

# A Review on Contribution Based Clustering Algorithm with Distinct Features

Chandrashekhar S Pawar

*Assistant Professor, Computer Department, R. C. Patel Institute of Technology, Maharashtra, India*

## ABSTRACT

*An image retrieval primarily based on texture features by the usage of the contribution based clustering method to form the cluster is an efficient way to extract the perfect image from database. Color features, shape features and texture feature approach is used for feature extraction. Clustering is an unmonitored category method which place the identical object in a cluster and dissimilar objects are place in unique clusters. Contribution based clustering algorithm is partition based clustering algorithm which offers the better results than that of k-mean clustering algorithm. This contribution based clustering algorithm process on 'contribution of data points'. In image retrieval, use of clustering method gives the better results. To match query image with the database image, feature extraction, similarity measures, indexing and retrieval etc. steps are applied.*

**Keyword:** - Contribution, Clustering, Feature extraction, Image retrieval

## 1. INRODUCTION

There are massive amount of data spread round us and that records are in various form like text, graphics, images, audio, video etc. Essential trouble in front of people is the, how way to manage that data and get it at the suitable time. The retrieval of images from the massive collection of database has turn out to be very essential for specific fields like medical, crime prevention, engineering, artwork, schooling and so on; to get right beneficial information, development in image retrieval method is required. So that to use the different techniques to enhance the managing and accessing method of the data. To reduced the managing and retrieving problem of image retrieval by using Content Based Image Retrieval (CBIR) technique. There are two steps: Feature extraction and similarity measurement [2] [3] [10].

Mostly when color based totally image retrieval is achieved, feature extraction is an vital step in image retrieval. When most effective the color feature extraction method is used that point it tests most effective color of an image and every now and then unwanted images get retrieved. To enhance the retrieval result, feature extractions of an image must improve so textures with the color must take to consider on the time of image feature extraction. After that feature extraction, clustering or classification techniques are used to improve the end result.

Clustering is nothing but partition data into groups on the basis of similar and dissimilar objects are placed in different groups. Clustering comes below the category of unsupervised category [9].

## 2. RELATED WORK

Harikrishna Narasimhan et. Al[3], proposed contribution based clustering algorithm to form clusters in CBIR to improve the end result of CBIR. This technique is used on historic database, natural scene database or medical databases. Visual feature extraction of query image and database image is done by means of the usage of RGB color histogram and then on that contribution based clustering algorithm is applied. Measuring of similarity of the feature

of a query image and database image is performed with the Euclidean distance measure. The comparison of clustering strategies: k-mean and contribution based totally for image retrieval is given.

P. Sankara Rao. Et. Al[4], proposed neural network to enhance effects of CBIR. Clustering is executed by hierarchical clustering algorithm and k-mean method after which on that neural network method is used. Radical basis function is used to find out the relevant images for query image.

P. Bradley. Et. Al [5], proposed scalable clustering framework application for huge databases. Clustering framework utility calls for at most one scan to form cluster. Comparison of clustering framework application and k-mean clustering algorithm is performed for huge database. This framework optimizes result by identifying regions of data which can be compressible, region that have to be maintained in memory and the region which can discard..

Ritu Shrivastava et.Al [6], proposed to evaluate the 2 clustering techniques: K- mean and C-mean clustering for image retrieval. In each technique the distance metric idea is used for the evaluation, each algorithms discover the distance between the centroid of the cluster and seed point. The point which has the minimal distance is included into that unique cluster. K- mean algorithm is simple and fast to compute and C- mean algorithm takes lengthy computational time. Both have the drawback of local minimum.

V.S.V.S Murthy et. Al [7], coloration function extraction of query image and database image is completed after that hierarchical clustering algorithm is used to shape the clusters of an image. Due to forming the clusters, it improves the result of retrieval of an image from big database. Hierarchical clustering algorithm assists faster image retrieval and find out the maximum relevant images from database. K-mean algorithm is used after that of hierarchical clustering algorithm which gives the accurate end result.

### 3. METHODOLOGY

#### 3.1 CLUSTERING ALGORITHM

Clustering algorithm is divided into five types:

1] Partitioned based clustering algorithm: It divides the data into different clustering. It construct partition of database D of n objects into a hard and fast of k-clusters then find out the similarity and dissimilarity of n objects and as in step with that similarity and dissimilarity flow that n<sup>th</sup> object into the k<sup>th</sup> cluster. K-means, K-medoids, contribution based clustering are the Partitioned based totally clustering algorithm.

2] Hierarchical based clustering algorithm: It creates a hierarchical decomposition of information. Establish the cluster classification (Create the tree of clusters). It has two processes:

- i. Agglomerative (bottom up) technique:- In this method, to start with each object is located in separate group after that the similar object merge with each other.
- ii. Divisive (top down) technique:- In this technique, initially all objects placed inside the same cluster after which that cluster is split into smaller clusters..

3] Density based clustering algorithm: Mostly it is beneficial for the arbitrary shape clusters. Cluster continuously grows, till it exceeds it a few threshold. It filters the noise and discover cluster of arbitrary shape.

4] Grid based clustering algorithm: It quantize the data into finite number of cells and form a grid shape. It's processing time is fast.

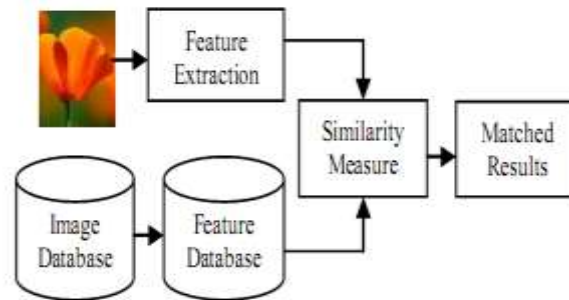
5] Model based clustering algorithm: This algorithm finds the best fit data for given model. It automatically determines the number of clusters based on standard statistics.

The selection of clustering method is depends on the form of the data available for clustering k-mean, k-medoids and contribution based clustering algorithm based on the partitioned based clustering algorithm. These clustering strategies try to increase the intra-cluster similarity and decrease inter-cluster similarity[2].

#### 3.2 CONTENT BASED IMAGE RETRIEVAL

In image processing the primary things are the browsing, searching and retrieving images from big database. Retrieving images from big database is a difficult. Describing images with the aid of textual content is not an efficient approach.

It has drawbacks like there are large amount of images available and to explain them manually is not feasible and the prosperous features of images are not feasible to describe by using keywords, so to avoid these drawbacks CBIR is used. In content based image retrieval, color, texture and shape feature extraction is used to describe the images. CBIR with clustering optimize the performance of CBIR system [1] [11].



**Fig -1:** Content Based Image Retrieval System

In image processing, user chooses a query image by clicking on images. And then content of that entire query image will compare with the content of the database image. If similarity is occurred in the content of query image and database image then simply retrieve that image from the database. Classification or categorization is the preprocessing steps used to enhance the speed and accuracy of image retrieval. In image processing feature extraction is done by using color, texture and shape techniques

### 3.3 FEATURE EXTRACTION TECHNIQUES

Visual feature extraction is the basis of content based image retrieval approach. Features may include both text content based features and visual features (color, texture, shape). Within the visual function scope, the features may be in addition categorized as low-level features and high level features. To choose the features of images by using a CBIR system [11].

#### 3.3.1 COLOR FEATURE

Color is one of the maximum used visual features in content-based image retrieval. Color is a notion that depends on the response of the human visual system to light and the interaction of light with items. Color features are described by using color space. The different color space like that RGB, HSV, CIE L\*a\*b\*, CIE L\*u\*v\* and opponent color. The color features represented by different methods are Color histogram, color correlogram, color coherence vector, and color moment under a certain color space [11].

#### 3.3.2 TEXTURE FEATURE

Texture features depends on visual pattern. It contains information about structure of surface and relationship to surrounding environment. Texture features are represented by Tamura feature, Wold decomposition, SAR model, Gabor and Wavelet transformation [11].

#### 3.3.3 SHAPE FEATURE

Shape is an important visual feature. It contains geometric information of an item in image. Some shape features likes perimeter, area, peculiarity, symmetry, etc. Shapes features can be represented in two categories are boundary based and edge based. Shape features representation techniques are moment invariants, turning angles, Fourier descriptors, circularity, eccentricity, and major axis orientation and radon transform[11].

### 3.4 SIMILARITY AND DIS-SIMILARITY MEASURE TECHNIQUES

After feature extraction similarity dis-similarity measure is locate. To discover that similarity and dis-similarity measure by the use of Minkowski-form distance, quadratic form distance, Mahalanobis distance, Kullback-Leibler divergence and Jeffrey divergence. For image retrieval, the Minkowski and quadratic form distance are the maximum generally used techniques.

For indexing, dimension reduction is the important element. PCA, ICA, Karhunen-Loeve (KL) transform, and neural network are some techniques used for dimension reduction.

After dimension reduction indexing is done by using R-tree, R\*-tree, quad-tree, K-d-B tree these tree structure.

Improvement of CBIR results can be done by improving any particular area of CBIR like visual content description, similarity and dis-similarity measures, indexing scheme. Performance measures of CBIR: Recall, Precision, F-Measure are calculated to evaluate the results of k-mean and contribution based clustering algorithm [2].

$$\text{Precision} = \frac{\text{Total number of retrieved relevant images}}{\text{Total number of relevant images}}$$

$$\text{Recall} = \frac{\text{Total number of retrieved relevant images}}{\text{Total number of relevant images}}$$

$$\text{F - Measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{Precision} + \text{recall}}$$

#### 4. CONCLUSION

Searching the accurate images from huge amount of database is the crucial thing. Texture feature extraction with classification and clustering techniques are used for the efficient searching. Contribution based clustering method optimize the inter-cluster and intra-cluster similarities. It takes fewer pass to search the data. It gives the better result than k-mean [2][3]. By using contribution based clustering algorithm to form clusters of relevant images due to that processing speed is get improved. Contribution based clustering algorithm gives improved result than that of k-mean algorithm and by using contribution of data with texture feature will improve the result of CBIR.

#### 5. REFERENCES

- [1] D Kinoshenko, V Mashtalir and E Yegorova “Clustering method for fast content-based image retrieval”, Computer Vision and Graphics, Vol 32. pp. 946-952, March 2006.
- [2] Snehal Mahajan, D. R. Patil, “Image retrieval using contribution-based clustering algorithm with different feature extraction techniques”, IT in Business, Industry and Government (CSIBIG) IEEE conference Indore, 2014.
- [3] Narasimhan and Ramraj, “Contribution-Based Clustering Algorithm for Content-Based Image Retrieval”, International Conference on Industrial and Information Systems, Mangalore, India, 2010.
- [4] P.Sankara Rao, E Vamsidhar and GVP Raju, “ An approach for CBIR system Through multi layer neural Network”, International Journal of Engineering Science and Technology. Vol. 2(4).559-563, 2010.
- [5] Bradley, UM Fayyad and C Reina, “Scaling clustering algorithms to large databases”, .in Proc. 4th Int. Conf. Knowledge Discovery and Data Mining (KDD).pp. 9–15, 1998.
- [6] Shrivastava, K Upadhyay and R Bhati, “Comparison between K-Mean and C-Mean Clustering for CBIR”. IEEE International Conference on Computational Intelligence, Modelling and Simulation, Tuban, Indonesia, 2010.
- [7] Murthy, E Vamsidhar, JS Kumar, “Content based image retrieval using Hierarchical and K-means clustering techniques”, International Journal of Engineering Science and Technology, Vol. 2(3), pp.209-212, 2010.
- [8] Long et.al. “Fundamentals of content-based image retrieval”, Springer-Verlag, New York, pp. 1–26, 2003.
- [9] J. Han and K. Micheline, “Data mining concepts and techniques” , Morgan Kauffman. 2006.
- [10] Fazal-e-Malik and Baharum Baharudin, “Efficient Image Retrieval Based on Texture Features” ,.National Postgraduate Conference, Kuala Lumpur, Malaysia, 2011.
- [11] Amandeep Khokher, Rajneesh Talwar, “Content-based Image Retrieval: Feature Extraction Techniques and Applications”, International Journal of Computer Applications, PP. 9-14, 2012.