Image Retrieval using Partitioning Based Clustering Methods

R.Kalpana, K.Shanthi

Abstract— **In current scenario, retrieval of images form huge amount of database is very important. Content Based Image Retrieval (CBIR) is a collection of techniques for retrieving images from large database. The images are retrieved based on content. The term "content" demonstrate to color, texture and shapes. In this system, color and texture features are retrieved from images. The color features are obtained using Dynamic Color Distribution Entropy of Neighborhoods (D_CDEN). The texture features are obtained using Gray Level Co-occurrence Matrix (GLCM). The mixture these two features are used for performance improvement of Content Based Image Retrieval (CBIR) system. CBIR system extracts the certain features from test images and compares these features with all the images in the database. Comparison of two images in the database is very easy but comparison of lot of images in the database is very difficult. The clustering technique is presented to solve the above problem. In this system, K-Means and Contribution based clustering techniques are used. The K-Means clustering algorithm optimizes only intra cluster similarity. Contribution based clustering enhance both intra and inter cluster similarity. The experimental results shows the comparison between average dispersion measures for both K-Means and Contribution based clustering.**

Index Terms— **Feature Extraction Approaches, Partitioning Based Clustering.**

I. INTRODUCTION

In real world, large amount of data scatter around us and this data may be in various kinds like audio, video, image and graphics etc. The major difficulty is that how to handle the data and obtain the data in correct time. Image retrieval from huge amount of database is essential for different fields. The content Based Image Retrieval (CBIR) system is a collection of techniques for retrieving images from large database [2].

The images are retrieved based on content. The term "content" demonstrates color, texture and shapes. In this paper, color and texture features are used. The color features are obtained using Dynamic Color Distribution Entropy of Neighborhoods (D-CDEN) and Texture features are obtained using Gray Level Co-occurrence Matrix (GLCM). Comparison of two images in the database is very easy but comparison of lot of images in the database is very difficult. The clustering technique is presented to solve the above problem. Clustering is the method of grouping the data point based on similarity. The data points in the one cluster are similar and data points in another cluster are dissimilar. The clustering algorithms can be classified into five categories. The five categories are Partitioning based clustering, Hierarchical based clustering, Grid based clustering, Density based clustering and Model based clustering. The Partitioning based clustering approaches are 1) K-Means 2) K- Medoid 3) CLARANS (Clustering Large Application with Randomized Search). In this paper, Partitioning based clustering is used. K-Means clustering algorithm optimizes only intra cluster similarity [12] and Contribution based clustering optimizes intra and inter cluster similarity [12]. Intra cluster similarity indicates the nearness between the elements of a cluster and inter-cluster similarity indicates the similarity between all the clusters present in the image. In this paper, the average dispersion measures of contribution based clustering are compared with dispersion measures of K-Means Clustering.

Manuscript received April, 2014.

***Kalpana.R**, Computer Science Engineering, United Institute of Technology, Coimbatore, India, 7418392601.*

***K.Shanthi**, Assistant Professor of computer science Engineering, United Institute of Technology, Coimbatore, India,*

II. BACKGROUND AND RELATED WORKS

The mixture of Region content based image retrieval (RCBIR) and Global content based image retrieval (GCBIR) [1] provides better retrieval results compared to different algorithms.

In this paper [3], The RGB colors and Textures (Entropy) features are acquired from images. The acquired features are grouped using Fuzzy C-Means algorithm. Entropy is used to differentiate two images with some threshold limit.

Hierarchical Divide and Conquer K-Means (HDK) and feature extraction methods are discussed [4].

Four clustering techniques are 1) K-Means 2) Ward's Linkage 3) Complete Linkage 4) Average Linkage are explored [5]. The resemblance between query and database images are compared using Euclidean distance and for classification K-Means clusters are used [6].

Here, using hierarchical clustering algorithm the images in the database are grouped into clusters of images depending on the similar color content [7].

Fuzzy Linking approaches are used [8]. It needs only some bins to describe the color distribution of the image. The mixture of Divisive and Agglomerative clustering algorithms is called hybrid clustering [9]. The combination of these two algorithms produces a good cluster quality.

The PSNR (Peak Signal to Noise Ratio) values and K-Means clustering algorithm are discussed [10].

ISSN: 2278 – 7798

Fuzzy C-Means (FCM), Multiple Kernel Fuzzy C-Means (MKFCM) and Fuzzy Local Information C-Means (FLICM) algorithm are reviewed [11].

III. Proposed System

The proposed system takes input as images. Dynamic Color Distribution Entropy of Neighborhoods and Gray Level Co-occurrence Matrix are used for feature extraction purpose. The clusters are formed based on the feature extraction using K-means and Contribution based clustering. Clustering will be more advantageous to reduce the search time of images from the database. K-Means and Contribution based clustering is a partition based clustering. Contribution based clustering optimizes inter cluster similarity and intra cluster similarity. K-Means optimizes only intra cluster similarity. Finally, the average dispersion measures of contribution based clustering are compared with dispersion measures of K-Means clustering.

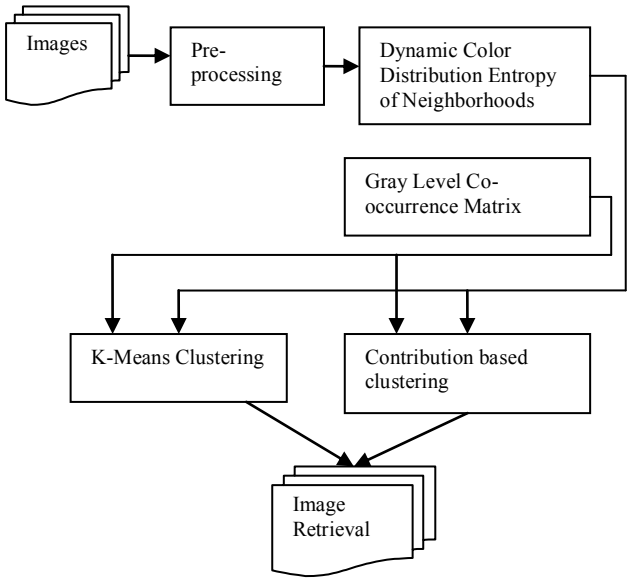

Fig.1. Overview of Proposed System Model

A. Dynamic Color Distribution Entropy of Neighborhoods (D-CDEN)

D-CDEN method illustrates spatial information of colors and based on Color Distribution Entropy (CDE). Color Distribution Entropy is based on Normalized Spatial Distribution Histogram (NSDH).

$$p_x = (p_{x1}, p_{x2,\ldots,} p_{xN})$$

$$p_{xy} = |A'_{xy}| / |A'_x| \quad (1)$$

for $1 \le y \le N$

The NSDH is described using (1). In (1), where N is the number of extracted neighborhoods for color bin x and is different for every color bins and A'_x is the set of pixels with color bin x of an image and $|A'_{xy}|$ is the count of the pixels of neighborhood y for color bin x. The D-CDEN is calculated using NSDH. The D-CDEN is given below,

$$E_X(P_X) = -\sum_{y=1}^{N} p_{xy} \log_2 p_{xy} \quad (2)$$

The D-CDEN is computed by Taking logarithm for p_{xy} and multiply with p_{xy} is illustrated in (2).

B. Gray Level Co-occurrence Matrix

Gray Level Co-occurrence Matrix is used for texture feature extraction purpose [3]. It is used to extort statistics from an image. Co-occurrence matrix works both distribution of intensities and intensity values. The co-occurrence matrix is constructed in four different directions like horizontal, vertical, left diagonal and right diagonal.

In this system, horizontal direction is used for construction of co-occurrence matrix. Many features are taken from this constructed co-occurrence matrix. In this system, Entropy is taken from this matrix. The texture denotes the energy content of the image. Entropy is used to determine the disorder of an image.

C. K-Means Clustering

It is a partitioning based clustering and commonly used simplest algorithm. The K-Means Clustering is used to

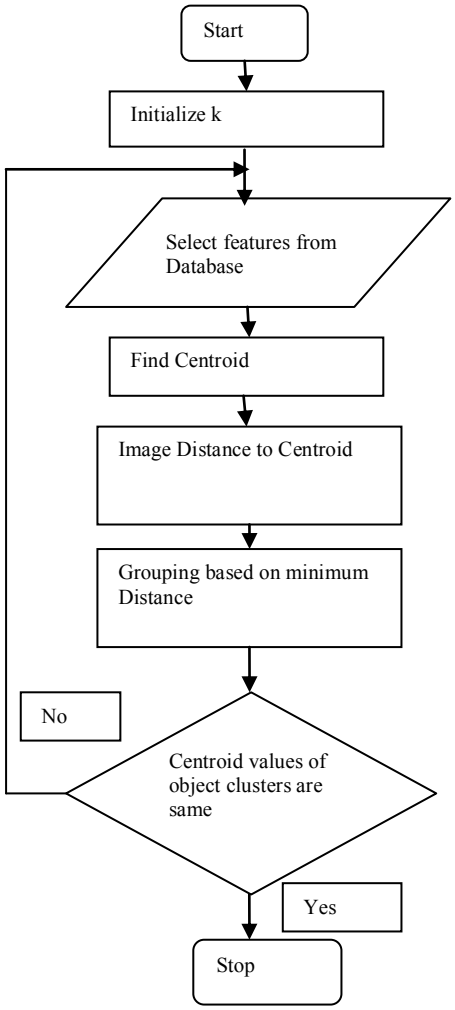

Fig.2. Flowchart for K-Means Clustering Algorithm

ISSN: 2278 – 7798

group n objects into k clusters to guarantee the resemblance among objects in the same cluster and the dissimilarity among samples in different cluster. K- Means Clustering optimizes only intra cluster similarity. The steps for k-means algorithm below,

1. Initialize number of clusters
2. Randomly choose centriod from database
3. Compute the Euclidean distance between data points and cluster centriod using (3)

$$Dist(o_x, o_y) = \sqrt{\sum_{k=1}^{n} (o_{xk} - o_{xk})} \quad (3)$$

4. The clusters are created based on minimum distance.
5. Update the cluster centroid by computing the mean for clusters
6. The procedure is continued until the mean values are same for consecutive iteration

D. *Contribution based clustering*

Contribution based clustering is a partitioning based clustering. Contribution based clustering optimizes both intra and inter cluster similarity. The steps for Contribution based clustering and flowchart is given below,

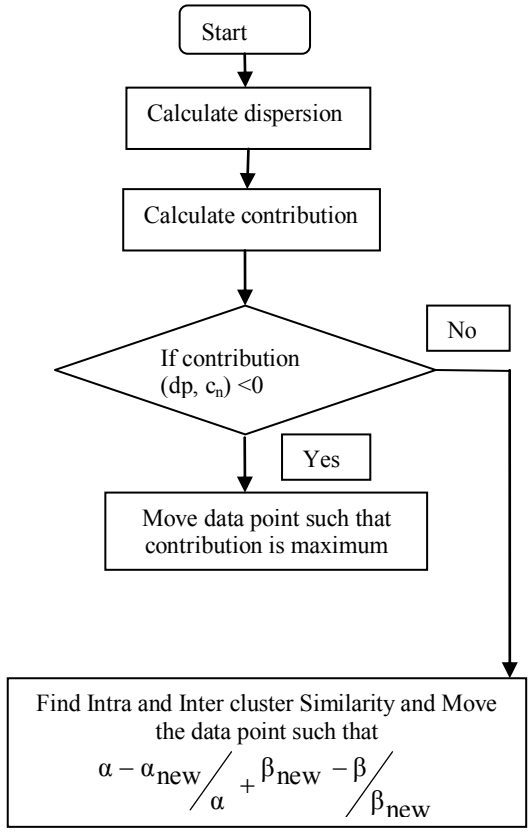

Fig.3. Flowchart for Contribution Based Clustering

1. Initialize number of clusters ($c_1, c_2, \ldots\ldots c_k$)
2. Randomly choose k centriod from image database ($m_1, m_2, \ldots\ldots m_k$)
3. For every data point (dp) do the following steps
   a. Calculate the minimum distance (dp, m_n) up to $1 \leq n \leq k$
   b. Insert the data point dp to cluster c_n and update centriod m_n
4. Calculate dispersion using

$$Dispersion(c_n) = \frac{1}{n} \sum_{dp \in c_n} (dp - m_n) \quad (4)$$

Given a cluster c_n with n points and centriod m_n.

5. Calculate Contribution using

$$Contribution(dp, c_n) = dispersion(c_n - \{dp\}) - dispersion(c_n) \quad (5)$$

6. If contribution(dp, c_n) <0
   a. Shift dp to a cluster c_p such that the contribution (dp, c_p) is maximum
   b. Update centriod m_p
7. If contribution(dp, c_n) >0
   a. Shift dp to a cluster c_p such that is maximum
   $$\alpha - \alpha_{new}/\alpha + \beta_{new} - \beta/\beta_{new}$$
   b. Update centriod m_p

α_{new} & β_{new} are values of α and β after the data point dp shifted to cluster c_p.

IV. Experimental Results

The performance evaluation is performed over 10000 images which are collected from Google. Intra cluster and inter cluster dispersion values for partitioning based clustering such as k-means and contribution based clustering are plotted against number of clusters are demonstrated in fig.4 and fig5.

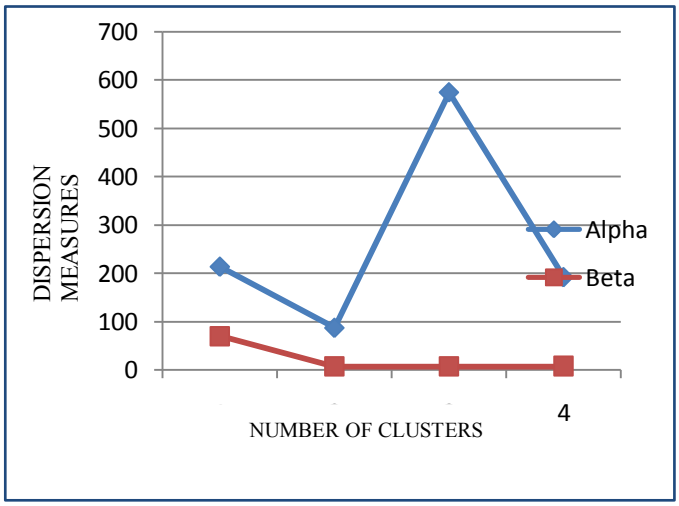

Fig.4. Values of α and β against number of clusters using K-means clustering

ISSN: 2278 – 7798

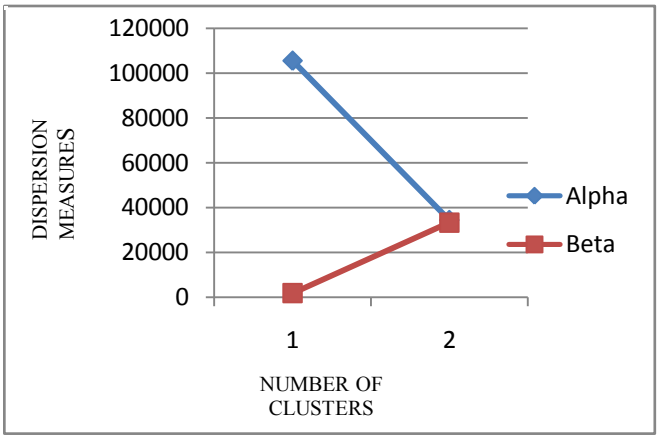

Fig.5. Values of α and β against number of clusters using Contribution based Clustering

We conclude that as the number of clusters increases, contribution based clustering performs both multi objective optimization such as intra Cluster and inter Cluster but K-Means optimizes only intra cluster similarity.

V. CONCLUSION

The images are retrieved based on minimum distance by using K-means and contribution based clustering. The experimental results indicate that Contribution optimizes multi objective optimization but k-means not optimizes the multi objective optimization. In future, valuable features like shape can be integrated for better image retrieval purpose.

REFERENCES

[1] Ibrahim S. I, Abuhaiba1Ruba A, and A. Salamah"Efficient Global and Region Content Based Image Retrieval", *I.J. Image, Graphics and Signal Processing,* no.5 , pp.38-46, 2012.

[2] Fuhui Long, Hongjiang Zhang , and David Dagan Feng," Fundamentals Of Content – Based Image Retrieval ",pp.1-23, 2002.

[3] Kannan.A, V.Mohan , and N. Anbazhagan "Image Clustering and Retrieval using Image Mining Technique",*IEEE International Conference on Computational Intelligence and Computing Research* ,pp.1-5, 2010.

[4] MonikaJain.S, and K.Singh,"Content Based Image retrieval Systems Using Clustering Techniques For Large Data sets ", *International Journal of Managing Information Technology (IJMIT)* Vol.3, No.4, pp. 24-39, 2011.

[5] Mesfin Sileshi, and Bjorn Gamback, "Evaluating clustering Algorithms:Cluster Quality and Feature Selection in Content-Based Image Clustering" , *Int.conf.Image Process*, pp.1-7, 2008.

[6] Ramamurthy.B, and K.R. Chandran, **"**Content Based Medical Image Retrieval With Texture Content Using Gray Level Co-Occurrence Matrix And K-Means Clustering Algorithm *", J. Computer Sci.,* 8 (7): pp.1070-1076, 2012.

[7] Santhana Krishnamachari, and Mohamed Abdel-Mottaleb," Hierarchical clustering algorithm for fast image retrieval **",***Part of the IS&T/SPIE Conference on Storage and Retrieval for Image and Video Databases VII San Jose*, Californi, pp. 427-435,1999.

[8] K. Konstantinidis, A. Gasteratos, and I. Andreadis ," Image retrieval based on fuzzy color histogram processing*" ,Optics Communications* 248 ,pp.375–386, 2005.

[9] Anderson Rocha, Jurandy Almeida1, Mario A, NascimenRicardo Torres, and Siome Goldenstein,"Efficient and Flexible Cluster-and-Search for CBIR",*http://www.liv.ic.unicamp.br /undersun /pub/ communications .html* pp. 1-12, 2008.

[10] JuliRejito,RetantyoWardoyo, Sri Hartati, and Agus Harjoko, "Optimization CBIR using K-Means Clustering for Image Database", *International Journal of Computer science and Information Technologies*, Vol. 3 (4) , pp.4789-4793,2012.

[11] K. Arun Prabha , and P.Malini , **"**An Efficient Algorithm For Clustering Of Images Using Fuzzy Local Information C Means*",The International Journal of Computer Science & Applications (TIJCSA)*, Volume 1, No. 11, ISSN – 2278-1080,pp. 39-50, 2013.

[12] Harikrishna Narasimhan , Purushothaman Ramraj," Contribution Based Clustering Algorithm for Content –Based Image Retrieval",pp.1-6.

ISSN: 2278 – 7798