Volume : 2 | Issue : 1 | Jan 2013 • ISSN No 2277 - 8179

Research Paper

Engineering

KEYWORDS : Image segmentation, clustering, k-mean algorithm, Pillar k-mean algorithm, adaptive k-mean algorithm.

Survey on Image Segmentation Using Different K-Mean Algorithms

Miki K. Patel PG Student, Department of Computer Engineering, Alpha college of Engineering, Khatraj, Gujarat, India.

Mitula H. Pandya Assistant Professor, Department of Computer Engineering, Alpha college of Engineering, Khatraj, Gujarat, India.

ABSTRACT

Image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Data clustering is a technique of data mining in which, the information which is logically similar is physically stored together. This paper presents the survey on image segmentation. For that the different k-mean clustering algorithms are described. The adaptive and pillar k-mean algorithms are found more efficient as those improve the segmentation quality in aspects of precision and execution time.

I. Introduction

Data mining is the process of extracting the information from a data set and transforming it into an understandable structure for further use. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records, unusual records and dependencies. Data mining is primarily used today by companies with a strong consumer focus - financial, retail, communication, and marketing Organizations. While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides link between the two. Data mining software analyzes relationships and patterns in stored transaction data base on open-ended user queries.

Clustering is the main part of the data mining. It is the task of assigning a set of objects into groups so that the objects in the same cluster are more similar to each other than to those in other clusters. Clustering is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify preprocessing and parameters until result achieves the desired properties. Clustering algorithms can be categorized based on their cluster model, as hierarchical clustering, centroid-based clustering, distribution-based clustering, density-based clustering etc.

Connectivity based clustering, also known as hierarchical clustering, is based on the core idea of the objects being more related to nearby objects than to farther objects. Such that, these algorithms connect "objects", to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. These algorithms do not provide single partitioning of the data set, but instead provide an extensive hierarchy of clusters that merge with each other at certain distances.

In centroid-based clustering, clusters are represented by central vector, which may not necessarily be a member of the data set. When the number of clusters is fixed to 'k' cluster, k-means clustering gives a formal definition as an optimization problem: find the ***k*** -cluster centers and assign the objects to the nearest cluster center, such that the squared distances from the cluster is minimized. The clustering model most closely related to statistics is based on distribution based clustering. Clusters can then easily define as objects belonging most likely to the same distribution. A nice property of this approach is that this closely resembles the way artificial data sets are generated by sampling random objects from the distribution.

In density-based clustering, the clusters are defined as areas of higher density than the remainder of the data set. Objects in these sparse areas that are required to separate clusters are usually considered to be noise and border points.

II. Basic Concepts

In data mining, k-means clustering is a method of cluster which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean value. k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes of image.

A set of observations (x1, x2,..., xn), where each observation is a d-dimensional real vector, k-means clustering partition the n observations into k sets (k ≤ n) S = {S1, S2, ..., Sk} so as to minimize the within-cluster sum of squares (WCSS):

$$\arg\min_{S} \sum_{i=1}^{K} \sum_{X_j \in S_i} \| X_j - \mu_i \|^2$$

Where μi is the mean of points in Si

The k-mean method randomly chooses k observations from the data set and uses these as the initial means. The random partition method first randomly assigns a k-cluster to each observation and then proceeds to the update step, thus the initial cancroids point of the cluster is assigned randomly. This method tends to spread the initial means out, while random partition places all of them close to the center of the data set. Euclidean is used as metric and variance is used as a measure of cluster scatter. The number of clusters k is an input parameter: an inappropriate choice of k may yield poor results. That is why, while performing k-means, it is important to run diagnostic checks for determining the number of clusters k in the data set.

Image segmentation is the process of partitioning of digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is easier and meaningful to analyze. Image segmentation is used to locate objects and boundaries in images. Image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. Application of image segmentation is medical imaging, locate objects in satellite images, face recognition, iris recognition, fingerprint recognition, traffic control systems, brake light detection, machine vision, agricultural imaging – crop disease detection.

III. Image Segmentation Using Clustering

In de-texturing and spatially constrained k-means approach[1] for image segmentation, first step relies on original de-texturing procedure which aims to converting the input natural textured color image into a color image without texture, that will be easier to segment. Once, this de-textured (color) image is estimated, a final segmentation is achieved by a spatially-constrained k-means segmentation is easy to implement and relatively fast compared to existing algorithm. This proposed method presents a simple, original and efficient segmentation procedure based on the k-means procedure. It overcomes the problems of this simple automatic clustering procedure, using spherical distributions on the difficult problem of the textured color image segmentation. First, the clustering is done on a simplified (de-textured) image, on which the search of well-separated cluster is easier. Implicitly, this procedure allows considering non-spherical clusters in the k-means clustering scheme, since the distribution of each textural feature vector is no more spherical after the spatial constraint. So its being simple to implement or to understand and having relatively fast execution time compared to the best existing algorithms.

In image segmentation of brain MRI using vessel segmentation method[2], does not require any user interaction, not even to identify the start point. Here seed points selected randomly which determines the main branches of the vessel algorithm. Random selection of seed points does not yield accurate segmentation of image. Accuracy of image segmentation can be improved by incorporating a priori information on vessel anatomy and let high level knowledge guide the segmentation algorithm. k-means algorithm is a popular clustering algorithm applied widely, but the standard algorithm selects K cluster from initial centroids, so that the large scale data can b handled.

Color Image Segmentation of using k-mean clustering method[3] divides into two processes. In first, pixels are clustered based on their color and spatial features, where the clustering process is accomplished. In second, the clustered blocks are merged to a specific number of regions. Using this two step process, it is possible to reduce the computational cost to avoiding feature calculation for every pixel in the image. Although the color is not frequently used for image segmentation, it gives high discriminative power of regions present in the image. This approach thus provides a feasible new solution for image segmentation which may be helpful in image retrieval by image segmentation. The experimental results clarify the effectiveness of our approach to improve the segmentation quality in aspects of computational time and precision result.

Spatial constrained k-mean approach to image segmentation[4] is a fast segmentation process in which first k-mean is applied on the image. Then, on image plane, the spatial constrains are adopted into the hierarchical k-means clusters on each level. The two processes are carried out iteratively and alternatively. An effective region merging method is proposed to handle over segmentation. The extensive experiments show the proposed approach is fast and generic, thus practical in applications. In hierarchical iteration, on each level, the spatial constrained region growing is carried out, whose results are used as a feed forwarded to the next level of k-mean algorithm. By this algorithm we handle over segmentation by combining color similarity with edge information, which is fast and generic.

Segmentation of medical image using k-mean clustering and marker controlled watershed algorithm[5], uses the conservative watershed algorithm for medical image analysis which does the entire division of the image. But its disadvantages are the over segmentation and sensitivity to false edges. For that here k-mean algorithm is used before using the watershed algorithm, to reduce the number of false edges and over-segmentation. The experimental results of the algorithm shows the above proposed method of using k-means clustering, which obtain a primary segmentation of MR brain images before applying the marker controlled watershed segmentation, which is effective as it reduces the amount of over segmentation.

Adaption of k-mean algorithm for image segmentation[6], tries to develop k-means algorithm to obtain high performance and efficiency. This method proposes initialization step in k-means algorithm. In addition, it solves a selection number of cluster by determining the number of clusters using datasets from images by frame size and the absolute value between the means. The additional convergence step added in k-means algorithm. Moreover, in order to evaluate the performance of the proposed method, the results of the proposed method and standard k-means and recently modified k-means are compared. The experimental results showed that the proposed method provides better output and increases the speed of the execution process. This paper shows the relationship between the frame sizes and the absolute values, which is used to estimate the number of cluster dependent on the values of pixels. The number of iteration of process affects its computational cost performance for reaching convergence.

Image segmentation using pillar k-mean algorithm[7], includes a new mechanism for clustering having the elements of high-resolution images in order to improve precision and reduce computational time. The system applies k-means clustering to image segmentation after optimizing by Pillar Algorithm. The pillar algorithm considers the pillars' placement which should be located as far as possible from each other to withstand against the pressure distribution of the roof, which is here identical to the number of centroids amongst the data distribution. This algorithm is able to optimize the k-means clustering for image segmentation in aspects of precision and computational time. It designates the initial centroids positions by calculating the accumulated distance metric between each data point and all previous centroids, and then selects data points which have the maximum distance as new initial centroids of image. This algorithm distributes all initial centroids according to the maximum distance metric and the experimental results clarify the effectiveness of our approach to improve the segmentation quality in aspects of precision and computational time.

Conclusion

After surveying the different k-mean clustering algorithms for image segmentation one can conclude that the pillar k-mean algorithm and adaptive k-mean algorithm are the best algorithms. These both algorithms are found efficient as they increase the computational time and improve the segmentation quality in aspects of precision and execution time.

REFERENCE

[1] Pattern Recognition Letters 32(2011)359-367, De-texturing and spatially constrained K -means approach for image segmentation | [2] International Journal Of Engineering trends and Technology, Image Segmentation of brain MRI using k-mean clustering method,Mrs.p.vijyalakhashmi | [3] International Journal Of Information and Communication Engineering, Color Image Segmentation of using k-mean clustering method,Anil H.Chitade | [4] Microsoft Research Area, Image Segmentation of brain MRI using k-mean clustering method, Ming Luo,Yu-Fei Ma,Hong-Jiang Zhang | [5] European Journal Of Scientific Research, segmentation of medical image using k-mean clustering and marker controlled watershed algorithm,M.C.Jobin Christ,R.M.S.Parvathi | [6] International Journal Of Information and Communication Engineering, Adaption of k-mean algorithm for Image Segmentation,Ali salem Bin samma and rosalina abdul salam | [7] International Journal Of Information and Communication Engineering,A new approch for image segmentation using pillar k-mean algorithm,Author:Ali Ridho Barakbah and Yasushi Kivoki |