International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 03, 2021

Categorization Of Land Area Using K-Means And Fuzzy C-Means Clustering Algorithms

R Gayathri[1], T Perarasi[2], M Leeban Moses[3], S Giriprasad[4]

[1,2,3] Bannari Amman Institute of Technology, Sathyamangalam, TamilNadu.
[4] Vishveshwaraiah Institute of Science & Technology, Chittoor, Andhra Pradesh.

Email: [1]gayathrir@bitsathy.ac.in, [2]perarasi@bitsathy.ac.in, [3]leebanmoses@bitsathy.ac.in, [4]toogiriprasad@gmail.com

Abstract. ***Division and request of significant standard satellite imagery is a troublesome issue because it's important to finish this errand on a pixel-by-pixel premise. The fine spatial aim implies that each item is made up of a collection of pixels in close spatial proximity and precise order, necessitates that this perspective be considered without delay. For arranging high aim satellite symbolism, K-implies grouping calculation is a superior technique. k-implies, also known as Lloyd's computation, is an iterative information parceling calculation that assigns perceptions to one of several clusters defined by centroids, with k chosen before the algorithm begins. A basic distance choosing principle is used to arrange the eliminated areas. The technique effectively eliminates the blended pixel problem that plagues most pixel-based techniques. To group satellite symbolism into explicit articles within its boundaries and for ecological organizing reasons, we used K-implies and fuzzy C techniques grouping calculation. Clusters are formed from the various pixels. For each centroid, a center point is discovered. This centroid is used to classify the entire cluster. In this research, the k-means clustering technique and fuzzy c-means were used to separate the pixels of satellite images and process them using MATLAB software before naming the clusters based on their color configurations. This clustering algorithm was used to color-code the locations in the satellite image. For example, green denotes a forest, blue denotes a body of water, brown denotes a muddy terrain, and so on. Fuzzy C-means is independent of the original groupings, resulting in improved clustering outcomes.***

Keywords: Satellite image processing, K-means algorithm, fuzzy c-means algorithm.

1. INTRODUCTION

Satellite image processing has become one of the most prominent tools in the field of research and scientific process. It has been used since early 1970’s. The various applications of satellite image processing are classifications of land surface, flood mapping, deforestation mapping etc. [1]. The satellite images were very expensive and confidential which is used for military purposes. After the launch of many satellites, the satellite images were commercially used for analyzing earth surface, high resolution images to government and other organizations. This technology would be useful in analyzing of land mainly in the destruction area and the place where the humans can’t enter. The high-resolution images would be useful in the applications of transport mapping, disaster management, telecommunications etc. The Low-resolution images were used in the application of urban monitoring, forest management etc. [2]. In image classifications the pixels in the image are named with meaningful information for the better information about the environment.

Image classification classified into pixel based or object-based methods. In pixel-based classification, individual pixels are accompanied with spatial information and spectral patterns are used. In object-based classification, pixels are grouped into objects based on chosen similarities like color, texture, intensity, etc... and mixed pixel problem will not occur. Object based classification requires more information [7]. In this paper, the satellite images of land are processed and classified into empty land, water, greenery areas etc. This process is made by K means and fuzzy C means algorithm.

2. METHODOLOGY

a. Data Collection

The satellite image of land surface is taken through zoom earth website. This satellite images will be provided with area in terms of square kilometer

Types of Algorithms used:

i. K-means Algorithm

ii. Fuzzy C-means Algorithm

2.2 *K-means Algorithm*

The K-means algorithm is one of the most popular for clustering the image. It can be easily implemented with faster rate and accuracy. The main disadvantage is the initial partition selection [3]. In k means, k is a user parameter. Initially we assume that we have data with data points (D=X1, X2, Xn). First, we have to find the initial centroid for the input satellite image. Then repeat the first step to find more centroids based on the image's pixel density. Then map the data points to nearby centroid. Then compute the cluster center of each cluster. The above step is done until all data points of same pixel values are clustered.

Methodology of K-means:

- Place K focuses into the space addressed by the items that are being grouped. These focuses address beginning gathering.
- Assign each object to the gathering that has the nearest centroid.
- When all articles have been allotted, recalculate the places of the K centroid.

K-implies grouping is one of the significant solo learning issues which manages the assortment of constructions in unlabeled information. The meaning of bunching is the way toward social event objects into bunches whose components are comparable in a particular point [4]. Therefore, the clustering is the process of collecting objects which are similar in their own cluster and are different when compared with other clusters.

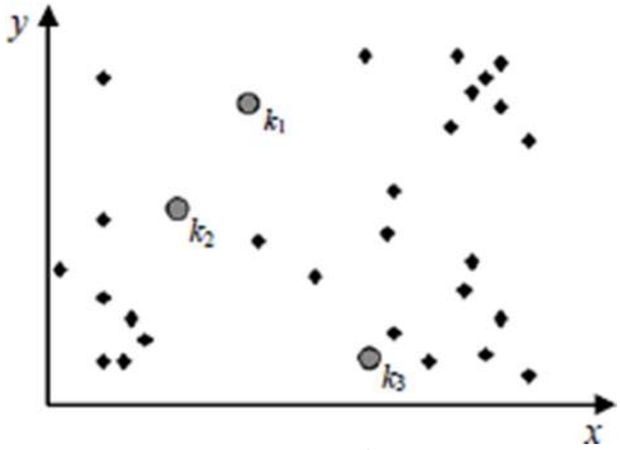

Fig. 1. Three centroids formed initially

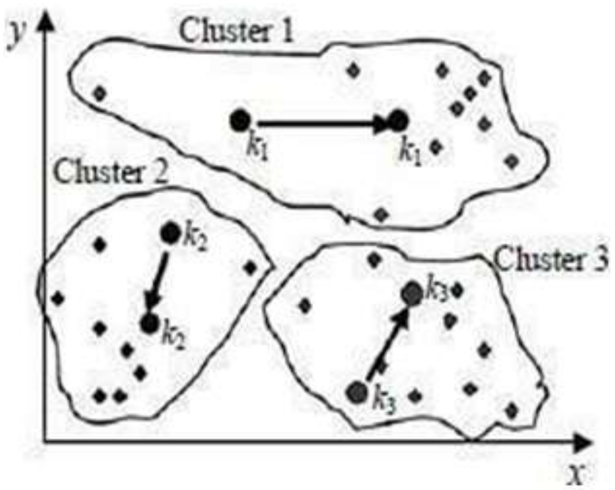

Fig. 2. Three more centroids

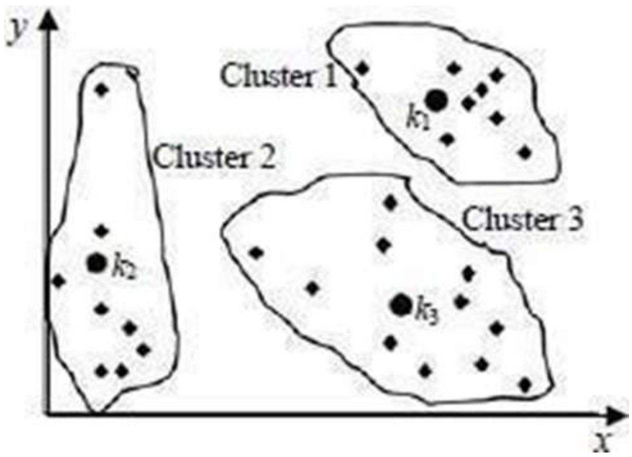

Fig. 3. Final cluster and their centroids.

Limitation of K-means:

- The mathematical calculation of k means is expensive. The time taken for computation is depends on number of iterations, number of cluster and data.
- The initial centroid selection determines the quality of cluster and reduces the number of centroids used.

• Empty clusters formed due to the fixed clusters will result in the starting problem of algorithm.

2.3 Fuzzy C-means Algorithm

The fuzzy C means is one of the widely use algorithm in fuzzy clustering. It is introduced by J. C. Dunn in 1973 based on the concept of fuzzy c partition. The fuzzy c-means is summarized as X= {x1, x2…, xn}.

$$M_{fcn} = \{U \epsilon U_{cn} \ : \ u_{ik} \epsilon [0.1] \quad \sum_{i=1}^{c} u_{ik} = 1, \qquad 0 < \sum_{k=1}^{n} u_{ik} < n\}$$

Here U_{cn} represent the set of real cxn matrices. The fuzzy C mean is similar to the K means algorithm.

The algorithm of Fuzzy C means follows,

• The number of clusters from the given input images is given by the programmer. The coefficients for each data point are given randomly. Due to these coefficients the data point will be located inside the cluster.

• Repeat the above process until all the data points are aligned inside the clusters.

• Find the centroid for each cluster.

$$c_k = \frac{\sum_x w_k(x)^m x}{\sum_x w_k(x)^m}$$

The coefficients of each data points inside the clusters should be computed after finding its centroids. The centroid can be found by using the formula here m is the hyper-parameter which controls the fuzzy cluster [6]. The higher of m will be in the end of the cluster.

3. EXPERIMENTAL RESULT

a. K-means Algorithm

In k-means algorithm the centroid is formed first. The data points in the input image are assigned to the nearby centroids. The first output image shows the data points that are nearer to a centroid value and it is shown. The second image shows the data points of other centroid in a cluster and it goes on for n number of clusters formed.

For the first iteration nearby data points moves towards the nearby centroid. Then it slowly gets closer to the centroid when the iteration increases. Thus, it forms a cluster of similar data points or pixels. In this kind of algorithm, the data points cannot have membership in two or more clusters.

Fig. 4. Input satellite image

Fig. 5. The vegetation area

Fig. 6. The empty land

Fig. 6. The water area

b. Fuzzy C-means Algorithm

This calculation works by relegating the information focuses to the relating group dependent on the distance between the bunch community and the information point. In this clustering algorithm the data points can have membership in two or more clusters [5]. As this algorithm undergoes many iterations the data points may vary from cluster to cluster to get perfection.

The cluster center and membership of data points are updated after each iteration. The initial centroid determines the cluster formation. The output image shows the partitioned data points after several iterations. Each data points in that image have centroid with respect to its distance between the centroid an itself.

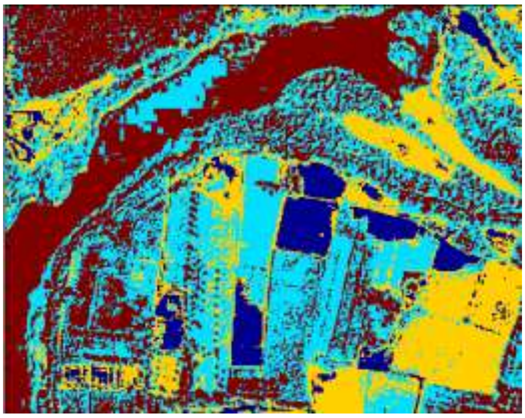

Fig. 6. Output image using fuzzy c-means

4. CONCLUSION

Clustering is one of the most important aspects in data mining because it has numerous numbers of applications in data evaluation, summarization, customer segmentation etc. Many numbers of clustering methods like fuzzy clustering, density-based clustering and model-based clustering etc. which are used for examining the data analysis, clustering analysis. In industries online streaming of data and data processing are very challenging, therefore the use of clustering algorithm should be recommended. By the above method used, both K means and Fuzzy c means are best in their aspects of the data given. K means will produce a better result on the large dataset while the fuzzy c means will produce better results for noisy cluster dataset. Since the technology is technology is developing day by day, the data is also going on diverse, so the algorithm which handles large data and which can overcome noise filled clustering data is highly recommended.

5. REFERENCES

[1]. P.Suganya, N.Mohanapriya, B.Kalaavathi,"Satellite image resolution enhancement using multi wavelet transform and comparison of interpolation techniques", International Journal of Research in Engineering and Technology, eISSN: 2319-1163 pISSN: 2321-7308, Volume: 03 Special Issue: 07 | May2014.

[2]. Arya P Unni, "Satellite Image Enhancement Using 2D Level DWT", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 3 Issue3, March 2014.

[3]. Lasmika, K. Raveendra, "Improving Quality of Satellite Image by Wavelet Transforming & Morphological Filtering", International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319- 8753, Vol.3, Issue 7, July 2014.

[4]. Wenkao Yang, Jing Wang and Jing Guo, "A Novel Algorithm for Satellite Images Fusion Based on Compressed Sensing and PCA", Hindawi Publishing Corporation, Mathematical Problems in Engineering, pp. 1 – 10, Volume 2013.

[5]. Beucher, S., and F. Meyer. 1992. "The Morphological Approach to Segmentation: The Watershed Transformation." In Mathematical Morphology in Image Processing, edited by E. R. Dougherty, 433. New York: Marcel Dekker.

[6]. Grau, V., A. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield. 2004. “Improved Watershed Transform for Medical Image Segmentation Using Prior Information.” IEEE Transactions on Medical Imaging 23 (4): 447–458. doi:10.1109/TMI.2004.824224.
[7]. Li, D., G. Zhang, Z. Wu, and L. Yi. 2010. “An Edge Embedded Marker-Based Watershed Algorithm for High Spatial Resolution Remote Sensing Image Segmentation.” IEEE Transactions on Image Processing 19 (10): 2781–278.