

Clustering for Content based Image Retrieval-A Survey

Vijay S Patil
Research Scholar
Department of E and TC
RCPIT, Shirpur, Maharashtra, India

P. J. Deore
Research Guide
Department of E and TC
RCPIT, Shirpur, Maharashtra, India

ABSTRACT

Clustering is the technique of classifying substance into sets of related or unrelated group of objects, basically Clustering is data analysis method for pattern recognition, feature extraction. Clustering perform very important task in CBIR to improve the accuracy in an image retrieval process.

General Terms

Pattern Recognition

Keywords

CBIR, Partitioning, Hierarchical Algorithms, Model Based.

1. INTRODUCTION

The term clustering is used in various research areas to describe the process of grouping unlabeled data? In the clustering process related and unrelated groups of data are cluster that would be helpful [1] for improve the sorting accuracy according to the image feature like colour, shape and texture. The aim of this paper is to survey the clustering ideas and methods helpful in the content based image retrieval process. Image retrieval is the process in which query image is provide as input and retrieved images from a large set of database. In fig.1 query image is applied to system after that the content of the images are compared with the available large number of data set accordingly image features. The system retrieves all images from the database that are similar in content with the query image. The techniques focus on the application of clustering to content-based image retrieval. A large collection of images is partitioned into a number of image clusters. The system retrieves all images from the cluster data that are similar in content with the query image. In distance based clustering similarity between two or more images depend upon they belong to same clustering group and how much close according to a given distance. Conceptual clustering is clustering in which two or more object or images groups having similar concept.

2. THE OBJECTIVE OF CLUSTERING

The objective of clustering is to determine the basic group in a set of unlabeled data, but always good clustering depends upon how to decide proper cluster and their group to find out similarity between clusters according to the user need.

3. NECESSITIES OF A CLUSTERING ALGORITHM

- Quantifiability.
- Dealing with special types of features.
- Determining clusters with random nature.
- Least requirements for region awareness to resolve input restriction.
- Ability to deal among noise.
- Tactlessness

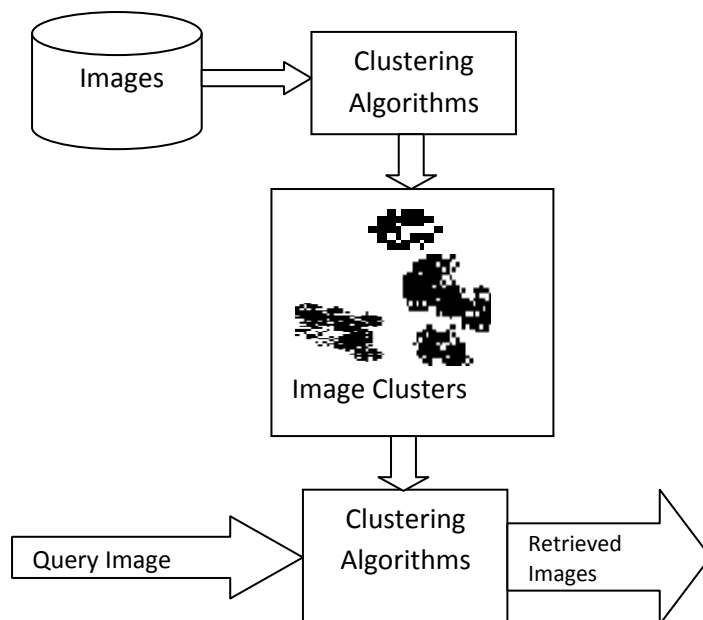


Fig.1 Clustering for CBIR

4. INTRICACY AMONG CLUSTERING ALGORITHMS

- Existing clustering methods does not attempt all the necessities effectively.
- Dealing with bulky amount of data items due to time complexity.
- The competence of the technique depends on the uses.

5. CLUSTERING METHODS

Fig.2 shows classification of clustering methods, the major clustering methods are partitioning, Hierarchical and model based.

5.1 Partitioning Methods

In this partitioning method constructs n number partitions of the data, in each partition represents a cluster. It differentiates the data into n number of groups which is jointly satisfying the following requirements are, one each group must contain either at least amount one entity or two each entity must belong to precisely one cluster. In [2] the direction of achieve desirable in partitioning-based clustering would require the meaning full possible partitions, most applications adopt one of a few popular methods such as one the k-means algorithm, where each cluster is represented by the mean value of the objects in the cluster and second the k-medoids [3] algorithm in which each cluster is represented by single of the objects situated near the inside of the cluster. K-medoids algorithm is [4] one of the most popular method for determined numbers of cluster as an input and the impact of initial value of cluster inside on clusters quality for similarity measurement.

5.2 Hierarchical Algorithms

In [5] cluster hierarchy indexing approach cluster is created by this method that would be useful and competent image retrieval. A hierarchical technique can be divided into two types one is agglomerative and another is divisive. The agglomerative method in which the clusters have sub-clusters and each object have its own cluster that clusters are sequentially combined up to the perfect cluster structure is obtained. In divisive method all the objects belongs to one cluster and then the cluster is divided into sub-clusters are sequentially separated and having sub clusters [6] to finding or improve the accuracy and reduce time complexity.

5.3. Model based methods

Mostly Maximum Likelihood Estimation is used in model-based clustering technique to obtain the perfect parameter inside the probability model. In [7] model-based clustering, data samples are represented by learning a mixture model usually a gaussian mixture model in which each mixture component represents a group or cluster. There are two types of learning methods for model-based clustering first is the mixture-likelihood in which the probability of each sample is a combination of all the probabilistic sample data, and another is classification-likelihood, in this method the probability sample data is produce only if they predict exact prediction of component. In this section, we study three examples of model-based clustering. 5.3.1 An [8] extension of the k-means partitioning algorithm, called Expectation-Maximization. Conceptual clustering is discussed in Section 5.3.2. A neural network approach to clustering is given in Section 5.3.3.

5.3.1 Expectation-Maximization

Maximization can be represented by a parametric probability

distribution. In the parametric probability distribution total data is a combination of these distributions that having individual distribution component of distribution for clustering. Content based image retrieval mainly contains two stages one is to signify an image and second is measure the dissimilarity between two images. Expectation-Maximization is a popular algorithm for clustering Gauss mixtures for the image illustration, but the insatiable character of EM make it durable to get a best possible model for content based image retrieval [9].

5.3.2 Conceptual Clustering

Conceptual clustering is a type of clustering use in machine learning that given a set of unlabeled objects which form a classified object. In this clustering technique finding attribute descriptions for all groups, where every group represents a concept or class of cluster. In[10] content based image retrieval technology is use to extract the image attribute of color, texture and shape of manufactured goods, calculated the goods cluster that having same feature of image content, after that this feature applied to the matching model, advice to the user in accordance with the users semantic perception.

5.3.3 Neural Network Approach

Neural network is locating link between input and output and each links has a weight or value associated with it. Neural networks have numerous properties that make them accepted for clustering, using [11] clustering based techniques with neural networks can improve the system performance on the basis of finding related cluster which reduce the retrieval time complexity, also [12] improve data processing in the multidimensional space requires soaring time and compute complexity by using neural network model.

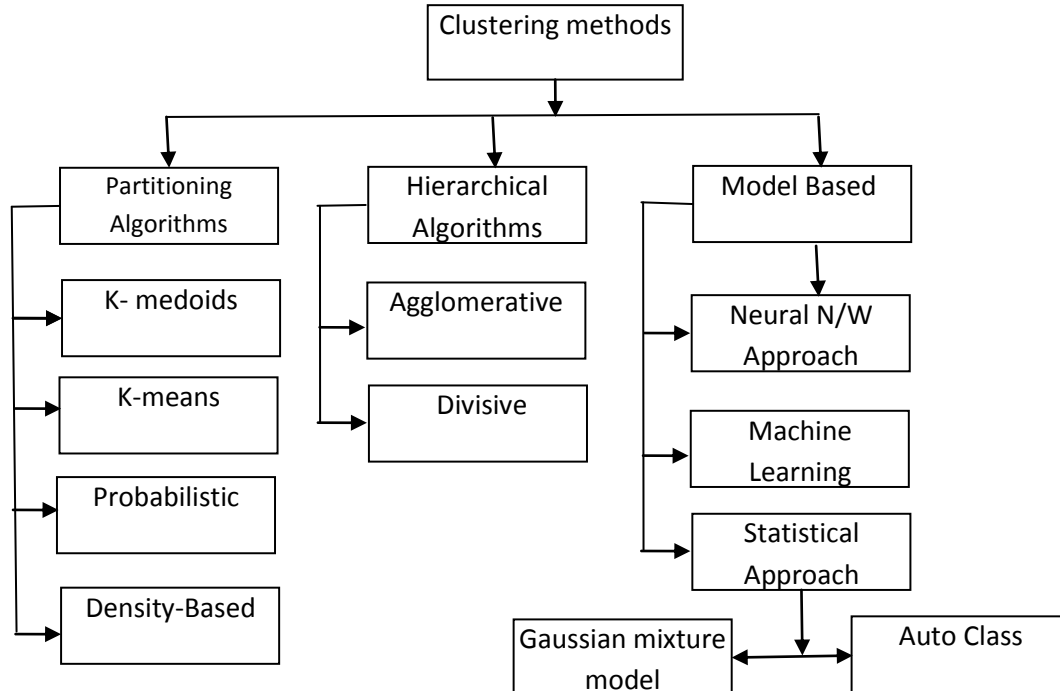


Fig.2 Classification of clustering methods

6. CONCLUSION

Clustering techniques can be helpful to finding a similarity between image (features like colour, texture and shape) in a collection of unlabeled data. Clustering method perform

crucial role in content based image retrieval to improve the accuracy and speed in an image retrieval process by reducing complexity to classified cluster in form of colour, texture and shape.

7. ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for insightful comments and helpful suggestions to improve the quality, which have been incorporated in this manuscript.

8. REFERENCES

- [1] A.K. Jain, M.N. Murtyin, P.J. Flynn, "Data Clustering: A Review", *ACM Computing Surveys*, Vol. 31, No. 3, 1999.
- [2] Kanungo Tapas, Mount D.M., Netanyahu N.S., Piatko C.D., Silverman R., Wu A.Y., "An efficient k-means clustering algorithm: analysis and implementation", *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol.24, no.7, pp.881,892, 2002.
- [3] Yuchai Wan, Xiabi Liu, Jie Bing and Yunpeng Chen, "Online image classifier learning for Google image search improvement", *IEEE International Conference on Information and Automation (ICIA)*, pp. 103 – 110, 2011.
- [4] Sabzi A, Farjami Y., ZiHayat M., "An improved fuzzy k-medoids clustering algorithm with optimized number of clusters", *Hybrid Intelligent Systems (HIS), 11th International Conference*, pp.206 – 210, 2011.
- [5] Hongli Xu, De Xu, Enai Lin, "An applicable hierarchical clustering algorithm for content-based image retrieval", *Computer vision/computer graphics collaboration techniques, 3rd international conference*, pp.82-92 Springer-Verlag Berlin, Heidelberg, 2007.
- [6] Jiawei Han. 2006, "Data Mining: Concepts and Techniques" Elsevier Inc.
- [7] Shih-Sian Cheng, Hsin-Chia Fu, Hsin-Min Wang, "Model-Based Clustering by Probabilistic Self Organizing Maps", *IEEE transactions on neural networks*, Vol.20, No. 5, 2009.
- [8] Khaled Alsabti, Sanjay Ranka, Vineet Singh, "An Efficient K-Means Clustering Algorithm", *Information Technology Lab (ITL) of Hitachi America, Ltd.*
- [9] Mingyan Yu , Haiyuan Liu, Yanhang Qiu, Ying Yan, "An improved EM algorithm for content based image retrieval", *Computer Science and Service System (CSSS), International Conference*, pp 2047 – 2050, 2011.
- [10] Guannan Zhao, Shijian Luo , Ji He, Ying Yan, "Style matching model-based recommend system for online shopping", *Computer-Aided Industrial Design & Conceptual Design, IEEE 10th International Conference*, pp. 1995 – 1999, 2009.
- [11] K L Du, "Clustering: A Neural Network approach", 2009 Elsevier, *Neural Networks* Vol. 23, No. 1, pp. 89-107, 2010.
- [12] Krakovsky, R., Forgac, R., "Neural network approach to multidimensional data classification via clustering", *IEEE 9th International Symposium on Intelligent Systems and Informatics*, pp 169 - 174, 2011.