Indian Institute Of Information Technology,

Nagpur

Segmentation Using k-means Clustering

Submitted by - **Dharmendra singh roj(BT20ECE074)**

Guided by - **Tapan Jain** (Department of ECE)

- **Abstract-**

  segmentation using k-means clustering in image processing is a technique used to divide an image into different segments based on their visual similarity. In this approach, the k-means clustering algorithm is applied to the image pixels to group them into k clusters, where k is a predefined number of segments.

1. Introduction - Image segmentation is an essential task in image processing that involves dividing an image into regions with similar visual characteristics. This process is particularly useful for applications such as object recognition, image compression, and medical imaging. While several segmentation techniques exist, k-means clustering is a popular method that has proven effective in segmenting images.

K-means clustering is an unsupervised machine learning algorithm that partitions a set of data points into k clusters based on their similarity. In image processing, k-means clustering is used to group pixels in an image into clusters based on their visual features such as color, texture, and shape. The resulting segmented image is divided into regions that share visual similarities, allowing for easy detection of objects and features of interest.

In this poster, we explore the use of k-means clustering for image segmentation. We will discuss the various steps involved in the process, including image preprocessing, feature extraction, and cluster determination. We will also present the results of applying the k-means clustering algorithm to an example image and discuss potential applications of image segmentation using this method. Overall, this poster highlights the importance of segmentation in image processing and the effectiveness of k-means clustering as a segmentation technique.

2. K means clustering - The k-means algorithm is an iterative algorithm that starts by randomly initializing k cluster centroids. The algorithm then assigns each pixel to the nearest centroid, based on the distance between the pixel's visual features and the centroid's visual features. After assigning all the pixels to the nearest centroid, the algorithm recalculates the centroid's visual features based on the pixels assigned to that centroid. This process is repeated until the algorithm converges and the centroids no longer change.

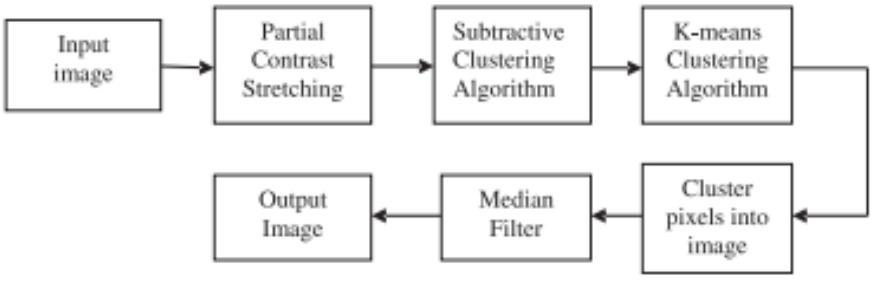
Once the k-means algorithm has converged, each pixel is assigned to the cluster with the closest centroid. This results in the image being segmented into k regions, with pixels in each region sharing similar visual characteristics. The number of clusters k is a parameter that must be defined before applying the algorithm, and finding the optimal value of k is an important consideration in the segmentation process.

K-means clustering has several advantages over other segmentation techniques. It is a simple and fast algorithm, making it suitable for real-time image processing applications. Additionally, it can be easily parallelized, making it possible to process large images quickly. However, k-means clustering also has limitations, including sensitivity to initial centroid locations and the need for prior knowledge about the number of clusters.

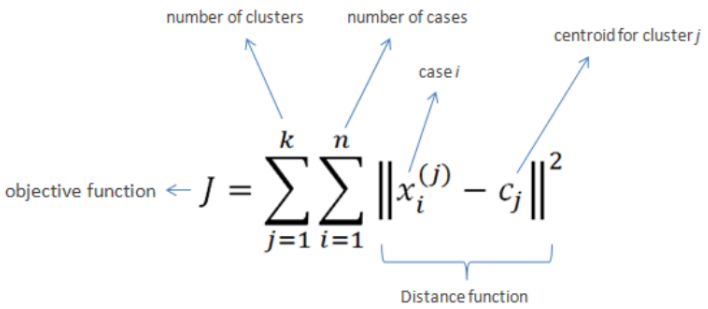
3 . Algorithm and preprocessing - In the k-means algorithm initially we have to define the number of clusters k. Then k-cluster center are chosen randomly. The distance between the each pixel to each cluster centers are calculated. The distance may be of simple Euclidean function. Single pixel is compared to all cluster centers using the distance formula. The pixel is moved to particular cluster which has shortest distance among all. Then the centroid is reestimated. Again each pixel is compared to all centroids. The process continuous until the center converges.

I. Give the no of cluster value as k.
II. Randomly choose the k cluster centers
III. Calculate mean or center of the cluster
IV. Calculate the distance between each pixel to each cluster center
V. If the distance is near to the center then move to that cluster.
VI. Otherwise move to next cluster.
VII. Re-estimate the center.

The proposed algorithm consists of partial contrast stretching, subtractive clustering, k-means clustering and median filter. Mostly the medical images which are used for segmentation have low contrast. So contrast stretching is used to improve the quality of the image. After improving the quality of image, subtractive clustering algorithm is used to generate the centers, based on the potential value of the image. Number of centre is generated based on number of cluster k. This centre is used as initial centre in k-means algorithm. Using the k-means algorithm, the image is segmented into k number of cluster. After the segmentation of image, the image can still contain some unwanted region or noise. These noises are removed by using the median filter



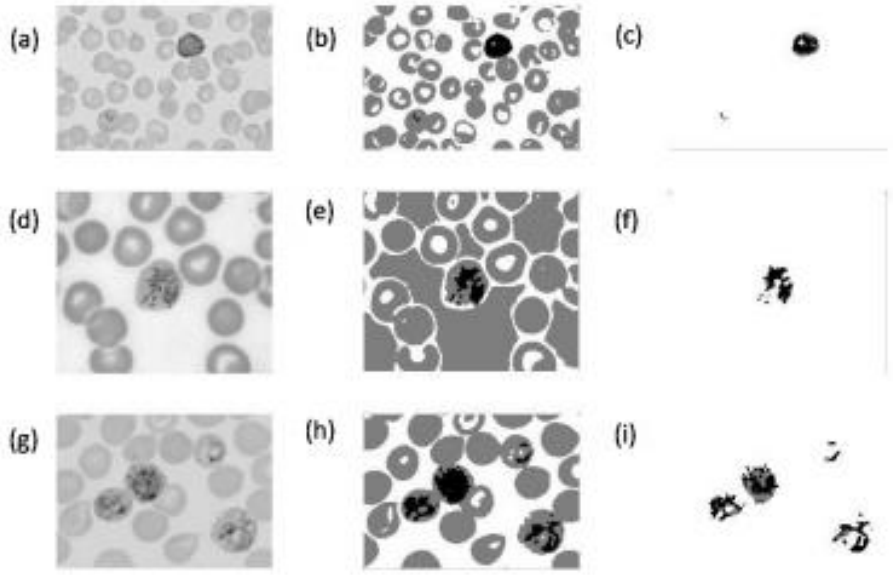
4. Mathematical Expression - The distance between the each pixel to each cluster centers are calculated. The distance may be of simple Euclidean function. Single pixel is compared to all cluster centers using the distance formula.

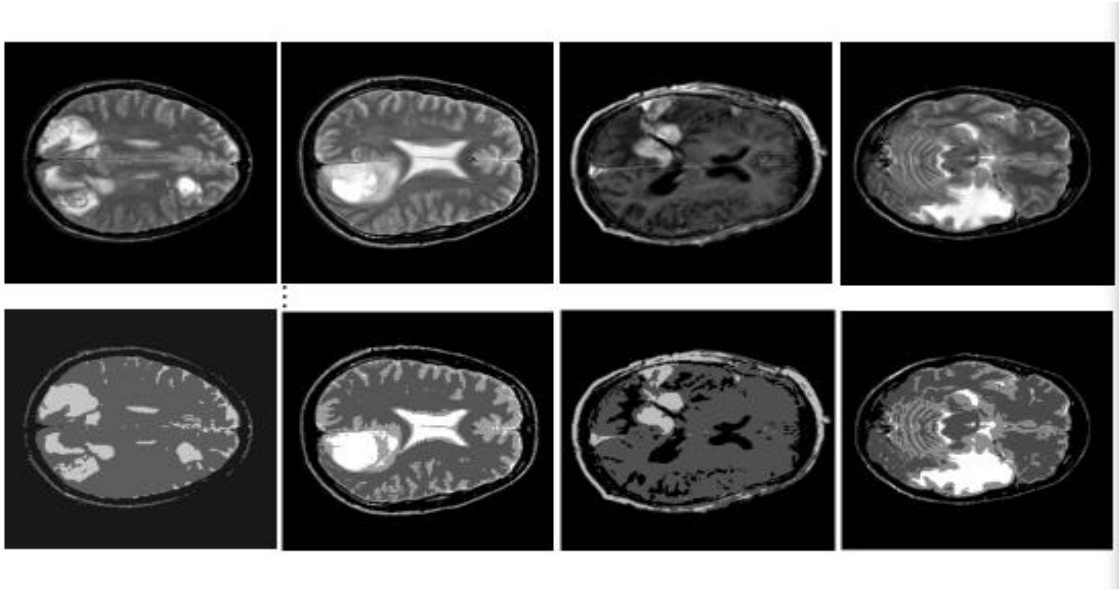


For a given cluster assignment C of the data points, compute the cluster means For a current set of cluster means, assign each observation as

$$C(i) = \arg \min_{1 \le k \le K} \left\| x_i - m_k \right\|^2, \; i = 1, \ldots, N$$

5. Implementation and Results - We used different types of medical images for the analysis. In medical image analysis, mostly the infected areas or area of interest are segmented from background. We used infected blood cell like malaria infected blood cell for the analysis. Matlab is used to implement the proposed algorithm.





6. Applications - Image segmentation using k-means clustering has a wide range of applications in various fields, including object recognition, image compression, and medical imaging.

In object recognition, segmentation using k-means clustering can be used to isolate an object of interest in an image, making it easier to detect and classify. For example, in autonomous vehicles, k-means clustering can be used to identify pedestrians or other objects on the road, allowing the vehicle to respond appropriately.

In image compression, k-means clustering can be used to reduce the size of an image by grouping similar pixels together and representing them with a single value. This reduces the number of pixels required to represent the image, resulting in a smaller file size. This technique is particularly useful in web applications where images need to be loaded quickly.

In medical imaging, segmentation using k-means clustering can be used to identify regions of interest in an image, such as tumors or other anomalies. This allows doctors to make more accurate diagnoses and plan treatment accordingly. For example, k-means clustering can be used in magnetic resonance imaging (MRI) to segment the brain into different regions for analysis.

Overall, the applications of image segmentation using k-means clustering are vast and varied. Its ability to identify regions of interest in an image based on visual features makes it a valuable tool in many fields, from computer vision to medical imaging.

7. conclusion - In conclusion, image segmentation using k-means clustering is a powerful technique that can be used to segment images based on their visual features. It is a simple and fast algorithm, making it suitable for real-time image processing applications. Additionally, its ability to identify regions of interest in an image has made it an important tool in various fields, from object recognition to medical imaging.

However, k-means clustering also has limitations, including sensitivity to initial centroid locations and the need for prior knowledge about the number of clusters. To overcome these limitations, various modifications to the k-means algorithm have been proposed, such as fuzzy k-means and spectral clustering.

Overall, image segmentation using k-means clustering is a valuable technique in image processing that has numerous applications in various fields. Its simplicity, speed, and effectiveness make it an attractive choice for segmenting images based on their visual features.