COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING
e-ISSN: 2582-7464

Computational Intelligence and Machine Learning
Vol-3, Issue-2, October 2022

Satellite Image Processing Using Fuzzy Logic and Modified K-Means Clustering Algorithm for Image Segmentation

Dewang Mehta [1*], Geerisha Jain [2], Dr. Anto S [2]

[1] Associate Consultant, Business Technology Solutions at ZS, Pune, Maharashtra, India
[2] Computer Science Engineering, Vellore Institute of Technology, Vellore, India
[3] Professor, Department of Computational Intelligence, Vellore Institute of Technology, Vellore, India
*Corresponding Author Email: dewang_mehta@hotmail.com

Abstract

Satellite images are useful in providing a real time dynamic picture of the earth and its environment. The large assemblage of remote sensing satellites orbiting the earth provide an extensive and periodic coverage of the planet through the capture of live images round the clock, in turn enabling numerous uses for the benefit of mankind. In the field of satellite image processing, image segmentation is one of the vital steps for extracting and gathering huge amount of information from the satellite images. The basic k-means clustering algorithm is simple and fast in terms of dealing with the required segmentation, but the limitation associated with this clustering is its inability to produce the same result for every run, as the resulting clusters depends on the initial random assignments. In this paper, an enhanced modified k-means clustering algorithm is proposed for the effective segmentation of the satellite images with an objective to overcome the demerits of the traditional k-means by combining fuzzy logic with the membership function. The proposed methodology continuously produces the same result for each run. As an outcome, the experimental results proved that the enhanced k-means algorithm is an effective and more efficient process for the precise and accurate segmentation of satellite images.

Keywords

Clustering, Fuzzy logic, Image Segmentation, K-Means, Satellite Imagery.

INTRODUCTION

In recent years, along with the breakthrough of machine sciences, image processing has become a field of interest. Image processing has been studied extensively and there have been many applications in practice: as in Medicine, detection and identification of tumor, radiographic imaging, identification of the vascular pathway from the X-ray imaging. One of the many applications cover, perhaps satellite imagery processing to calculate urbanization speed and tightening deforestation and forest cover.

Satellite image processing to calculate the speed of urbanization, forest erosion, deforestation on an area, according to traditional methods from the past was done using drawings, maps, and field trips, but it could not be carried out in areas where human land cannot be set up to check the measurements. Today with modern scientific instruments, we can obtain images from anywhere on the Earth. Through the satellite images captured of a forest, the trial image segmentation (using the K-means algorithm) into different areas, can help us calculate the percentages of changes in the area as depicted by the images over the years. This can in turn aid in concluding urbanization speed, forest cover rates, deforestation rates, change over the period and much more.

The proposed methodology is based on Image segmentation which deals with objects dividing the image into regions that can be considered homogenous according to the most standard feature, such as color, texture, text etc. The result of image segmentation is a set of common areas covering the whole image, or a set of lines mined from the image. Each pixel in the pixel set of a region is similar to each other with attention to some properties or computational attributes such as color sharpness, intensity, and texture. Segmentation techniques are divided into 3 groups: Clustering, edge detection, growing region. Some common clustering algorithms such as K-means is commonly used in image segmentation. Segment of images refers to the process of partitioning a digital image into several different clusters (the pixel set). In this paper, we will be discussing K-Means algorithm, its merits, demerits and proposing a new modified approach to K-Means overcoming the limitations of the traditions K-means model for satellite image processing.

RELATED WORK

There are many ways to modify and improve the traditional K-Means clustering algorithm. It consists of four basic steps: (i) initialization of the clustering, (ii) classification on the data member, (iii) computational stage and (iv) convergence condition. As compared to other stages, initialization stage of the KMeans clustering is the one in which most of the modifications and improvement have been performed. The detailed survey of the previous work on modifications on the K-Means clustering algorithm is shown in Table 1 below.

Table 1. Prior Related Work on K-Means Clustering Algorithm

Reference, Author, Year	Modifications on the traditional K- Means Clustering Algorithm
[1] J.B.McQueen, 1967	Suggested a learning strategy to determine a set of cluster seeds for segmentation.
[2] J.Tou and R.Gonzales, 1974	Proposed simple cluster seeking method.
[3] Y. Linde, A. Buzo, R. M. Gray, 1980	Proposed binary splitting method for segmentation.
[4] G. P. Babu, M. N. Murty, 1993	Suggested method of genetic programming based near optimal seed selection. In this, the selection of the size of the population, mutation and crossover probabilities influence on the result
[5] C.Huang, R. Harris, 1993	Based on principle component analysis, proposed the direct search binary splitting method
[6] I. Katsavounidis, C. C. J. Kuo, Z. Zhen, 1994	Suggested a method starts with selection of a point as the first seed on the edge of the data. The point which is furthest from first seed is consider and selected as the second seed.
[7] M. B. A. Daoud, S. A. Roberts, 1996	Introduced a method to divide the entire data into two different groups and the points are randomly distributed in the group.
[8] P. S. Bradley, U. M. Fayyad,1998	Proposed a method for choosing the most centrally located instance as the first seed.
[9] A. Likas, N. Vlassis, J.J. Verbeek, 2003	Proposed a global K means algorithm in which gradually increase the number of seeds till K is found that is till convergence
[10] S. S. Khan, A. Ahmad, 2004	Introduced a centroid initialization method based on a density- based multi scale data condensation. In this method, the density of the data at a point is estimated, and then sorting the points based on their density.

METHODOLOGY

K-means Clustering

K-Means is a very important algorithm and is commonly used in clustering techniques. The K-means algorithm takes the input parameter K and divides a set of n objects in K clusters, the homology in clusters is high while homology out of clusters is low. Cluster homogeneity is measured when evaluating the mean value of these objects in the cluster which then can be observed as the "focus" of the cluster. The algorithm handles the following: it randomly selects one K object, one out of these points represent a cluster average or cluster data center. For the objects left, each object will be assigned to the cluster that it is most closely based on distance between object and cluster mean. Then the average cluster will be recalculated for each cluster. This treatment will be repeated until the standard convergence function jars.

The purpose of the K-means algorithm is to generate K data sets {C 1 , C 2 , ..., C k } from a data set containing n objects in d space X i = {x i1 , x i2 , ..., x id }, i = 1 ÷ n, such that the standard function is given by (1):

$$E = k \text{ i= 1} \sum x \in C \quad |x - mi|2\mathbf{min} \qquad \textbf{(1)}$$

'|x - mi|' is the Euclidean distance between Data point x and center

$\{m1, m2, m3, , mk\}$ mi is the center of the cluster C

$\{C1, C2, C3,, Ck\}$ is a set of clusters

'k' is the number of clusters, cluster centers.

The Algorithm for traditional K-means is given below.

Begin

1. Initialization:

Randomly selects the centroid for the K cluster. Each cluster is represented by the center of the cluster.

2. Distance calculation:

Calculate the distance between objects to K hearts (usually Euclidean distance) using (1):

$$Dk \sqrt{\sum_{i=1}^{n}(xi - mj)} \qquad \textbf{(2)}$$

For each point (1 ≤ i ≤ n), calculating its distance to each center (1 ≤ j ≤ k). Then find the closest focus for each point and group them into nearest groups.

3. Update focus:

For every 1≤ j ≤ k, update the cluster focus mj by identifying Averages plus data object vectors using (2).

$$vj = (1/cj) \sum_{i=1}^{Cj} xi \qquad \textbf{(3)}$$

here, ' cj ' represents the number of data points in jth cluster.

4. Reassign the points near the new group center

Group the objects into the nearest group based on the focus of the group.

Stop condition: Repeat steps 2 and 3 until there are no group changes of objects

End.

Limitations of Traditional K-means Clustering

The K-means algorithm is proved to be convergent and has a computational complexity of O (tkn) where t is the number of iterations, k is the number of clusters, n is the number of objects of the data set. Usually, k << n and t << n end at a local optimization point.

However, the disadvantage of K-means is that it is very sensitive to interference and extraneous elements in the data. Moreover, the quality of data clustering of the K-means algorithm much depends on the input parameters such as the number of K-clusters and K-centered initialization boards. In the case of the initial initialization focus that is too deviated from the weights, the cluster result of K-means is very low, that is, clusters of data is very different from the actual clusters. Hence, there is not one optimal solution to choose input parameters, the most commonly used way out is to experiment with different K input values and then select the best solution.

Proposed Modified K-means Clustering

The K-Means clustering algorithm is widely used for the segmentation of various images such as medical or satellite images for its fast convergence and simplicity. However, the application and performance of the K-means clustering algorithm is still limited due to several disadvantages as indicated in Section II. In this section, modifications on the conventional KMeans clustering algorithm are introduced to overcome the disadvantages and weakness and improve the segmentation performance. For the modification on the K-Means clustering, consider an image which has N data that have to be clustered into n number of centers. Let Xi be the i-th data and Cj be the j-th center with predetermined initial value where i = 1,2,3,4...,N and j = 1,2,3,4,...,n . In this paper, the concept of fuzzy logic is introduced to modify the k-means clustering algorithm. In fuzzy logic, each member has varying membership contrast to crisp logic wherein each member has clearly defined boundary (its membership strictly either 0 or 1). When fuzzy logic applied to the image, each data member can be assigned simultaneously to more than one cluster or group with different degree of membership. The above mentioned process of fuzzy based approach can be obtained based on the membership function as given by (2)

$$m_{ck} = \frac{1}{\sum_{j=1}^{i}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{1}{q-b}}} \quad \textbf{(4)}$$

where dik is the distance from point k to the current centroid djk is distance from point k to other centroid j and q is the fuzziness exponent where the typical value is 1. After assigning the membership for each data in the image, then we have to apply the fitness calculation process for all the data member using (3)

$$F(c_j) = \sum_{k_{ij}}(x_i - c_j)^2 \quad \textbf{(5)}$$

As part of processing the satellite images, we need to consider the color component of the images as the real time images are multichromatic and not monochromatic.

The basic purpose is to segment the colors automatically using the L * a * b * color space and K-means clustering. The whole proposed methodology of processing satellite images can be summarized in the following steps:

Step 1: Read the image

Step 2: Convert the image from RBG color space to L * a * b *color space. The space L * a * b* consists of a 'L' layer, a '*' color layer, and a 'b *' color layer. All color information is in layers '* a' and 'b *'. The difference between the two colors can be measured using the standard Euclidean distance.

Step 3: Color Classification in Space "* a * b *" uses K-means clustering.

Step 4: Label each pixel in the image using the result from K-means. For all objects in the input , K-means returns an index corresponding to a cluster. Label each pixel in the image with its cluster index.

Step 5: Create images that are segmented by color. Using the pixel labels, divide the object in the image by color, which will result in three different images.

Step 6: Segment the kernel into a separate image. Then, the output determines that the clustered indexes contain blue objects because K-means will not return the same cluster_idx value each time. This can be done using the cluster center values, rather than a 'a *' and 'b *' values for each cluster.

Assume that the surface color of the objects in the image is a constant property and that the color is mapped to a 2D space and color. Then, using the K-means clustering algorithm, the cluster of colors has a set of similar pixels. After segmenting the image, each pixel belongs to a single region. These unique areas usually correspond to the whole or part of the actual objects in the image.

EXECUTION RESULTS

The database consisting of satellite images was created to test the proposed algorithm. The proposed algorithms was coded in MatLab 7.10(R2010a) and executed in Intel core i3 system with 2GB RAM. The performance of this algorithm is assessed with both quantitative comparison and visual judgement as shown in Figure 1 and Figure 2.

e-ISSN: 2582-7464

Computational Intelligence and Machine Learning
Vol-3, Issue-2, October 2022

Figure 1. Test Satellite Image

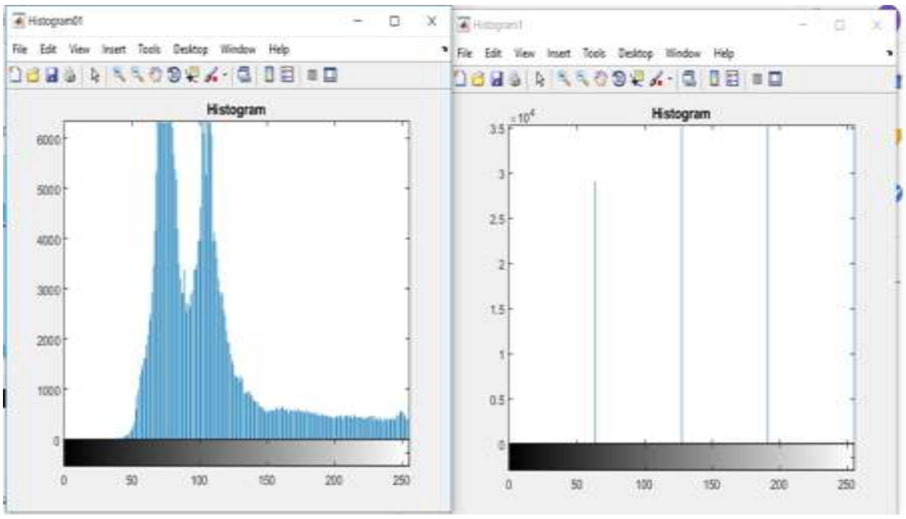

Figure 2. Before and after histogram of the test satellite image

CONCLUSION

The modified K-Means clustering algorithm for satellite image segmentation was proposed in this paper. The characteristic of the Fuzzy logic is applied to modify the membership function of the traditional K-Means Clustering algorithm. A number of test satellite images were segmented using the proposed algorithm and experimental results proved that the modified algorithm was an effective method for the segmentation of satellite images. This algorithm could segment object precisely, decrease the segmentation time, and enhance the segmentation effect.

REFERENCES

[1] J. B. MacQueen, "Some methods for classification as well as analysis of multivariate observation", In: Le Cam, L.M., Neyman, J. (Eds.), University of California, 1967.

[2] J. Tou, R. Gonzales, "Pattern Recognition Principles", Addison-Wesley, Reading, MA., 1974.

[3] Y. Linde, A. Buzo, R. M. Gray, "An algorithm for vector quantizer design", IEEE Trans. Commun. 28, 1980, 84– 95.

[4] P. S. Bradley, U. M. Fayyad, "Refining initial points for Kmeans clustering", In: Proc. 15th International Conf. on

Machine Learning. Morgan Kaufmann, San Francisco, CA, 1998, pp. 91–99.
[5] C. Huang, R. Harris "A comparison of several codebook generative approaches", IEEE Trans. Image Process. 2 (1), 1993, 108–112.
[6] I. Katsavounidis, C. C. J. Kuo, Z. Zhen, "A new initialization technique for generalized lloyd iteration", Signal Process. Lett. IEEE 1(10), 1994, 144–146.
[7] G. P. Babu, M. N. Murty, "A near optimal initial seed value selection in K-means algo using a genetic algorithm", Pattern Recognition Lett. 14 (10), 1993, 763–769.
[8] M. B. A. Daoud, S. A. Roberts, "New methods for the initialization of clusters", Pattern Recognition Lett.17 (5), 1996, 451 45.
[9] A. Likas, N. Vlassis, J. J. Verbeek, 2003. "The global K-means clustering algorithm", Pattern Recognition 36, 451–461.
[10] S. S. Khan, A. Ahmad, 2004 "Cluster center initialization algorithm for k means clustering", Pattern Recognition Lett. 25 (11), 1293–1302.
[11] Bo Zhao, Zhongxiang Zhu, Enrong Mao and Zhenghe Song "Image Segmentation Based on Ant Colony Optimization and K-Means Clustering" Proceedings of the IEEE International Conference on Automation and Logistics August 18 - 21, 2007, Jinan, China.
[12] J. Pérez, R. Pazos , L. Cruz, G. Reyes, R. Basave, and H. Fraire "Improving the Efficiency and Efficacy of the K-means Clustering Algorithm through a Convergence Condition", Gervasi and M. Gavrilova (Eds.): ICCSA 2007, LNCS 4707, Part III, Springer-Verlag Berlin Heidelberg 2007, pp. 674–682.
[13] S. J. Redmond, C. Heneghan, "A method for initializing the K-means clustering algorithm using kd - trees. Science direct", Pattern Recognition Letters 28 (2007), 965–973.
[14] Nor Ashidi Mat Isa,Samy A. Salamah,, Umi Kalthum Ngah "Adaptive Fuzzy Moving K- means Clustering Algo for Image Segmentation " , IEEE Transactions on Consumer Electronics, Vol. 55, No. 4, Nov 2009.