ISSN:0975-9646
N.Naveen Kumar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (4) , 2011, 1717-1719

Image Clustering Technique for Web Search Engine Retrieval System

N.Naveen Kumar, S.Ramakrishna

Department of Computer Science, Sri Venkateswara University, Tirupathi, A.P., India.

***Abstract*-In Web Search Engine, Clustering is an efficient way of reaching information from raw data and K-means is a basic method for it. Although it is easy to implement and understand, but it has serious drawbacks. So we go for some other techniques for filtering process like greedy global algorithm. These types of algorithms are also work as a text mining techniques over the web and also cluster the relevant data according to the input query. Using this web mining process we can download relevant documents only. Even though, it can't produce exact result for the query. Image search process is also work as the text mining techniques, during this image clustering or search process we only have some basic methodology like size, type of Image (jpg, bmp). In our process we go to implement an algorithm for image retrieval system over the web. Here, image as input query for the web search engine. According to this technique we improve the precision of web based image retrieval system. Generally, Co-clustering technique is used for text comparison only here we use image comparison process and also performs advance fuzzy c-means for clustering process. These concepts are smooth to grouping images as well as downloading process.**

***Keywords*— Image clustering technique, Web search engine retrieval, Image texture segmentation, Optimize time processing**

1. INTRODUCTION

The World Wide Web is an interlinked collection of billions of documents formatted using HTML. It provides a vast source of information of almost all type. Finding the right information from such a large collection is extremely difficult. However, this information is often scattered among many web servers and hosts, using any different formats. We all want that we should have the best possible search in less time. A user's interaction with the search output is often far from optimal. If the few sampled output are not found relevant, it is very likely that the rest of the output will not be inspected at all. To avoid this kind of result, we need methods about efficient image indexing and retrieval from image databases have been proposed for the applications such as digital library or train the image database. Low-level visual features such as color, texture, and shape are often employed to search relevant images based on the query image. Among these features, color constitutes a powerful visual cue and is one of the most salient and commonly used features in color image retrieval systems.

The proposed method is Modified Fuzzy C-Means technique with an adjustable central weight obtained by partitioning the observation vector space. Dominant points of the proposed approach are partitioning of observation vector space using fuzzy c-means clustering method, training procedure using this method and then applying the Co-clustering algorithm. It is one of the simplest clustering algorithms in document retrieval process. In this paper, Co-clustering technique was implemented for image clustering. Generally, any unsupervised classification method applied to contend based image retrieval is to gather considered images to be similar. In this case, algorithms treat the data in only one direction: lines or columns, but not both at the same time. Contrary to the one dimensional clustering, Co-clustering proposes to process the data tables by taking of account the lines and the columns in a simultaneous way. We propose in this paper a new Co-clustering modeling which introduces a two levels similarity concept. The aim of this method is to improve the image retrieval accuracy and to optimize time processing.

2. CLUSTERING PROCESS

Clustering analysis plays an important role in web mining research. A widely adopted definition of optimal clustering is a partitioning that minimizes distances within a cluster and maximizes distances between clusters. In this approach the clusters and, to a limited degree, relationships between clusters are derived automatically from the data to be clustered, and the data are subsequently assigned to those clusters. It made huge progress in such aspects as image texture segmentation, remote-sensing image segmentation and edge detection etc.

2.1. Image segmenting/training by Fuzzy C-Means

Image segmentation means that image is indicated as set of physically meaningful connected areas. Generally we achieve image segmentation purpose through analyzing such different image characteristics as edge, texture, color and brightness etc. Integration of fuzzy logic with data mining techniques has become one of the key constituents of soft computing in handling the challenges posed by massive collections of natural data. The central idea in fuzzy clustering is the non-unique partitioning of the data in a collection of clusters.

The fuzzy c-means algorithm is based on minimization of the following objective function, with respect to μ, a fuzzy c-partition of the data set, and to v,μ a set of c prototypes

$$J_{FCM} = \sum_{X=1}^{N} \sum_{I=1}^{C} \mu_{X,i}^{m} d^2(Z_X, V_1) \qquad (1)$$

where $\mu_{x,i}$ (x=1,2,...,N, i=1,2, ...,c) is membership value, it denotes fuzzy membership of data point x belonging to class i, v_i (i=1, 2, ..., c) is centroid of each cluster and z (x=1,2,..., N) is data set (pixel values in image), m is fuzzification parameter, (z_x, v_i) is Euclidean distance between z_x and v_i, N is the number of data points, c is number of clusters.

Fuzzy partition is carried out through an iterative optimization of equation (1) according to [4]

1) Choose primary centroids v_i (prototypes).

2) Compute the degree of membership of all data set in all the clusters

$$\mu_{X,i} = \frac{\left[\frac{1}{d^{2(Z_X,V_i)}}\right]^{1/(m-1)}}{\sum_{i=1}^{C}\left[\frac{1}{d^{2(Z_X,V_i)}}\right]^{1/(m-1)}} \qquad (2)$$

3) Compute new centroids v_i^1

$$V_i^1 = \frac{\sum_{X=1}^{N}\mu_{X,i}^m Z_X}{\sum_{X=1}^{N}\mu_{X,i}^m} \qquad (3)$$

and update the degree of memberships, $\mu_{x,i}$ to μ^i x,i according to equation

4) if $\max_{x,I}[|\mu_{x,I} - \mu^1_{x,i}|] < Є$ stop otherwise go to step 3.

Where Є is a termination criterion between 0 and 1, and d^2 (z_x, v_i) is given by

$$d^2(z_x\ v_i) = \|z_x - v_i\|^2 \qquad (4)$$

use this method ,trine all database images. Images are separate as segments using this FCM algorithm. How the system process the segmentation technique is explained [1].

2.2. Co-clustering Process

Co-clustering use a bipartite graph to model the images and their contents simultaneously. For the two sets of vertices of a bipartite graph, we associate one set with image content descriptors and the other with the images. The edges linking these two sets of vertices measure the degrees of association between the content descriptors and the images. We then introduce graph partitioning to cut the graph such that images and the features that are most strongly associated with each other are clustered into the same group. Each group of images formed in this way manifests a certain visual theme which is strongly linked to its associated visual features, which in turn provides the content semantics for the images, thus facilitating image indexing and retrieval.

We propose a recently developed image clustering technique, to first cluster a large image database into image clusters; each contains images of certain homogeneous property such as color or texture. We then transfer the color statistics of the image clusters into the input image to correct and enhance the color appearances of the input image.

In Fig.1 for a given input image (to be processed or query), we use query based image retrieval to select example images either automatically or interactively. The purposes of the processing can be varied from making the input image to take on some of the statistical characteristics of the retrieved example images to using the example images to relative the appearances of the input image.

3. EXPERIMENTAL RESULTS

We used nearly 200 images for our proposed technique. Server maintain image table for image selection process according to the user's query. In our propose system user send the query through mobiles, server accept query. Before that server train the master image DB. Trained information maintain by the server (as shown in Fig. 2). Using Image Table has the input image information. Using this, Image selection process done by the server. After that selected image (Byte Stream) and trained image as the input for the Co-clustering technique. Finally it identifies the relevant image which present in the Image Database and then server automatically transmit result to the user (mobile).

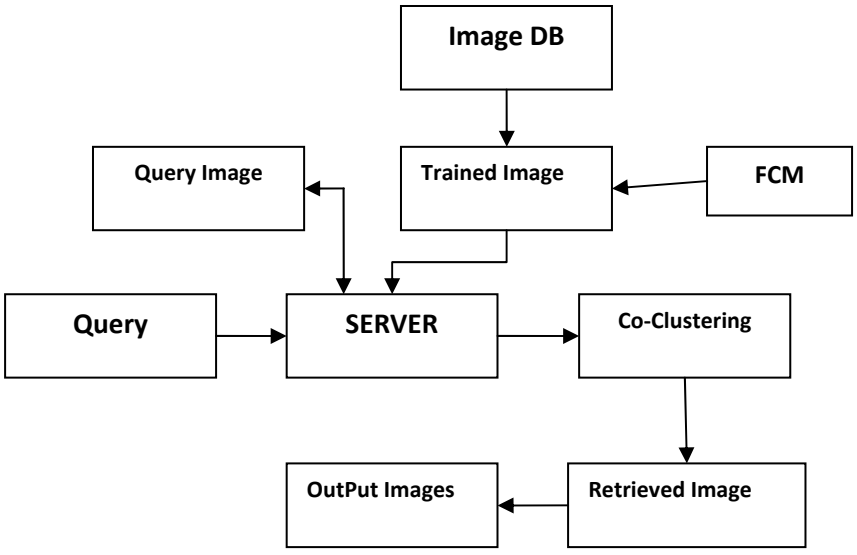

Fig. 1. Image Clustering Technique for Web Search Engine Retrieval **System**

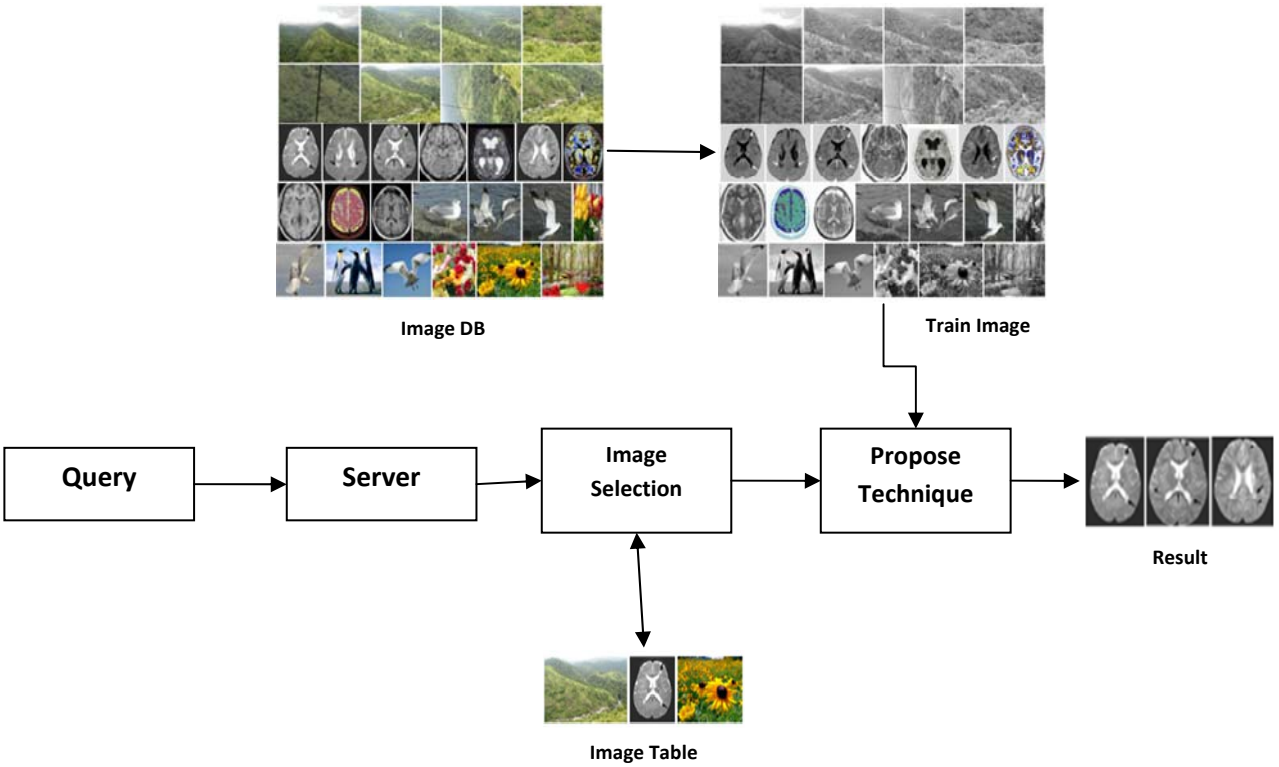

Fig. 2. Process flow for Web Search Engine Retrieval System

4. CONCLUSION

Our Proposed system provides efficient image search from large scale database. However due to its fix duration time interval it generates many results. Whereas it provides an efficient image search without generating many intermediate query results in larger image database. Secondly a client user can define the time when he wants to see the query results regardless of database size and the feature selected in the database. Unlike previous approaches to image clustering, our method associates features with clusters and explicitly identifies which features are more important to which cluster.

REFERENCES

[1] M. Ameer Ali, Gour C Karmakar and Laurence S Dooley, "Fuzzy Image Segmentation using Suppressed Fuzzy C-Means Clustering (SFCM)," *in: International Conference on Computer and Information Technology*, Dhaka, Bangladesh, December 2004.

[2] A. W. M. Smeulders et al, "Content-based image retrieval at the end of the early years", *IEEE Transaction PAMI*, vol. 22, pp. 1349 – 1380, December 2000.

[3] M. Hanmandlu, O.P. Verma, N.K. Kumar, M.Kulkarni, "A Novel Optimal Fuzzy System for colour image Enhancement using bacterial foraging," *IEEE Trans. Measurements and Instrumentation*, August 2009.

[4] P. Gupta, K. Johari, "Implementation of Web Crawler," *Second International Conference on Emerging Trends in Engineering and Technology, ICETET-09*, Nagpur, pp. 838-843, January 2010.

[5] I. Ahmad, M. Gabbouj, "Content-based Interactive Image Retrieval from Java Enabled Mobile Devices," *Proceedings of the 4th International Symposium on Communications Control and Signal Processing*, *SCCSP 2010*, Limassol, cyprus, pp. 1-4, March 2010.

[6] S. Kiranyaz, M. Gabbouj, "An Interactive Query Implementation over High Precision Progressive Query Scheme", *In Proc. of WIAMIS Workshop*, Korea, April 2006.

[7] A.L. Oliveira, S.C. Madeira, "Biclustering algorithms for biological data analysis: A syrvey,", *EEE/ACM Trans. Computational Biology and Bioinformatics* , vol. 1, no. 1,pp. 24-45, 2004.