Research Article Impact Factor: 0.621 ISSN: 2319-507X
Ankur Mahalle, IJPRET, 2014; Volume 2 (9): 68-74 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

IMAGE DATABASE CLUSTERING AND RETRIEVAL USING IMAGE MINING TECHNIQUES

ANKUR S. MAHALLE[1], PROF. M. A. PUND[2]

1. Research Scholar, Department of Information Technology, PRMIT&R, Badnera, Maharashtra, India
2. Assistant Professor, Department of Computer Science And Engineering, PRMIT&R, Badnera, Maharashtra, India

Accepted Date: 27/02/2014 ; Published Date: 01/05/2014

Abstract: Image plays vital role in every aspect of business such as business images, satellite images, medical images and so on. If we analysis these data, which can reveal useful information to the human users. Image retrieval is the fast growing and challenging research area with regard to both still and moving images. Many Content Based Image Retrieval (CBIR) system prototypes have been proposed and few are used as commercial systems. CBIR aims at searching image databases for specific images that are similar to a given query image. It also focuses at developing new techniques that support effective searching and browsing of large digital image libraries based on automatically derived imagery features. It is a rapidly expanding research area situated at the intersection of databases, information retrieval, and computer vision. The CBIR focuses on Image 'features' to enable the query and have been the recent focus of studies of image databases. The features further can be classified as low-level and high-level features. Users can query example images based on these features such as texture, colour, shape, region and others. By similarity comparison the target image from the image repository is retrieved. Meanwhile, the next important phase today is focused on clustering techniques. Clustering algorithms can offer superior organization of multidimensional data for effective retrieval. Clustering algorithms allow a nearest neighbour search to be efficiently performed.

Keywords: Content Based Image Retrieval (CBIR),Clustering Algorithms.

Corresponding Author: MR. ANKUR S. MAHALLE

Access Online On:

www.ijpret.com

How to Cite This Article:

Ankur Mahalle, IJPRET, 2014; Volume 2 (9): 68-74

PAPER-QR CODE

INTRODUCTION

Image plays vital role in every aspect of business such as business images, satellite images, and medical images and so on. Image retrieval is the fast growing and challenging research area with regard to both still and moving images. Many Content Based Image Retrieval (CBIR) system prototypes have been proposed and few are used as commercial systems. CBIR aims at searching image databases for specific images that are similar to a given query image [2]. It also focuses at developing new techniques that support effective searching and browsing of large digital image libraries based on automatically derived imagery features. It is a rapidly expanding research area situated at the intersection of databases, information retrieval, and computer vision. Meanwhile, the next important phase today is focused on clustering techniques. Clustering algorithms can offer superior organization of multidimensional data for effective retrieval. Clustering algorithms allow a nearest neighbour search to be efficiently performed [2]. Hence, the image mining is rapidly gaining more attention among the researchers in the field of data mining, information retrieval and multimedia databases. Spatial Databases is the one of the concepts which plays a major role in Multimedia System. Image mining normally deals with the extraction of implicit knowledge, image data relationship, or other [3].

I. LITERATURE REVIEW

Here we will come to know what are the different techniques to extract the features of the image as well how to extract the images from the image database and then display the results according to the human expectations.

Image Database

Image database is a collection of image data, typically associated with the activities of one or more related organizations. It focuses on the organization of images and its metadata in an efficient manner. Sometimes delves more thoroughly into an image's content. It efficiently stores images in database. We need database because images stored in the database can be directly linked with metadata, Fine grained security is possible. Access to an image can be restricted to individual users and it also offers other restriction controls. Backing up the database will backup every image so it simplifies the process. Only one recovery procedure needed in an event of failure. An image can be converted from one format to another. Metadata can be extracted from it. It can be copied, resized and the image quality controlled.

Sets of images can be deleted, updated or copied as easy as it is to write a query. Images can be linked together and metadata can be easily attached to them. All data related to an image or set of images can logically co-exist[1].

Clustering

Clustering can be considered the most important *unsupervised learning* problem; so, as every other problem of this kind, it deals with finding a *structure* in a collection of unlabeled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way". A *cluster* is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. But how to decide what constitutes a good clustering? It can be shown that there is no absolute "best" criterion which would be independent of the final aim of the clustering. Consequently, it is the user which must supply this criterion, in such a way that the result of the clustering will suit their needs. Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold or the number of expected clusters) depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data preprocessing and model parameters until the result achieves the desired properties[2].

Image Mining Techniques

Advances in image acquisition and storage technology have led to tremendous growth in very large and detailed image databases. A vast amount of image data such as satellite images, medical images, and digital photographs are generated every day. The World Wide Web is regarded as the largest global image repository. An increasing proportion of the contents in digital libraries are images [3]. These images, if analyzed, can reveal useful information to the human users. Unfortunately, it is difficult or even impossible for human to discover the underlying knowledge and patterns in the image when handling a large collection of images.

Image mining is rapidly gaining attention among researchers in the field of data mining, information retrieval, and multimedia databases because of its potential in discovering useful image patterns that may push the various research fields to new frontiers. Image mining systems that can automatically extract semantically meaningful information (knowledge) from image data are increasingly in demand. The fundamental challenge in image mining is to determine how low-level, pixel representation contained in a raw image or image sequence can be efficiently and effectively processed to identify high-level spatial objects and relationships. In other words, image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the image databases [3].

The images from an image database are first pre processed to improve their quality. These images then undergo various transformations and feature extraction to generate the important features from the images. With the generated features, mining can be carried out using data mining techniques to discover significant pattern.

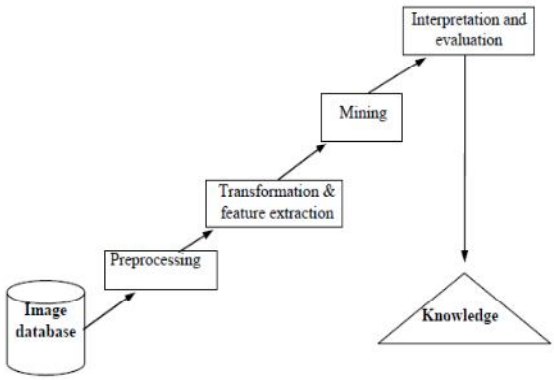

Fig 1: Image Mining Process[3]

Early work in image mining has focused on developing a suitable framework to perform the task of image mining. The image database containing raw image data cannot be directly used for mining purposes. Raw image data need to be processed to generate the information that is usable for high-level mining modules. An image mining system is often complicated because it employs various approaches and techniques ranging from image retrieval and indexing schemes to data mining and pattern recognition. A good image mining system is expected to provide users with an effective access into the image repository and generation of knowledge and patterns underneath the images. Such a system typically encompasses the following functions: image storage, image processing, feature extraction, image indexing and retrieval, patterns and knowledge discovery. At present we can distinguish two kinds of frameworks used

to characterize image mining systems: function-driven versus information-driven image mining frameworks. The former focuses on the 6 functionalities of different component modules to organize image mining systems while the latter is designed as a hierarchical structure with special emphasis on the information needs at various levels in the hierarchy.

II. SYSTEM DESIGN

Design Engineering deals with the various UML [Unified Modeling language] diagrams for the implementation of project. Design is a meaningful engineering representation of a thing that is to be built. Software design is a process through which the requirements are translated into representation of the software. Design is the place where quality is rendered in software engineering. Design is the means to accurately translate customer requirements into finished product. This project has following phases

1. Input query.
2. Searches image in the database.
3. Image clustering.
4. Uses RGB features to extract the image.
5. Displays the results.

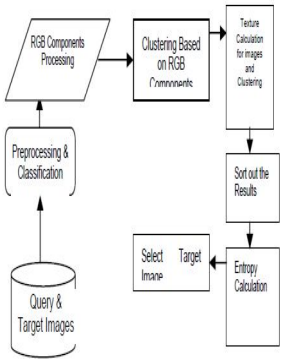

Fig. Block Diagram with components

Fuzzy C- Means Algorithm

FCM is one of the most popular fuzzy clustering techniques. It is an approach, where the data points have their membership values with the cluster centers, which will be updated iteratively. The FCM algorithm consists of the following steps[6]:

Step 1: Let us suppose that M-dimensional N data points represented by xi

(i = 1, 2, . . . ,N), are to be clustered.

Step 2: Assume the number of clusters to be made, that is, C, where 2 ≤ C ≤ N.

Step 3: Choose an appropriate level of cluster fuzziness f > 1.

Step 4: Initialize the N × C ×M sized membership matrix U, at random, such

that Uijm ∈ [0, 1] and∑Cj=1 Uijm = 1.0, for each i and a fixed value of m.

Step 5: Determine the cluster centers CCjm, for jth cluster and its mth dimension

by using the expression given below:

$$CC_{jm} = \frac{\sum_{i=1}^{N} U_{ijm}^{f} x_{im}}{\sum_{i=1}^{N} U_{ijm}^{f}}.$$

Step 6: Calculate the Euclidean distance between ith data point and jth cluster center with respect to, say mth dimension like the following:

$$D_{ijm} = \|(x_{im} - CC_{jm})\|.$$

Step 7: Update fuzzy membership matrix U according to Dijm. If Dijm > 0,then

$$U_{ijm} = \frac{1}{\sum_{c=1}^{C} (\frac{D_{ijm}}{D_{icm}})^{\frac{2}{f-1}}}.$$

If Dijm= 0, then the data point coincides with the corresponding data point of jth cluster center CCjm and it has the full membership value, that is, Uijm = 1.0.

Step 8: Repeat from Step 5 to Step 7 until the changes in U ≤ ϵ, where ϵ is a pre-specified termination criterion.

III. CONCLUSION

This Project has enabled us to visualize a system that proves to be more accurate in Searching and Extracting Images from databases. Unlike the conventional search methods this technique has enabled us to achieve accuracy in data mining techniques and a feedback system that enables users to get a panoramic view of how and why the specific results were achieved thus

making this system more reliable with transparent operations. This project's implementation in the real will help us to find the image more faster in the image database in effective way as well as by extracting the features it will be easy to display the results according to human expectations which was given as the input query.

REFERENCES

1. Integrated Browsing and Querying for Image Databases, Simone Santini and Ramesh Jain, University of California, San Diego .

2. Image Clustering and Retrieval using Image Mining Techniques, A. Kannan, Dr. V. Mohan Dr. N. Anbazhagan, 2010 IEEE International Conference on Computational Intelligence and Computing Research.

3. Image Mining: Trends and Developments, Ji Zhang Wynne Hsu Mong Li Lee.

4. A. Kitamoto. Data Mining for Typhoon Image Collection. In Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), San Francisco, CA,USA, August, 2001.

5. O. R. Zaiane, J. W. Han et al. Mining MultiMedia Data. CASCON'98: Meeting of Minds, pp 83-96, Toronto, Canada, November 1998.43

6. A Comparative Study Of Fuzzy C-Means Algorithm And Entropy-Based Fuzzy Clustering Algorithms, Subhagata Chattopadhyay,2011.

7. M. C. Burl et al. Mining for Image Content. In Systemics, Cybernetics, and Informatics /Information Systems: Analysis and Synthesis, (Orlando, FL), July 1999.

8. J. S. D. Bonet. Image Preprocessing for Rapid Selection in "Pay attention mode". *MIT*, 2000.

9. S. Berchtold, D. A. Keim and H. P. Kriegel. The X-tree: An Index Structure for High dimensional Data. *In Proceedings of the 22nd VLDB Conference*, pages 28-39, Mumbai, India, September 1996.

10. Y. Rui, T. S. Huang et al. Image Retrieval: Past, Present and Future. Invited paper in *IntSymposium on Multimedia Information Processing*, Taipei, Taiwan, Dec 11-13, 1997.