ISSN: 2321-7782 (Online)
Volume 3, Issue 3, March 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Image Mining and Clustering Based Image Segmentation

Ankur S.Mahalle[1]
Department of Information Technology,
PRMIT&R, Badnera Amravati(MH), India

Snehal H. Kuche[2]
Department of Computer Science And Engineering,
PRMIT&R, Badnera Amravati(MH), India

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. 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. mage segmentation is the method of dividing. A digital image into several segments. The aim of segmentation is to simplify or modify the signification of an image into meaningful form that is more significant and easier to examine. It is generally used to put objects and edges in images.

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

I. 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]. Image Segmentation is an important process of image processing and understanding. Basically it is defined as the process of dividing the image into different parts of homogeneity. The aim of image segmentation is to simplify the representation of an image into something that is more meaningful and easier to understand [12].

II. 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 Clustering Techniques

Histogram Thresholding Based Methods:

In histogram thresholding method operation of converting a multilevel image into a binary image is performed, where it as signs the value of 0 (background) or 1 (objects or foreground) to each pixel of a digital image based on a comparison with some threshold value T (intensity or color value). If the T is constant, the approach is called global thresholding otherwise, it is called local thresholding. Global thresholding methods can fail when the background illumination is uneven so to compensate for this uneven illumination we can use multiple thresholds and the threshold selection is typically done interactively. These methods are popular because of their simplification and efficiency. The problems that arise in this kind of method is that basic histogram based thresholding algorithm do not process those images which have histograms that are unimodal when the target segment is much smaller than the background area[11].

Edge Detection Based Methods:

Edge detection method is widely used in the field of medical image segmentation. This method helps to locate the pixels in the image that correspond to the edges of the objects seen in the image and the result is a binary image with the detected edge pixels. The algorithms that are commonly used are Sobel, Prewitt and Laplacian operators. These algorithms are best applied to images that are simple and do not carry any noise. But this method does not work well when images have too many edges and noise so it will be unable to identify a closed boundary easily[11].

FELICM:

It stands for Fuzzy C-Means with Edge and Local Information, which introduce the weights of pixels within local neighbours windows to reduce the edge degradation. [1] Basically this method has somehow tried to overcome the isolated distribution of pixels inside segments of image. The basic process of FELICM is that in this method firstly the original image is being converted into gray image and then the principal components analysis is taken. Then in the next step edges are obtained by adjusting two threshold values that is a high threshold value and low threshold value in canny edge detection algorithm. After analyzing the edges ,different weights are set to the neighbours within the local windows. Then the clustering is done with the FELICM method by using the spatial and spectral information[11].

III. SYSTEM DESIGN

Design Engineering deals with the various UML [Unified Modeling language] diagrams for the implementation. 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 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.

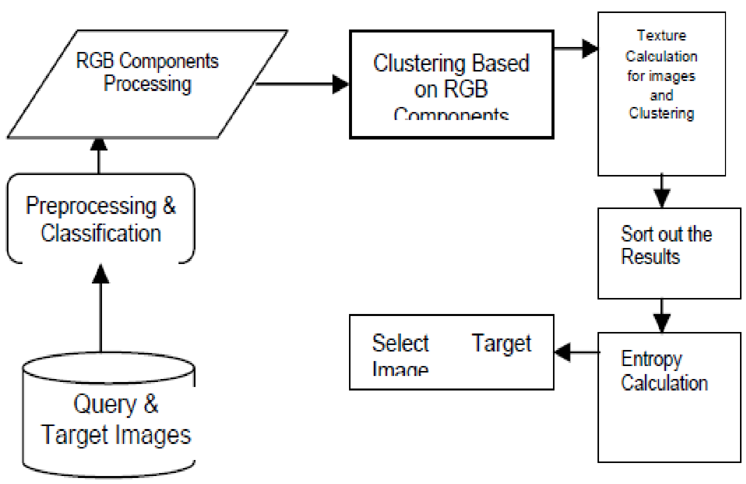

Fig. Block Diagram with components

Proposed Algorithm

Step1. Input image:

First of all, Image "I" will be passed to a proposed algorithm as an input image.

Step2. In this step PCA and OTSU threshold are applied separately.

(a) Principle Component Analysis :

Apply RGB2PCA to convert given image in PCA plane.

$$[M,N,D] = \text{size}(I(x,y)) \ldots\ldots(1)$$

Where M represent rows, N represent columns and D represent dimensions. I(x,y) is an input image.

(1) To convert RGB image to PCA, first each component of an RGB image should be converted into vector. Then all these vectors are concatenated by using following equation

$$IIV = cat(2, R, G, B)........ (2)$$

Where IIV represent the Input Image Vector and cat represent the concatenate function.

(2) Then Eigen values are computed by using principal component function which is given by following equation

$$VV = princomp(IIV)....... (3)$$

Where VV represent the vector values and princomp is inbuilt function in MATLAB.

(3) PCA vector is obtained from vector values by using following equation

$$Vector = VV/(\sum(VV)........ (4)$$

(4) Finally PCA image is obtained from vector representation by using following function

$$OVI = IIV*Vector......... (5)$$

Where OVI represent the output vector image and IIV represent the Input Image Vector

(b) OTSU Thresholding: Otsu's method is employed to automatically execute clustering-based image thresholding or the diminution of a gray level image to a binary image .In Otsu's techniquethe threshold that reduces the intra-class variance defined as a weighted sum of variances of the two classes.

Step3. Canny Edge Extraction:

It is an edge detection worker that employs a multi-stage technique to notice a large variety of edges in images.

1) Noise diminution. The gray image is convolved with a 5×5 Gaussian filter by standard deviation σ= 0.4.

2) Locating the intensity gradient of the image.

3) Non maximum suppression decides if the pixel is a enhanced applicant for an edge than its neighbours.

4) Drawing edges throughout the image and hysteresis thresholding.

Step4.

It has included two steps i.e. Neighborhood Weighting and FELICM.

(a) Neighborhood Weighting: In this step, if the direct line between two pixels is cut off by an edge, these two pixels go to dissimilar regions.

(b) FELICM: It stands for Fuzzy C - Means with Edge and Local Information, which initiate the weights of pixels inside local neighbor's windows to decrease the edge poverty.

Step 5. Edge Pixel Processing:

In this step, edges are extracted with the help of FELICM and Canny edge extraction.

Step 6. Clustering Result:

In this step, Clustering is done with the FELICM method by using spatial and spectral information. Basically clustering is essentially a group of such clusters, frequently containing every object to each other[12].

IV. Conclusion

This Paper 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 paper'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.
11. Evaluating the Short Comings of Clustering based Segmentation Techniques Amanjot Kaur Randhawa, Dr. Rajiv Mahajan IJESRT June,2014.
12. Improved Felicm based Underwater Color Image Segmentation by using Gradient Minimization and DBPTGMF, Rozy Kumari, Narinder Sharma, International Journal of Computer Applications (0975 –8887) Volume 105 –No. 16, November 2014.

Author(s) Profile

Prof. Ankur S. Mahalle, received the M.E degree in Information Technology from PRMIT&R, Badnera in 2014. Now working in Department of Information Technology in Porf. Ram Meghe Institute of Technology & Research, Badnera,Amravati(MH), India.

Prof. Ms. Snehal H. Kuche, received the M.E degree in Wireless Communication & Computing from G. H. Raisoni, Nagpur in 2012. Now working in Department of Computer Science & Engineering in Porf. Ram Meghe Institute of Technology & Research, Badnera,Amravati(MH), India.