speed and accuracy. We can use any clustering algorithm instead of relevance feedback. It is based on the features extracted from the image themselves and also allocates those images into the nearest cluster. The algorithm computes and allocates until there is little bit variation in the movement of feature points in each cluster. Clustering is used to classification of patterns into groups. Its main work is to assigning a set of objects into groups, so that the objects in the same cluster are more similar to each other than the to those in other cluster.

In this paper, Color, Texture and Shape features were extracted and combined to form feature vector of image. For color features, the moments of the color distribution were calculated from the images and used as color descriptor. For texture features, we used Gabor filter, which is a powerful texture extraction technique in describing the content of image. For shape features, edge histogram features that include five categories were used as shape descriptor. These three descriptors were combined and optimized using GA with H A R P c l u s t e r i n g accuracy as a fitness function to select optimum weights of features. We performed GA with H A R P c l u s t e r i n g on the database as an offline step, and the system does not need to search the entire database images; instead just a number of candidate images are required to be searched for image similarity.

II. PREVIOUS STUDY

There are various approaches are present for CBIR. Some of the important literature which covers the more important CBIR System is discussed below.

Chin-Chin Lai et.al. have proposed an *interactive genetic algorithm* (IGA) to reduce the gap between the retrieval

results and the users' expectation .They have used color attributes like the mean value, standard deviation, and image bitmap.They have also used texture features like the entropy based on the gray level co-occurrence matrix and the edge histogram.

Zhang Xu-bo *et.al.* have published a paper on *improved K-means clustering and relevance feedback* to re-rank the search result in order to remedy the rank inversion problem in CBIR. Experimental results show that the reranking algorithm achieves a more rational ranking of retrieval results and it is superior to Reranking via partial Grouping method

Lijun Zhao *et.al* .have proposed a *multi-round relevance feedback* (RF) strategy based on both support vector machine (SVM) and feature similarity to reduce the gap between query and retrieve result.

SharadhRamaswamy *et.al.*have published a paper on a fast *clustering-based indexing technique*. In this method relevant clusters are retrieved till the exact nearest neighbors are found. This enables efficient clustering with low preprocessing storage and computation costs.

Nhu-Van Nguyen *et.al.* have proposed *Clustering and Image Mining Technique* for fast Retrieval of Images. The main objective of the image mining is to remove the data loss and extracting the meaningful information to the human expected needs. The clustering-repeat gives good result when the number of examples of feedback is small.

Hua Yuan *et.al.* have presented a new *statistical model-based image feature extraction* method in the wavelet domain and a novel Kullback divergence-based similarity measure. The Gaussian Mixture Models(GMM) and Generalized GMM are presented to help extract new image features.

From the literature survey it is concluded that a wide variety of CBIR algorithms have been proposed in different papers. The selection feature is one of the important aspects of Image Retrieval System to better capture user's intention. It will display the images from database which are the more interest to the user.

III. ARCHITECTURE OF NEW CBIR SYSTEM

Training Image Input : The learning phase tells about the training process which a huge amount sample images are input in the first step. The genetic algorithm is used to train the features with different weights. For optimizing the feature weights and for fitness function, HARP algorithm is used. The training part outputs the classifying result and stores it in the feature database. All these steps performed offline and each class will be indexing along with its associated classID in the index files.

Feature Extraction (Image Signature): There are various kinds of visual features to represent an image, such as color, texture, shape, and spatial relationship. Since one type of features can only represent part of the image properties, a lot of work done on the combination of these features. The feature of each image is very much smaller in size compared to the image data, so the feature database contains an abstraction of the images in the image database.

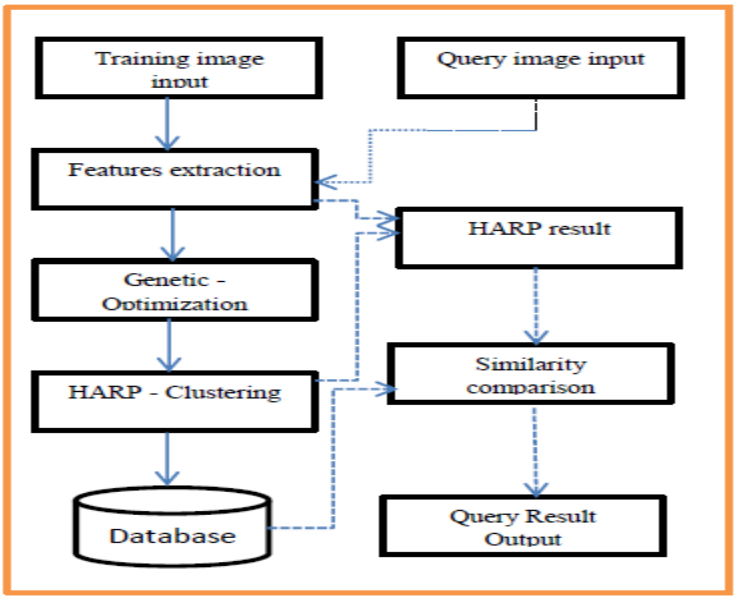

Fig.1 A New architecture for CBIR

Colors: are defined on a selected color space. Varieties of color spaces include, RGB, LAB, LUV, HSV (HSL), YCrCb and the HMMD. Common color features or descriptors in CBIR systems include color-covariance

matrix, color histogram, color moments and color coherence vector storing, filtering and retrieving audiovisual data. The emerging MPEG-7 is a new multimedia standard, which has improved CBIR by providing a rich set of standardized descriptors and description schemas for describing multimedia content. MPEG-7 has included dominant color, color structure, scalable color, and color layout as color features. Here we used Color Structure Descriptor (CSD) as color feature. The CSD represents an image by both the color distribution of the image or image region and the local spatial structure of the color. CSD used a 8 × 8 structure to scan the total image. This descriptor counts the number of times a particular color is contained within the structuring element while the image or image region is scanned by this structuring element. It has used HMMD color space.

Texture: There exist different approaches to extract and represent textures. They can be classified into space-based, frequency-based models, and texture signatures. Some popular techniques i.e wavelet transform, co-occurrence matrix, and Gabor filters are applied to express texture features for image.

Shape: It is seen that natural objects are primarily recognized by their shape. Two main types of shape feature are commonly used; *global features* such as aspect ratio, circularity and moment invariants and *local features* such as sets of consecutive boundary segments.

A. Genetic Algorithm–for Optimization*:*

It is used to find optimization solution to complicated problems. It is also known as heuristic searching method. There are three operators such as selection, mutation and crossover. These concepts are used in the new architecture of CBIR systems. This improves the good searching ability and gave the optimization solution. It contains the following operators:

Selection: In the each successive iteration existing input iterations are refreshed through a fitness process. If any iteration is best fit, then it is the solution and no further selection is needed.

Mutation is occurs during evolution according to user definable mutation probability. This probability is set low.

Cross over is a process n-parent solutions used to derive a child solution.

B. HARP - A Clustering Algorithm: The algorithm is based on bottom-up manner. Initially put each element in its own cluster. Among all current cluster on the basis of smallest distance by merging all the selected and related data on various clusters. In HARP algorithm, the accuracy level of clustering is more by using relevance indexing and merge score. The scalability level is also very high. The time taken for finding the closest cluster is very less.

Database: A database containing number of images with any one of the formats such as .bmp, .jpg, .tiff. is required.

Query: The user provides a sample image or sketched figure as the query for the system. This phase describes the images searching process. The user enters a query image for which the system extracts color, texture and shape features the features vectors of database images are previously extracted and stored.

Similarity Matching*:* Using the similarity metrics defined for color, texture and shape, the similarity distances between the query image and the centroid image of each class are calculated. The smallest distance (most similar) will determine to which the image belongs. The class with the smallest distance is returned and the images in this class will be compared with the query image.

Retrieval*:* The most matching images will be retrieved and then they are sorted in ascending order. The first *N* similar target images with smallest distance value to the query are retrieved and shown to the user.

IV. PERFORMANCE EVALUATION

We select to test our system, to compare the new system results with some other existing CBIR systems. The images database that we used in our evaluation. It is a sub set of the Corel database of 1,000 images in JPEG format.1,000 image database went through our implemented system to extract the features and stored them. The extracted features are weighted by GA and they are used for classification by using the HARP algorithm (Hierarchical Approach with Automatic Relevant dimension selection for projected clustering). The level of retrieval accuracy is a factor to influence the performance. In CBIR, the most commonly used performance measures are *Precision* and *Recall*. Precision is defined as the ratio of the number of retrieved relevant images to the total number of retrieved images. This means that precision measures the accuracy of the retrieval. Recall is defined as the ratio of the number of retrieved relevant images to the total number of relevant images in the database. The recall measures the robustness of the retrieval.
In CBIR, if the precision score is 1.0 then every image retrieved by a search is Relevant. If the recall score is 1.0 then all relevant images are retrieved by the search is robust. We evaluate the new system by using two metrics viz: the *Retrieval Effectiveness* and the *Retrieval Efficiency.*

a. ***Retrieval Effectiveness:*** A retrieved image is considered a match if it is in the same class as the query image. The system works well and it retrieves better results over the randomly selected images as queries by using GA and HARP algorithm.
b. ***Retrieval Efficiency*****:** By assigning different weights to each feature to improve the efficiency we have used GA with a HARP algorithm to select optimum weights of features to get the accuracy.

Here by using clustering pre-process of the database image via HARP algorithm decreases the average query response time, the similarity search time for image matching and increases the efficiency of the system.

V. COMPARISON OF THE NEW SYSTEM WITH OTHER EXISTING SYSTEMS

.

For each class in the database, we randomly selected 20images as queries. Since we have 3 classes in the database, we have 60 query images. For each query, we calculate the precision and recall of the retrieval. The average precisions and the average recall for each class based on the returned top 20 images were recorded. Moreover the new system result is compared against the performance of three methods.

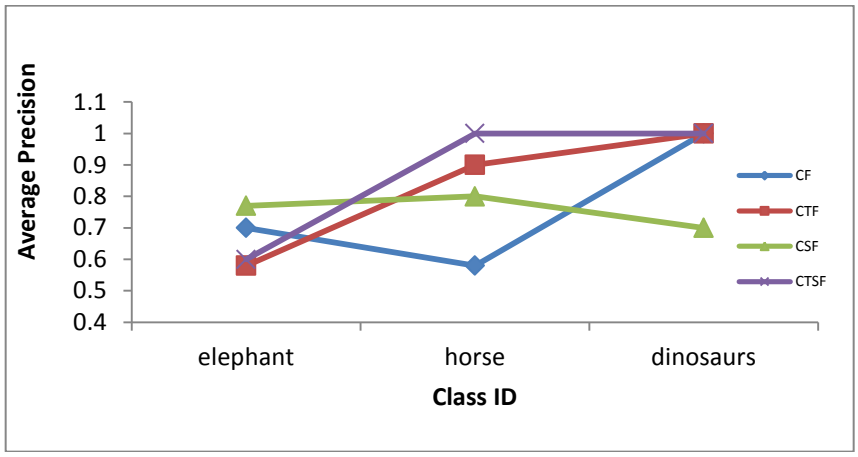

Fig. 2 Comparison of Precision of the New System with various types of features.

The below figure shows that the new system performs significantly better in all three classes except elephant class. This result confirms that a fusion of multiple feature scan increase the performance of the system.

The below figure shows that the new system performs significantly better than other existing systems for all classes except elephant class. This is a good indicator for the effectiveness of our system. The reason behind the limitation in two classes is that those classes' images are very similar in term of the dominant color, texture and shape so, our new system may confused between them.

.

The above figure shows that the new system performs significantly better than other systems, for all classes. This means that the new system can retrieve most of database images that match query image. The new system works well in the classification part of using GA with HARP algorithm. The average precision and the average recall increased from 79.2% to 89.1% and 52.3% to 69.8% respectively and obtained an average reduction in 6.18 seconds

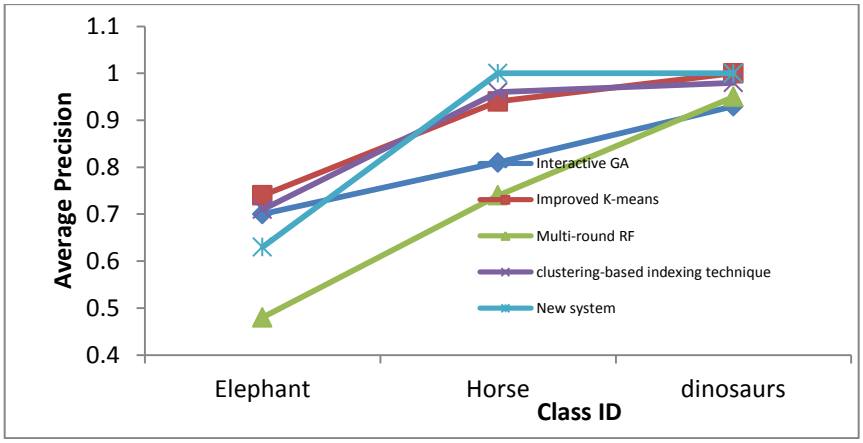

Fig.3 Comparison of Precision of the New System with some Existing Systems

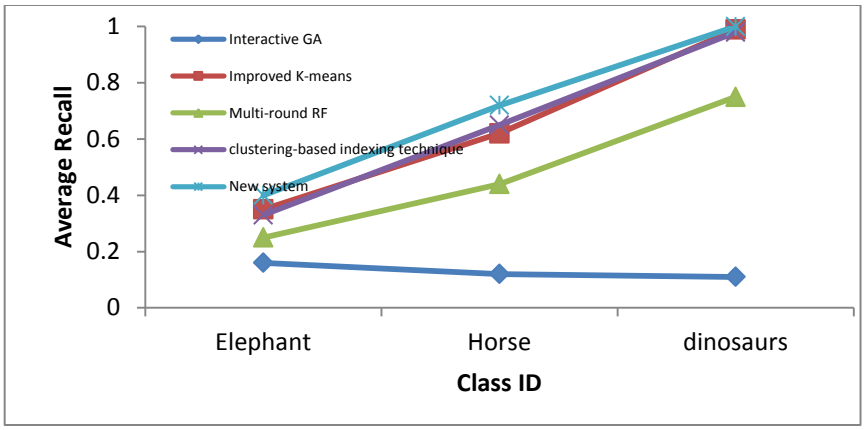

Fig.4 Comparison of Recall of the New System with some Existing S ystems

VI. CONCLUSION

The growth of image data leads to the need of research and development of Image Retrieval. CBIR is currently a keen area of research in the area of multimedia databases. Various research works had been undertaken in the past decade to design efficient image retrieval techniques from the image databases. More précised retrieval techniques are needed to access the large image archives being generated, for finding relatively similar images. In this work the GA is combined with HARP clustering algorithm to improve the retrieval accuracy of the system. Getting lower computational time and retrieving relevant and accurate image is possible by using CBIR. In future we have a proposal to disseminate the features selections and use other distance measures to improve the overall results.

The efficiency of the new system is improved by considering candidate images for similarity computation purpose i.e. not considering the whole database images. A candidate image lies in the same cluster with the query image the benefit of the clustering process clearly proved the retrieval accuracy.

REFERENCES

[1] I.El-Naqa, Y.Yang, N.Galatsanos, R.Nishikawa and M.Wernick, "A Similarity Learning Approach to Content-Based Image Retrieval: Application to Digital Mammography," *IEEE Transactions on Medical Imaging*, 2009.

[2] R.Datta,J.Li,andJ.Wang,"Content-asedImageRetrieval-ApproachesandTrends of theNew Age,"*ACMComputing Surveys*, vol. 40, no. 2, pp. 1-60,April 2008.

[3] S.Selvam and Dr.S.Thabasu Kannan,"Design of an Effective Method for Image Retrieval", published IJIRAE, International Journal of Innovative Research in Advanced Engineering, Volume-1, March 2014, pp.51-56.

[4] V.Gudivada and V.Raghavan, "Content-based Image Retrieval Systems," *IEEE Computer*, vol. 28, no 9, pp18-22, Sep. 1995.

[5] R.Chang,J.Ho,S.Lin,C.FannandY.Wang,"ANovelContentBasedImageRetrievalSystemusingK-means with Feature Extraction, "*International Conference on Systems and Informatics*,2012.

[6] B.WANG,X.ZHANG,andN.LI,"Relevance Feedback Technique For Content-Based Image Retrieval Using Neural Network Learning, "*Proceedings of the Fifth International Conference on Machine Learning and Cybernetics*, Dalian, 2006.

[7] P.JeyanthiandV.JawaharSenthilKumar, "Image Classification by K-means Clustering. "*Advances in Computational Sciences and Technology*, 2010.

[8] F.Long,H.Zhang,H.Dagan,andD.Feng, "Fundamentals of Content Based Image Retrieval," *Multimedia Signal Processing Book*, Chapter1, Springer-Verlag, Berlin Heidelberg New York, 2003.

[9] J.HanandM.Kambr,"DataMiningConceptsandTechniques,"2ndEd., MorganKaufmannPublisher, 2006.

[10] S.Selvam and Dr.S.ThabasuKannan, "An Empirical Review on Image Retrieval System by using Relevance Feedback" proceeding of International Symposium on "Research innovation for quality improvement in Higher Education" conducted by Bharathiar University, Coimbatore, October 2014 and published in "Research and Trends in Data mining and Image Processing Technologies and Applications", Bloomsbury publishing India, London, New Delhi, New York, Sydney.