TEST Engineering & Management

November/December 2020
ISSN: 0193-4120 Page No. 163-167

Image Segmentation using K-means clustering

Vijay Bhardwaj, Harjinder Kaur

Guru Kashi University, Talwandi Sabo

Article Info
Volume 83
Page Number: 163-167
Publication Issue:
November/December 2020

Abstract

Clustering is the process of dividing the datasets or objects into groups, consisting of similar data points. Points in the same group are as similar as possible. Points in the different groups are as dissimilar as possible. K-means clustering is one of the popular algorithms in clustering and segmentation that is used to separate objects from the surrounding background. K-means clustering treats each feature point as having a location in space. The basic K-means algorithm then arbitrarily locates, that number of cluster centers in multidimensional measurement space. Each pixel in the image is then , assigned to the cluster whose arbitrary mean vector is closest. The procedure continues until there is no significant change in the location of class mean vectors between successive iterations of the algorithms. However, the K-means algorithm is very sensitive in initial starting points. K-means generates initial cluster randomly. When random initial starting points close to the final solution, K-means has a high possibility to find out the cluster center. Otherwise, it will lead to incorrect clustering results. K-means clustering is a partitioning method. The function K-means partitions data into k mutually exclusive clusters and returns the index of the cluster to which it has assigned each observation. Each cluster in the partition is defined by its member objects and by its centroid or center.

Keywords: K-means clustering,

Article History
Article Received*: 25 October 2020*
Revised*: 22 November 2020*
Accepted*: 10 December 2020*
Publication*: 31 December 2020*

1. Introduction

Image segmentation is one of the techniques most used to correctly identify the pixels of an image in a decision Application driven[1]. It divides an image into several distinct regions, so that the pixels are extremely similar In each region, and high regional contrast In several fields, it is a powerful resource like health care, image processing, traffic signal, pattern recognition etc[2]. There are various image segmentation techniques including threshold-based, edge-based, cluster-based, neural network-based techniques. One of the most effective approaches from the various techniques is the clustering approach. Once again different forms of clustering exist: K-means clustering, Fuzzy C means clustering, mountain clustering method and subtractive clustering[3]. But for various optimal cluster numbers it gives different cluster results[4].

Nowadays the segmentation of images is one of the essential medical techniques where it is

used to isolate or area of interest from the context. Medical images are then segmented using various methods, and process outputs are used for further medical study[5]. But medical images are represented by the arrays of numbers in their raw form[6]. The number on the computer indicates the values of the relevant physical quantities showing the

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

contrast between the different types of body parts[5]. Medical image processing and analysis are useful for converting raw images into a quantifiable symbolic form, extracting relevant qualitative information to assist in diagnosis, and incorporating complementary data from multiple imaging modalities[7]. And one of the fundamental problems in the medical analysis is image segmentation which identifies the boundaries of objects such as organs or abnormal regions in images. Results from the segmentation make it possible for shape analysis, detecting volume change, and making a precise radiation therapy treatment plant.

2. Proposed K-Means Clustering Algorithm

Inputs are k and training data of different classes where k represents the number of clusters.Find the centroids of different classes separately i.e. if the set of points are x1, x2,…… xn then

Centroid

$$C_j(a) = \frac{1}{n}\sum X_i(a) \;\; for\; a = 1,2,.....d$$

Find the nearest centroid for each point xi by using Euclidean distance.

Euclidean distance: if p = (p1, p2,..., pn) and q = (q1, q2,..., qn) are two points in Euclidean space, then the distance (d) from p to q, or from q to p is given by

Assign each point to the cluster whose mean has the least squared Euclidean distance, this is the nearest mean. Repeat the same procedure for every iteration till the clusters formed.

Assign each point to the cluster whose mean has the least squared Euclidean distance, this is the nearest mean[2]. Repeat the same procedure for every iteration till the clusters formed. In computer vision, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). 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[8]. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images[5][9]. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image[10]. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic. Clustering refers to the grouping of samples process, so that the samples within each group are similar. These classes are called Clusters[11][8]. Clustering is a technique of data mining used in statistical analysis, data mining, pattern recognition, image analysis etc[12]. Different methods of clustering include hierarchical clustering

which constructs a hierarchy of clusters from individual elements. Different methods of clustering include hierarchical clustering which constructs a hierarchy of clusters from individual elements[7]. In partitional clustering; the objective is to build one collection of clusters that divide the data into similar classes[9]. Some clustering methods are distance-based according to which if two or more objects belonging to the same cluster are nearby a given distance, then clustering is called distance-based. In our work we have used the K-means clustering approach for performing image segmentation using Matlab software. A good clustering method will produce high-quality clusters with high intra-class similarity and low inter-class similarity. The quality of the results of the clustering depends on both the method's similarity measure and its implementation. It also measures the quality of a clustering method by its ability to discover some or all of the hidden patterns[7]. Image segmentation is the basis of image analysis and understanding, and is a crucial part of image processing and the oldest and hardest problem. Clustering means classifying and distinguishing things for which similar properties are provided[13]. The techniques of clustering classify the pixels with the same characteristics into one cluster, thus forming different clusters according to the coherence between pixels in a cluster. It is an unsupervised learning method and a common technique used in many fields, such as pattern recognition , image analysis, and bioinformatics, for statistical data analysis.

3. Experimental Results:

In this, we are going to segmentation of image data by using iterative nature. The random initialization of centroids at the start of the algorithm, different initializations may lead to different clusters since the k-means algorithm may be stuck in a local optimum and may not converge to the global optimum. Activity 1: In this we are taking the aerial image of NIT Rourkela, and apply image segmentation with K=2 and K=4, here Figure-1 represents the original image and Figures 2 and 3 represent the segmented image with K=2 and K=4 respectively. Note that K tells the number of clusters.

Figure-1: Original image of NIT Rourkela.

Figure-2: Segmented Image of NIT Rourkela with K=2.

Figure-3: Segmented Image of NIT Rourkela with K=4

Activity 2: In this we are taking the aerial image of Beach Image, and apply image segmentation with K=2 and K=4, here Figure-4 represents the original image and Figures 5 and 6 represents the segmented image with K=2 and K=4 respectively. Note that K tells the number of clusters.

Figure-4: Original image of the beach

Figure-5: Segmented Image of Beach with K=2.

Figure-6: Segmented Image of Beach with K=4.

Activity 3: In this we are taking the sample image (fewer pixels) and apply image segmentation with K=2 and K=4, here Figure-7 represents the original image and Figures 8 and 9 represent the segmented image with K=2 and K=4 respectively. Note that K tells the number of clusters.

Figure-7: Original sample image.

Figure-8: Segmented Image of Sample Image with K=2.

Figure-9: Segmented Image of Sample Image with K=4

4. Conclusion:

Segmentation is used for class quantification of Aerial images in which we can classify buildings, soil, water etc. here we can notice that K represents the clusters and classes which belong to each type of area, and we can calculate how much percentage of particular area occupied in that Aerial image. By using the K-means algorithm for segmentation of the image, one disadvantage is there. That is the number of computations for distance functions is more. To avoid this, the proposed (MEMBERSHIP K-MEANS) algorithm, in this method number of distance computations as well as time reduced and accuracy also improved. If the number of pixels of Image is increased then time for execution also increased. The two Arial images (more pixels) and one sample image (fewer pixels) are segmented by using K-means clustering algorithm with different K values.

References:

1. and A. T. Michailovich, Oleg, Yogesh Rathi, “"Image segmentation using active contours driven by the Bhattacharyya gradient flow.,” *IEEE Trans. Image Process.*, vol. 16.11, pp. 2787–2801, 2007.
2. F. jian. and M.R.Frater, “Threshold-based image segmentation through an improved particle swarm optimization,” in *Digital image computing techniques and applications (DICTA)*, 2012.
3. P. Y. and G. R. Reddy, “Segmentation of noisy images using improved distance regularized level set evolution,” in *IEEE International conference on circuits, power and computing technologies*, 2017.
4. and A. S. Jumb, Vijay, Mandar Sohani, “"Color image segmentation using K-means clustering and Otsu’s adaptive thresholding.",” *Int. J. Innov. Technol. Explor. Eng.*, vol. 3.9, pp. 72–76, 2014.
5. E. Deligiannidis, Leonidas, and Hamid R. Arabnia, “Emerging trends in image processing, computer vision and pattern recognition.,” in *Morgan Kaufmann*, 2014.
6. and S. B.-A. Al Bashish, Dheeb, Malik Braik, “‘Detection and classification of leaf diseases using K-means-based segmentation and.,’” *Inf. Technol. J.*, vol. 10.2, pp. 267–275, 2011.
7. and H. K. H. Wong, Stephen TC,

"'Design methods and architectural issues of integrated medical image data base systems.,'" in *Computerized Medical Imaging and Graphics*, 1996, p. 285–299.

8. K. V. and B. K. Singh, "An enhancement in adaptive median filter for edge preservation," in *ScienceDirect, Computer science procedia*, 2015, pp. 29–36.
9. P. Vijayalakshmi, "Segmentation of Brain MRI using K-means Clustering Algorithm," *Int. J. Eng. Trends Technol.*, pp. 113–115, 2011.
10. K. M. and Y. J. chanu N. Dhanachandra, "," Image segmentation using K-means clustering algorithm and subtractive clustering algorithm",, 2015, pp. 764-771.," in *ScienceDirect, Computer science procedia*, 2015, pp. 764–771.
11. and C.-C. C. Wu, Ming-Ni, Chia-Chen Lin, "'Brain tumor detection using color-based k-means clustering segmentation.,'" in *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, 2007, p. vol. 2.
12. G.N. Lee and H. Fujita, "K-means clustering for classifying unlabeled MRI data," *Digit. Image Comput. Appl.*, pp. 92–98, 2007.
13. and A. A. K. Chandhok, Chinki, Soni Chaturvedi, "'An approach to image segmentation using K-means clustering algorithm.,'" *Int. J. Inf. Technol.*, vol. 1.1, pp. 11–17, 2012.