ISSN:0975-9646

Extracting More Relevant Features from Color Image for Enhancing the Image Retrieval Using Genetic Algorithm

Aruna Bajpai
Assistant Professor,
Department of Computer Science Engineering
Institute of Technology & Management, Gwalior

Abstract— **soft computing and machine learning techniques are help to find the optimum patterns from raw data and utilize with different application to solve the real world problem. Thus the manner and the aspect of learning are changed as the application and their orientation is changed. In this presented paper we discuss a data mining technique which is used to find the relevant data more specifically the image data. The proposed model of image retrieval is based on the contents of the image therefore a review of the CBIR is prepared first and using the observed concepts a new method is tried to develop for enhancing the CBIR for obtaining the more relevant image from huge databases. The implementation and obtained performance of the proposed system demonstrate the efficiency and effective precision and recall rate during the classification analysis of the data. Additionally that improves the search time of the images using the query by image context.**

Keywords— **CBIR, genetic algorithm, feature extraction, image data analysis, performance evaluation**

I. INTRODUCTION

Data mining is a technique that is used to extract knowledge from the data. By nature data is found in structured format or in unstructured format therefore the different learning approaches are implemented to find the mathematical correlation among data attributes. These estimated attributes are help to evaluate the patterns for classification, prediction and other applications [1]. Among various applications of data mining the information retrieval is an essential application domain. Using this approach a number of web clients are able to search the valuable data from knowledge base.

In this presented work the utilization of data mining technique is performed with the content based image retrieval technique. The content based image retrieval is one of the most popular technique is information retrieval. In this approach the image features or contents are recovered and these features are used as the content of image. The image data is a kind of numerical matrix contains some values for representing the real world objects [2]. Thus the retrieval process is deferent from the text data or other formats of data. In this presented work the content based image retrieval process is studied. The image contains the real world objects in form of matrix thus to identify the hidden information in the retrieval process three key features are utilized for discovering the data patterns in image data. The key features include the shape, texture and color distribution in image data. All these features are associated in the image to recognize the images more accurately.

An image retrieval system can be defined as searching, browsing, and retrieving images from massive databases consisting of digital images. Although Conventional and common techniques of retrieving images make use of adding metadata namely captioning keywords so as to perform annotation of words. However image search can be described by dedicated technique of search which is mostly used to find images. For searching images user provides the query image and the system returns the image similar to that of query image [3].

Image Retrieval has been adopted in most of the major search engines, including Google, Yahoo!, Bing, etc. A large number of image search engines mainly employ the surrounding texts around the images and the image names to index the images. Because there are only two main places where anyone can place text first in title (Name of image) and second in the tags which are proposed and implemented using web 2.0 concepts? Most of the time user make query in the text format for search contents over any search engine.

However, this limits the capability of the search engines in retrieving the semantically related images using a given query. On the other hand, the current contemporary in content-based image retrieval is rolling; it has not yet succeeded in bridging the semantic gap between human concepts, e.g., keyword-based queries, and low-level visual features that are extracted from the images. Hence, it has become an urgent need for developing novel and effective methods and techniques by which we fill the gap between the image contents and user query that go beyond these conventional approaches or retrieval models.

There are three methods for image retrieval: text-based method, content-based method and hybrid method. This section explains in details each method. Image retrieval system can be classified as:

- Text based Image retrieval system
- Content Based Image retrieval system

II. LITERATURE SURVEY

This section provides the understanding about the recently developed approaches and techniques for improving the content based image retrieval process.

The CBIR tends to index and retrieve images based on their visual content. CBIR avoid many evils connected with conservative ways of retrieve images by keywords. Thus, a rising interest in the area of CBIR has been known in current years. The arrangement of a CBIR system mainly depends on the particular image illustration and similarity matching function working. The CBIR tends to index and retrieve images based on their visual content. CBIR avoids many problems associated with traditional ways of retrieving images by keywords. Thus, a growing interest in the area of CBIR has been established in recent years. The performance of a CBIR system mainly depends on the particular image representation and similarity matching function employed. So a new CBIR system is proposed by ***Sandeep Singh et al [4]*** which will provide accurate results as compare to previous developed systems. Soft system will be used in this system. Based Image recovery system which evaluates the similarity of each image in its data accumulate to a query image in terms of various visual features and return the image with desired range of similarity. To develop and put into practice an efficient feature extraction NN and SVM to extract features according to data set using Auto calculate the feature weight by neural network. The precision and recall graph in gui according to the retrieved contents of the images from the datasets. To Apply back propagation or feed forward algorithm for neural network classification and calculate cross relationship and apply weakening model for feature matching.

The emergence of cloud data centers enhances the capability of online data storage. Since massive data is stored in data centers, it is necessary to effectively locate and access interest data in such a distributed system. However, traditional search techniques only allow users to search images over exact-match keywords through a centralized index. These techniques cannot satisfy the requirements of content based image retrieval (CBIR). In this paper, ***Jianxin Liao et al [5]*** propose a scalable image retrieval framework which can efficiently support content similarity search and semantic search in the distributed environment. Its key idea is to integrate image feature vectors into distributed hash tables (DHTs) by exploiting the property of locality sensitive hashing (LSH). Thus, images with similar content are most likely gathered into the same node without the knowledge of any global information. For searching semantically close images, the relevance feedback is adopted in our system to overcome the gap between low-level features and high-level features. Author's show that given approach produces high recall rate with good load balancing and only require a few number of hops.

Hanwang Zhang et al [6] present a novel attribute-augmented semantic hierarchy (A2SH) and demonstrate its effectiveness in bridging both the semantic and intention gaps in content-based image retrieval (CBIR). A2SH organizes semantic concepts into multiple semantic levels and augments each concept with a set of related attributes. The attributes are used to describe the multiple facets of the concept and act as the intermediate bridge connecting the concept and low-level visual content. A hierarchical semantic similarity function is learned to characterize the semantic similarities among images for retrieval. To better capture user search intent, a hybrid feedback mechanism is developed, which collects hybrid feedback on attributes and images. This feedback is then used to refine the search results based on A2SH. They use A2SH as a basis to develop a unified content-based image retrieval system. We conduct extensive experiments on a large-scale dataset of over one million Web images. Experimental results show that the proposed A2SH can characterize the semantic affinities among images accurately and can shape user search intent quickly, leading to more accurate search results as compared to state-of-the-art CBIR solutions.

Effective and efficient image retrieval techniques have become the fast and active research area because of explosive use of digital images. User interaction in CBIR system consists of a query formation. The user has problems in declaration of a query with different schemes which has been introduced in literature. In this paper, ***Sunkari Madhu [7]*** analyze the two retrieval method, query by texture and query by color .Texture features involves the invariant histogram characteristics to retrieve the images and color features carry the color histogram in RGB color space to retrieve the images. It is observed from the experimental results, that the query by texture is more effective than the query by color for retrieving the general images.

Automatic analysis of histopathological images has been widely utilized leveraging computational image-processing methods and modern machine learning techniques. Both computer-aided diagnosis (CAD) and content-based image-retrieval (CBIR) systems have been successfully developed for diagnosis, disease detection, and decision support in this area. Recently, with the ever-increasing amount of annotated medical data, large-scale and data-driven methods have emerged to offer a promise of bridging the semantic gap between images and diagnostic information. In this paper, ***Xiaofan Zhang et al [8]*** focus on developing scalable image-retrieval techniques to cope intelligently with massive histopathological images. Specifically, they present a supervised kernel hashing technique which leverages a small amount of supervised information in learning to compress a 10 000-dimensional image feature vector into only tens of binary bits with the informative signatures preserved. These binary codes are then indexed into a hash table that enables real-time retrieval of images in a large database. Critically, the supervised information is employed to bridge the semantic gap between low-level image features and high-level diagnostic information. Author build a scalable image-retrieval framework based on the supervised hashing technique and validate its performance on several thousand histopathological images acquired from breast microscopic tissues. Extensive evaluations are carried out in terms of image classification

(i.e., benign versus actionable categorization) and retrieval tests. Framework achieves about 88.1% classification accuracy as well as promising time efficiency. For example, the framework can execute around 800 queries in only 0.01 s, comparing favourably with other commonly used dimensionality reduction and feature selection methods.

III. PROPOSED WORK

The proposed work is indented to design more accurate image retrieval technique for extracting the image from large databases. Therefore system architecture is presented in this section that utilizes the different techniques and consumes to find more accurate images. The given figure 1 shows the proposed architecture of the system.

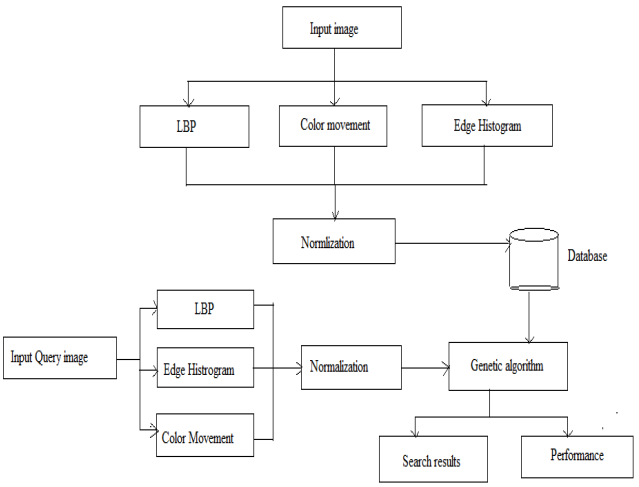

Figure 1 system architecture

The proposed technique contains a number of sub-components to organize the entire efficient and accurate image search system. In order to implement the system is implemented in two major modules both modules are described as:

1. Training module

During the training of the system the process is made to find the essential features from the input training and store into a database for utilizing with the search system. In first the *input image* is produced to the system and their feature vectors are extracted by evaluation of the image. In first the *LBP (local binary pattern)* method is utilized that provides texture analysis of the input image and returns a vector of values. In the next the same image is analysed using the *grid color movement* analysis that provides the color feature of the input training image. Finally the *edge histogram* method is utilized with the input image. That provides the edge feature of image. All the feature vectors are then *normalized* to obtain a common feature from the input image. This common feature is developed by combining the extracted features of image and preserved into the database.

2. Testing module

For retrieving the image from the data base this module is prepared. In this phase the query image is produced to the system and their features are extracted as defined in the training module therefore the edge feature, color movement analysis and the local binary pattern is estimated from the image. These features are normalized first and produced into the genetic algorithm. The genetic algorithm is a kind of heuristic search technique which promises to provide the accurate search results. Therefore in this search technique the query image features are produced as the input query and the population is used as the database features.

For finding the more similar images the Euclidian distance of features are used as the fitness function for evaluating the target outcomes. The proposed method then returns the images and the performance of the system. According to the different experiments the performance of the image retrieval system is estimated and reported in the next section.

IV. RESULTS ANALYSIS

The implemented enhanced image retrieval technique is evaluated on the basis of the different experimental scenarios and different sets of data. The evaluated performance of the obtained system is described in this chapter with their evaluation and outcomes.

A Precision

In any data retrieval or search applications the precision is a fraction of search results which is most relevant to the input query. The provided precision of the proposed content based image retrieval system are given using figure 2.

Figure 2 precision rate

This can be evaluated using the user feedback basis and can be evaluated by the following formula.

$$precision = \frac{relevent\ document\ \cap\ retrieved\ documets}{retrieved\ documents}$$

The precision rate of the implemented system is described in the figure 2, the computed precision values are demonstrated using the Y axis of the given figure and the X axis of the figure shows the amount of training images in the database. According to the obtained results the performance of the proposed system is increases as the amount of data in database is increases. In addition of the precision rate is growing continuously as the similar kinds of images are also increases in data base.

B. Recall

In data retrieval application or the search application recall values are measured for accuracy measurement in terms of relevant document retrieved or relevant data obtained according to the input user query. This can be evaluated using the following formula.

$$recall = \frac{relevant\ doucment\ \cap retrieved\ documents}{relevant\ documents}$$

The figure 3 shows the recall values of the proposed image retrieval application. In order to represent the performance of the proposed image retrieval system the X axis contains the amount of images in database and the Y axis reports the obtained recall rate of the implemented system. According to the obtained results the performance of the proposed system is enhances as the amount of data is increases in the database. The retrieval accuracy with the increasing amount of data is also increases thus the proposed concept is adoptable for the image search applications.

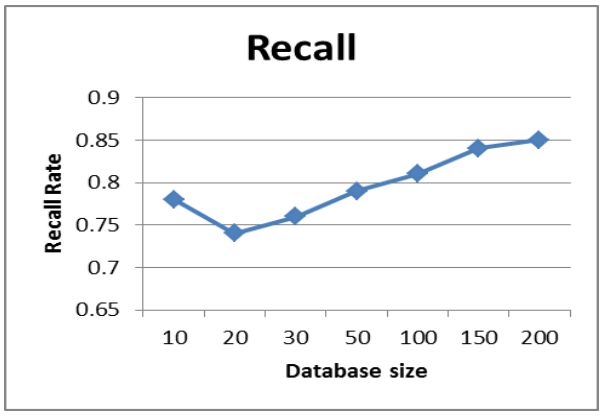

Figure 3 recall rate

C. F-measures

The f-measures of the system demonstrate the fluctuation in the computed performance in terms of precision and recall rates. The f-measures of the system can be approximated using the following formula.

$$F - measures = 2.\frac{precision\ X recall}{Precision + recall}$$

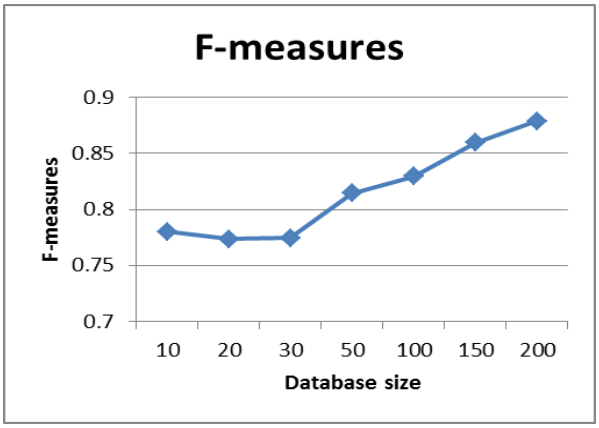

Figure 4 f-measures

The figure 4 shows the performance of the system in terms of f-measures. To demonstrate the performance of the system the X axis shows the amount of data is placed in storage during experiments and the Y axis shows the obtained performance in terms of f-measures. According to the obtained results the performance of the system is much stable and enhancing in progressive manner as the amount of data base is increases. Thus the obtained results are adoptable and efficient for the image retrieval applications.

D. Memory used

The memory used sometimes also called the memory consumption or the space complexity. That amount of main memory required to execute a given algorithm with the amount of data is known as the memory consumption or space complexity of algorithm.

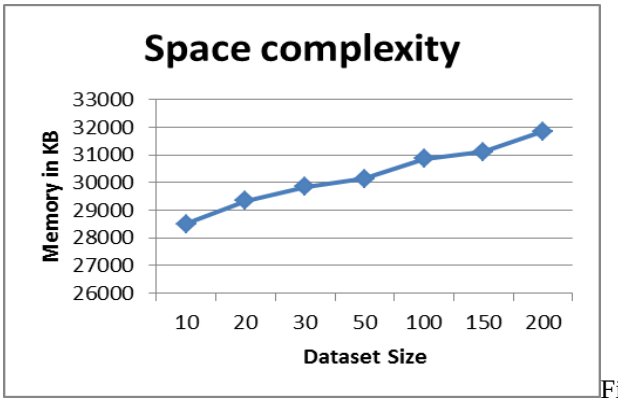

Figure 5 space complexity

The figure 5 shows the performance of the system in terms of space complexity, in this diagram the X axis shows the amount of data available in data base and the Y axis shows the amount of memory consumed in terms of KB (kilo bytes). According to the obtained results the performance of the system becomes consistent and not consuming more memory even when the amount of data to be process is increases in the database but that produces a small amount of effect in memory consumption.

E. Time consumption

The amount of time required to complete the retrieval task after providing input to the system is termed as time consumption of the algorithm. The time consumption of the proposed technique is given using figure 6.

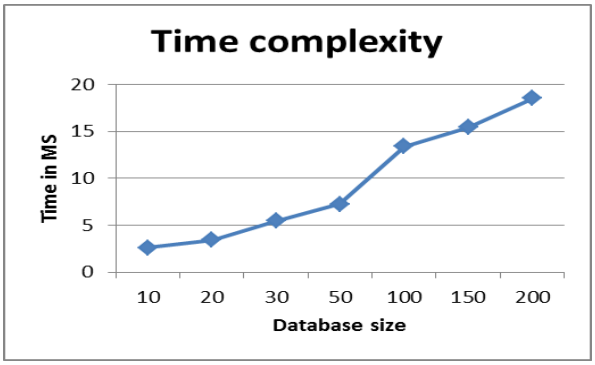

Figure 6 time complexity

According to the demonstrated results the X axis contains the amount of images available in the database and the Y axis shows the amount of time consumed during the retrieval process in terms of milliseconds. According to the obtained results the performance of the system is fluctuating with the amount of data produced in the data base thus as the amount of data is increases the amount of comparison time is increases. Therefore the outcomes of the retrieval system take long time as the amount of data in database is increases.

V. CONCLUSIONS

In this presented paper the main aim is to enhance the technique of image retrieval. Therefore the genetic algorithm based search algorithm is applied on the data for retrieving the images. Therefore a brief literature survey is performed first and the recent development on the CBIR technique is reported in further the using the suitable modeling a new retrieval system is developed and implemented. For finding the significance form the image the color grid movement, edge histogram and the local binary pattern is analyzed and using the image query the genetic algorithm based search system is implemented. The implemented search system returns the promising search outcomes which is demonstrated using the precision rate, recall, and f-measures that demonstrate the accurate search results. Additionally the performance of the search is also provided with the help of space and time complexity. According to the performance outcomes the proposed system is adoptable and efficient thus the given model is providing more accurate results with less resource consumption.

REFERENCES

[1] Jiawei Han and Micheline Kamber, “Data mining Concepts and Techniques”, http://akademik.maltepe.edu.tr/~kadirerdem/772s_Data.Mining.Concepts.and.Technique-s. 2nd.Ed.pdf

[2] P. S. Hiremath and Jagadeesh Pujari. “Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement”, International Journal of Computer Science and Security, Volume (1) : Issue (4)

[3] Minsu Cho, Jungmin Lee, and Kyoung Mu Lee,“Reweighted Random Walks for Graph Matching”, Computer Vision–ECCV 2010, 2010 – Springer

[4] Sandeep Singh, Er.Rachna Rajput, “Content Based Image Retrieval using SVM, NN and KNN Classification”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 6, June 2015

[5] Jianxin Liao, Di Yang, Tonghong Li, Jingyu Wang, Qi Qi, Xiaomin Zhu, “A scalable approach for content based image retrieval in cloud data center”, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, P.O. Box 296, Beijing 100876, China

[6] Hanwang Zhang, Zheng-Jun Zha, Yang Yang, Shuizheng Yan, Yue Gao, and Tat-Seng Chua, “Attribute-Augmented Semantic Hierarchy: Towards a Unified Framework for Content-Based Image Retrieval”, ACM Trans. Multimedia Comput. Commun. Appl. 11, 1s, Article 21 (September 2014), 21 pages.

[7] Sunkari Madhu, “Content based Image Retrieval: A Quantitative Comparison between Query by Color and Query by Texture”, Journal of Industrial and Intelligent Information Vol. 2, No. 2, June 2014

[8] Xiaofan Zhang, Wei Liu, Murat Dundar, Sunil Badve, and Shaoting Zhang, “Towards Large-Scale Histopathological Image Analysis: Hashing-Based Image Retrieval”, IEEE Transactions On Medical Imaging, Vol. 34, No. 2, February 2015