Content Based Image Retrieval Using Hierachical and Fuzzy C-Means Clustering

Prof.S.Govindaraju[#1], Dr.G.P.Ramesh Kumar[#2]

[#1]Assistant Professor, Department of Computer Science, S.N.R. Sons College, Bharathiar University, Coimbatore, Tamil Nadu, India-641006.
[#2]Assistant Professor, Department of Computer Science, Govt. Arts College, Kulithalai, Bharathidasan University, Tamil Nadu, India-639120.

Abstract - *Grouping images into semantically meaningful categories using low-level visual feature is a challenging and important problem in content based image retrieval. CBIR is a part of image processing. We know that with the development of the internet and the availability of image capturing devices such as digital cameras, image scanners, and size of the digital image collection is increasingly rapidly and hence there is a huge demand for effective image retrieval system. Normally CBIR is retrieving/ searching stored images from a collection by comparing features automatically extracted from the image themselves. The most common features used are mathematical measure is texture, color and shape. Clustered images are utilized by content-based image retrieval and querying system that requires effective query matching in large image database. Particularly, Inthis paper we are using HFCM Algorithm. It has the combinational advantage of both fuzzy and possiblistic approaches. The experimental results suggest that the proposed image retrieval technique results in better retrieval.*

Keywords *-Query, Hybrid Fuzzy C-Means, Content Based Image Retrieval.*

I. INTRODUCTION

The growing amount of digital images caused by more and more ubiquitous presence of digital cameras and, as a result, many images on the world wide web confronts the users with new problems. Normally, the retrieval of the content based image involves the following systems [11].

A. COLOR –BASED RETRIEVAL

Color feature is the most sensitive and obvious feature of the image, and generally adopted histograms are describing it. Color histogram method has the advantages of speediness, low demand of memory space and non-sensitive with the images. Change of size and rotation, it bins extensive attention consequently [5].

B. RETRIEVAL BASED ON TEXTURE FEATURE

When it refers to the description of the image's texture, it usually adopt texture's statistic feature and structure feature as well as the features that based on spatial domain are changed into frequency domain[9]. The homogeneous texture descriptor describes a precise statistical distribution of the image texture. It enables to classify images with high precision and it is to be used for similarity retrieval applications.

C. THE RETRIEVAL BASED ON SHAPE FEATURE

Here, there are some problems needs to be solved during the image retrievalbased on shape feature. Firstly, shape usually related to the specifically object in the image, so shape's semantic feature is stronger than texture [12].

This paper focuses on using Fuzzy C-Means algorithm which is typical clustering algorithm that has been widely utilized in engineering and scientific disciplines such as medicine imaging, bio-informatics, pattern recognition and data mining. As the basic FCM clustering approach employs the squared – norm to measure similarity between prototypes and data points, it can be effective in clustering only the spherical clusters and many algorithms are derived from the FCM to cluster more general dataset[14].

II. METHODOLOGY

In this work, the main focus is the application of clustering algorithm for content based image retrieval. A large collection of images is partitioned into a number of image clusters. Given a query image, the system receives all images from the clusters. Given a query image, the system retrieves all images from the cluster that is closest in content to the query image. The overall system is shown in Fig-1.

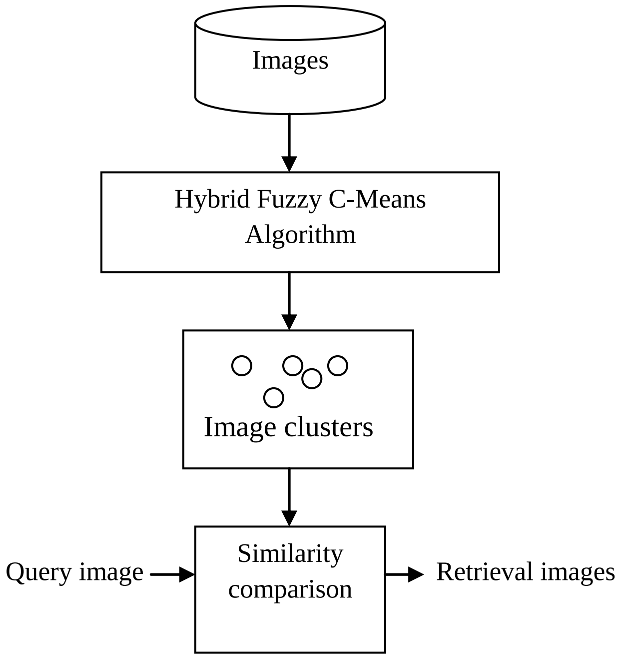

Fig-1 Content – Based Image Retrieval System

The proposed clustering algorithm is applied to image retrieval and compared its performance with Fuzzy possiblistic Clustering algorithm. Each image in the database is represented by a visual content descriptor consisting a set of visual features. A similarity/ dissimilarity measure is then used to retrieve images. Whose features are closest to that query image? A common distance/ dissimilarity metric is the Euclidean distance, which is used in this work.

III. IMAGE CLUSTERING BASED ON HYBRID FUZZY C-MEANS(FCM) CLUSTERING ALGORITHM

The choice of an appropriate objective function is the key to the success of the cluster analysis and to obtain better quality clustering results; so the clustering optimization is based on objective function. To meet a suitable objective function, we have started from the following set of requirements. The distance between clusters and the data points assigned to them should be maximized and the distance between clusters is modeled by term; it is the formula of the objective function [11].

Fuzzy C-means is a clustering method which allows a piece of data to belong to two or more cluster, which is frequently used in computer vision, pattern recognition and image processing. The FCM algorithm obtains segmentation results by fuzzy. Color based classification methods which group a pixel belong exclusively to one class.FCM approach is quite effective for color based image segmentation. [10]Several segmentation algorithms are based on fuzzy set theory. Fuzzy C-means is a clustering algorithm that used membership degree to determine each data point belongs to a certain cluster. FCM divided the n vectors X_i(i=1,2,3…..n) into C fuzzy group and computing the cluster center of each group making value function of non-similarity index to achieve the minimum.[6]Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. It is based on minimization of the following objective function:

$$J_m = \sum_{i-1}^{N} \sum_{j-1}^{C} U_{ij}^{m} ||x_i - c_j||^2$$

,$1 \leq m < \infty$.where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center.[11] Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k-1}^{c} \left(\frac{||xi - cj||}{||xi - cj||} \right)^{2/m-1}}$$

$$c_j = \frac{\sum_{i-1}^{N} U_{ij}^{m} \cdot Xi}{\sum_{i-1}^{N} U_{ij}^{m}}$$ This iteration will stop when

$\mathrm{Max}_{ij}\{| U_{ij}^{(k+1)} - U_{ij}^{(k)} |\} < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of J_m. the steps are given below

Step1 : Initialize U=[u_{ij}] matrix, $U^{(0)}$
Step2: At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i-1}^{N} U_{ij}^{m} \cdot xi}{\sum_{i-1}^{N} U_{ij}^{m}}$$

Step:3 Update $U^{(k)}$, $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k-1}^{c} \left(\frac{||xi-cj||}{||xi-cj||} \right)^{2/m-1}}$$

Step4: If $|| U^{(k+1)} - U^{(k)} ||< \varepsilon$ then STOP; otherwise return to step 2.

IV. EXPERIMENTAL RESULTS

The test data consists of 777 images belonging to 18 categories obtained from the University of Washington's object and concept recognition for CBIR research project image dataset. Each category contained varying number of images. All the images contained a textual description mentioning the salient foreground objects. The images were clustered using the algorithm with the centroids chosen at random. The cluster whose centroid was closest in distance to the given test image was determined and the images belonging to the cluster were retrieved. The results were then compared with images retrieved using the Fuzzy C-Means algorithm with the same set of initial Centroids. The following performance measures used to evaluate the performance of the algorithm are Precision and Recall is given below.

$$\text{Precision} = \frac{totalnumberofreterivedrelevantimages}{totalnumbe\ \ rofreterivedimages}$$

$$\text{Recall} = \frac{totalnumberofreterivedrelevantimages}{totalnumberofrelvantimages}$$

A. PRECISION

The precision value resulted for using various techniques is presented in figure-2. The number of clusters is varied for determining the performance of reterival techniques. Different number of clusters used in this paper for evaluation is 1, 2,3,4,5,16,7 and 8. From the result, it can be observed that the usage of HFCM results in lesser precision when compared to the usage of K-Means algorithm for clustering. For the considered number of clusters maximum

precision value for proposed approach is 0.81 and minimum precision value is 0.69 where as the usage of K-Means results in higher precision value that is maximum of 0.91 and minimum of 0.79

B. RECALL

The recall value resulted for using various techniques is presented in fig-3. From the result, it can be observed that the usage of HFCM results in higher recall value when compared to the usage of K-means algorithm for clustering. For the considered number of clusters, maximum recall value for proposed approach 0.71 and minimum recall value is 0.31 where as the usage of K-Means algorithm results in lesser precision value that is maximum of 0.61 and minimum is 0.25

Table 1 PrecisionAnd Recall Values in %

Query Image	Hierarchical with K-Means		Hierarchical with Fuzzy C-Means	
	Precision	Recall	Precision	Recall
1	92.71	83.62	96.33	92.12
2	95.6	90.84	99.65	98.66

IV. CONCLUSION

This paper proposes a new method for unsupervised image clustering using probabilistic continuous models and information theoretic principles. Image clustering relates to Content Based Image Retrieval systems. It enables the implementation of efficient retrieval algorithms and the creation of a user friendly interface to the database. This paper uses Hybrid Fuzzy C-Means Clustering algorithm for retrieving the relevant images. An experimental result shows that the proposed technique results in better retrieval of relevant images when compared to the existing approach.

(a) 5 matches out of 12; 11 out of 24

(a) 9 matches out of 12; 17 out of 24

Fig. 2 The retrieved images when the Hierarchical and K-Means are used for clustering. The query image is the upper-left corner image of each block of images

(a) 11 matches out of 12; 20 out of 24

Fig. 3 The retrieved images when the Hierarchical and Fuzzy C-Means are used for clustering. The query image is the upper-left corner image of each block of images

V. REFERENCES

[1] Li, J., Wang, J. Z. and Wiederhold, G.,"*Integrated Region Matching for ImageRetrieval*," ACM Multimedia, 2000, p. 147-156.

[2] Flickner, M., Sawhney, H., Niblack, W.,Ashley, J., Huang, Q., Dom, B., Gorkani, M.,Hafner, J., Lee, D., Petkovic, D., Steele, D.andYanker, P., "*Query by image and videocontent: The QBIC system*," IEEE Computer,28(9), 1995,pp.23-32

[3] Pentland, A., Picard, R. and Sclaroff S.,"*Photo book: Content based manipulation of image databases*", International Journal ofComputer Vision, 18(3), 1996, pp.233–254

[4] Smith, J.R., and Chang, S.F., "*Single color extraction and image query,"* In ProceedingIEEE International Conference on ImageProcessing, 1997, pp. 528–531

[5] Gupta, A., and Jain, R., "*Visual information retrieval*," Comm. Assoc. Comp. Mach., 40(5), 1997, pp. 70–79.

[6] M. Saadatmand-Tarzjan and H. A. Moghaddam, "*A Novel Evolutionary Approach for Optimizing Content-Based Image Indexing Algorithms*", IEEE Transactions OnSystems, Man, And Cybernetics—Part B: Cybernetics, Vol. 37, No. 1, February 2007, pp. 139 153.

[7] N. Vasconcelos, "*From Pixels to Semantic Spaces: Advances in Content-Based ImageRetrieval*",Computer Volume: 40, Issue: 7, 2007, pp. 20-26.

[8] N. Rasiwasia and N. Vasconcelos, "*A Study of Query bySemantic Example*", 3rd International Workshop onSemantic Learning and Applications in Multimedia,Anchorage, June 2008, pp. 1-8.

[9] N. Rasiwasia, P. J. Moreno and N. Vasconcelos, *"Bridging the Gap: Query by Semantic Example*", IEEETransactions On Multimedia, Vol. 9, No. 5, August 2007,pp. 923-938.

[10] S. Cheng, W. Huang, Y. Liao and D. Wu, "*A ParallelCBIR Implementation Using Perceptual Grouping OfBlock-based Visual Patterns",*IEEE InternationalConference on Image Processing – ICIP, 2007, pp. V -161 - V - 164.

[11] D. Tao, X. Tang, and X. Li "*Which Components are Important for Interactive Image Searching?",*IEEETransactions On Circuits And Systems For VideoTechnology, Vol. 18, No. 1, January 2008, pp. 3-11.

[12] Datta, R., Li, J., Wang, J.Z.: Content-based image retrieval: approaches and trendsof the new age. In: MIR 2005: Proceedings of the 7th ACM SIGMM internationalworkshop on Multimedia information retrieval, pp. 253–262. ACM Press, New York(2005)

[13] Chang, H., Yeung, D.Y.: Kernel-based distance metric learning for content-basedimage retrieval. Image Vision Comput. 25, 695–703 (2007)

[14] Cz´uni, L., Csord´as, D.: Depth-based indexing and retrieval of photographic images.In: Garc´ıa, N., Salgado, L., Mart´ınez, J.M. (eds.) VLBV 2003.LNCS, vol. 2849,pp. 76–83. Springer, Heidelberg (2003)

[15]. Zhang, D.S., Lu, G.: A comparative study on shape retrieval using fourier descriptorswith different shape signatures. In: Proc. of International Conference on Intelligent Multimedia and Distance Education (ICIMADE 2001), Fargo, ND, USA,pp. 1–9 (2001)

[16] Hiremath P.S. and JagadeeshPujari, "*Content Based Image Retrieval using Color Boosted Salient Points and Shape features of an image*", International Journal of Image Processing, Volume (2): Issue(1).

[17] Dahlhaus, E.: "*Parallel Algorithms for Hierarchical Clusterings and Applications to Split Decomposition and Parity Graph Recognition*". Journal of /Algorithms 36, 205-240(2000).

[18] P.Jeyanthi and V.JawaharSenthil Kumar, "Image Classification by K-Means Clustering", Advances In Computational Sciences and Technology, Vol:3, No:1, pp. 1-8, 2010.