EasyChair Preprint

№ 6018

Segmentation of Images Using K-Means Clustering Algorithm

Durganala Srenitha

July 5, 2021

IMAGE SEGMENTATION USING K-MEANS CLUSTERING ALGORITHM.

***Abstract*---** **In a decision-oriented application, image segmentation is one of the most commonly used methods for correctly classifying the pixels of an image. The process of partitioning an image into multiple segments is known as Image segmentation. An image is divided into distinct regions with high similarity between pixels in each region and high contrast between regions. Threshold based, edge based, cluster based, neural network based are some of the techniques used for image segmentation. From these different techniques one of the most efficient methods is the clustering method. K-means clustering, Fuzzy C-means clustering, mountain clustering method, and subtractive clustering method are all methods used for image segmentation using clustering. Here, in this project we use one of the most efficient method called K-means clustering algorithm to find a segmented image. Here we take number of clusters as input and image segmentation is done based on it. Image segmentation has become a popular technique in the medical field, where it is used to isolate a region of interest from a background image.**

Keywords— Image segmentation, k-means clustering algorithm, clusters, segmented image.

I. INTRODUCTION

A. Introduction

Image segmentation is one of the mostly used methods to classify the pixels of an image correctly in a decision-oriented application. divides an image into distinct regions with high similarity between pixels in each region and high contrast between regions. It's useful in a variety of fields, including health care, image processing, traffic image analysis, and pattern recognition.

K-means clustering is a common clustering algorithm. It is easier to use and quicker to compute than hierarchical clustering. It can also be used for a large number of variables. However, it generates different cluster results depending on the number of clusters. They are programmed to identify certain words and patterns that enable you to address preset responses.

Nowadays image segmentation becomes one of important tool in medical area where it is used to extract or region of interest from the background. So medical images are segmented using different technique and process outputs. The other type uses machine learning and artificial intelligence to provide the best response.

B. Objectives:

1. Here in this project, we take an image and number of clusters as input and generate an segmented image using k-means clustering algorithm.

2. This segmented image is further useful for analysis (like detecting diseases in human or plants).

II. RELATED WORK:

Many studies have been conducted in the field of image segmentation using various methods. And several of them are focused on various image segmentation applications. The K-means algorithm is one of the most basic clustering algorithms, and there have been several variations on how to initialize the core.

III. METHODOLOGY

A. K-means clustering algorithm

Clustering is a technique for dividing a collection of data into a set of categories. K-means clustering is one of the most widely used methods. The k-means clustering algorithm divides a collection of data into k

number groups. It divides a given set of information into k disjoint clusters.

K-means algorithm consists of two separate phases. In the first phase it calculates the k centroid and in the second phase it takes each point to the cluster which has nearest centroid from the respective data point. There are different methods to define the distance of the nearest centroid and one of the most used methods is Euclidean distance. Once the grouping is done it recalculate the new centroid of each cluster and based on that centroid, a new Euclidean distance is calculated between each center and each data point and assigns the points in the cluster which have minimum Euclidean distance. Each cluster in the partition is defined by its member objects and by its centroid.

The centroid for each cluster is the point to which the sum of distances from all the objects in that cluster is minimized. So K-means is an iterative algorithm in which it minimizes the sum of distances from each object to its cluster centroid, over all clusters.

Let us consider an image with resolution of x×y and the image has to be cluster into k number of cluster. Let p(x, y) be an input pixels to be cluster and ck be the cluster centers. The algorithm for k-means clustering is following as:

Step1: Initialize number of cluster k and centre.

Step-2: For each pixel of an image, calculate the Euclidean distance d, between the centre and each pixel of an image using the relation given below

$$d = ||p(x, y) - ck||$$

Step-3: Assign all the pixels to the nearest centre based on distance d.

Step-4: After all pixels have been assigned, recalculate new position of the centre using the relation given below.

$$V=\sum(xi - \mu i)2$$

Step-5: Repeat the process until it satisfies the tolerance or error value.

Step-6: Reshape the cluster pixels into image.

Implementation of Algorithm

1) Randomly select ‘c’ cluster centers.

2) Calculate the distance from each data point to the center of the cluster.

3) Assign a data point to the cluster center that has a minimum distance from the cluster center of all the cluster centers.

4) Recalculate the new cluster center using:

$$v_i = (1/c) \sum_{j=1}^{c} x_i$$

Where, 'c' reflects the sum of data points in.

5) The distance between each data point and the new cluster centers obtained is re-calculated.

6) Then stop if no data point has been reassigned, or repeat from phase 3.

Algorithm

k-means (k, D)

1.Choose random k data points as initial cluster mean (cluster centers)

2. Repeat

3. For each data point x from D

4. Compute the distance between x and each cluster mean(centroid)

5. Assign x to the nearest cluster

6. End for

7. Recompute the mean for the cluster collections

8. Until reaching stable clusters (current clusters means equals last clusters means)

Responses: The response is provided by the chatbot. It identifies from its source the necessary predefined answers. The most suitable response is then selected based on intent and context.

IV. Requirements:

A. Hardware Requirements:

Pentium pro-processor: Intel produced and manufactured Pentium Pro, the sixth generation x86 microprocessor was launched on November 1st, 1995. The P6 microarchitecture has been implemented and was initially designed in a whole variety of applications to replace the existing Pentium.

RAM 512MB or more: the minimum size of random-access memory is512 MB and maximum can be anything above that.

B. Software Requirements:

Language: Javascript, NodeJS

OS: Windows/Linux.

V. PROS AND CONS

PROS:

1. K-Means is comparatively fast algorithm when compared to other algorithms.

2. K-means algorithm generalizes the clusters into many shapes and sizes, generally into elliptical clusters.

CONS:

1. The quality of the final clustering results is depended on the arbitrary selection of initial centroid.

2. So if the initial centroid is randomly chosen, it will get different result for different initial centers. Therefore the initial center will be carefully chosen so that we get our desire segmentation.

3. And also computational complexity is another term which we need to consider while designing the K-means clustering. It relies on the number of data elements, number of clusters and number of iterations.

VI. RESULTS AND DISCUSSION:

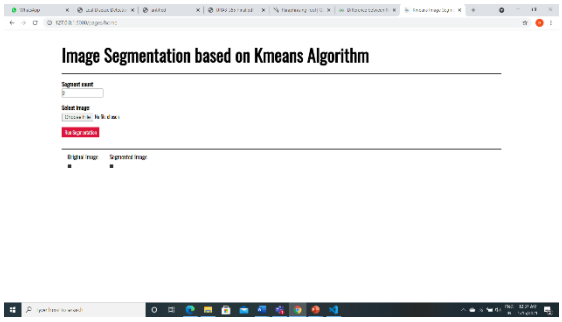
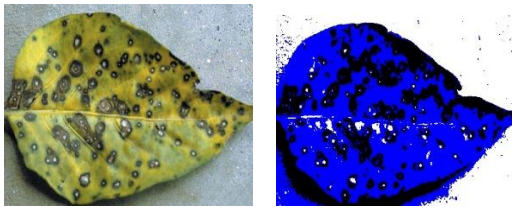


Image should be selected at choose file option and we need to give number of segments #segment_count which describes into how many segments the image should be segmented.



(a) (b)

(a) Original image.
(b) Segmented image.

VII. CONCLUSION:

We have segmented an image by using k-clustering algorithm, using subtractive cluster to generate the initial centroid. Although k-means has the great advantage of being easy to implement, it has some drawbacks. The quality of the final clustering results is depends on the arbitrary selection of initial centroid. So if the initial centroid is randomly chosen, it will get different result for different initial centers. So the initial center will be carefully chosen so that we get our desire segmentation. And computational complexity is another term which we need to consider while designing the K-means clustering

VII. REFERENCES:

[1] Nameirakpam Dhanachandra*, Khumanthem Manglem and Yambem Jina Chanu, “Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm”,2015.

[2] Ch. Usha Kumari, S. Jeevan Prasad, G. Mounika, “Leaf Disease Detection: Feature Extraction with K-means clustering and Classification with ANN”,2017.

[3] Atwell, E. (2005). Web chatbots: the next generation of speech systems? European CEO,November-December:142–144.