ISSN No. 0976-5697
Volume 4, No. 5, May 2013 (Special Issue)
International Journal of Advanced Research in Computer Science
TECHNICAL NOTE
Available Online at www.ijarcs.info

Application of Modified K-Means Clustering Algorithm for Satellite Image Segmentation based on Color Information

Ganesan P
Research Scholor
Sathyabama University
Chennai, India
gganeshnathan@gmail.com

Dr V.Rajini
Dept of Electrical and Electronics Engg
SSN College of Engineering
Chennai, India
rajiniv@ssn.edu.in

Abstract: The Image segmentation is the process of clustering or partitioning the image into number of sub images based on any one characteristics of the image such as colour, intensity or texture. The segmentation is the one of the middle level important process in the image analysis. The number of segmentation algorithms has been developed for various applications. In the field of satellite image processing, the segmentation is one of the vital step for gathering huge amount of information from the satellite images. The basic k-means clustering algorithm is simple and fast. The main problem associated with this clustering is not producing the same result for every run because the resulting clusters depends on the initial random assignments. In this paper, a modified k-means clustering algorithm is proposed for the effective segmentation of the satellite images. This proposed method always produces the same result for every run. The experimental results proved that the improved k-means algorithm is an efficient and effective method for the satellite image segmentation for the exact and accurate segmentation of satellite images.

Keywords: image segmentation; satellite image; kmeans clustering; centroid; fuzzy logic

I. INTRODUCTION

The segmentation is the process of grouping image pixels according to any one characteristics of the image. The goal of the segmentation process is to simplify and change the representation of an image into more meaningful and make easier to analyze [8]. The segmentation of satellite images requires automated or semi-automated analysis because the satellite images have huge volumes of data and detailed information. Generally the images received from satellite contains huge amount of data to decipher and process. But our human eye is insensitive to realize subtle changes in the image characteristics such as intensity, color, texture or brightness. So the manual human processing is not successful to retrieve the hidden treasures of information in the satellite image. The optimal solution is the processing of satellite images with digital computers. To retrieve the information or extract region of interest (ROI) from images, we need a segmentation method which is most important and difficult task in the image analysis. Even though an intensity image has only 256 variations, a color satellite image may contain more number of colors. For example a RGB image may contain 256*256*256 colors. In this case, setting up crisp boundaries for color is impossible. So a fuzzy logic based approach, fuzzy-k-means algorithm, is the best solution for the segmentation of the satellite images to gather more information. The paper is organized as follows. The basic principle of the *K*-means clustering and its application in the image segmentation were introduced in Section II. The detailed survey on previous work on the modification of k-means clustering algorithm for image segmentation is presented in Section III. The modified K-means clustering algorithm for satellite image segmentation using fuzzy logic was proposed in Section IV. The experimental simulation results are presented in Section V. Finally, some conclusions of the experimental results were given in Section VI.

II. KMEANS CLUSTERING ALGORITHM

K-means clustering algorithm, proposed by Mac Queen, is numerical, unsupervised, non-deterministic and iterative to segment or classify or cluster or to group the objects based on characteristics or features into K number of clusters [1] [2]. This is widely used because of its simpleness and fast convergence. In this algorithm, the clustering is done by minimizing the distances between data and the corresponding cluster centroid [7]. The basic K-Means clustering algorithm is simple. This algorithm starts with finding number of cluster (K) and assuming the center of these clusters (centroid). The working principle of K-means clustering algorithm can be explained as follows: If the number of data is less than the number of cluster then consider each data as the centroid of the cluster. If the number of data exceeds the number of cluster, we have to calculate the distance to all centroid and get the minimum distance for each data. Now the data belongs to the cluster that has minimum distance from this data. If we have no clue about the location of the centroid, we have to adjust the centroid location based on the current updated data. Then we have to assign all the updated data to this new centroid. This process is repeated until there is no data moving to another cluster anymore. The most important properties of this clustering algorithm is listed as (i) There is always 'K' number of clusters (ii) No overlapping of clusters (iii) No empty clusters i.e., atleast one data in each cluster (iv) The data in cluster is very close to its 'home' cluster than any other clusters (v) Apply k-means clustering only if the number of data is many. If the number of data is very few, when the same data is applied as input in different ways may produce different clusters.

This algorithm aims at minimizing an objective function, in this case a squared error function which is given by

© 2010, IJARCS All Rights Reserved

CONFERENCE PAPER
"National Conference on Networks and Soft Computing"
On 25-26 March 2013
Organized by
Vignan University, India

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (1)$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j , is an indicator of the distance of the *n* data points from their respective cluster centers.

Step 1: Determine the number of clusters, K, and assume the center of these clusters (centroid)

Step 2: Choose any random objects or the first K objects as the initial centroids

Step 3: Calculate the distance of each object to the centroids by using euclidean distance measure. This measure uses the same equation as the euclidean distance metric, except the square root. So the clustering with the euclidean squared distance metric is faster than clustering with the regular euclidean distance. Moreover the output of the K-Means clustering is not severely affected when the euclidean distance is replaced with euclidean squared distance.

Step 4: Assign each object to the group that has the closest centroid i.e., minimum distance

Step 5: After all objects were assigned, recalculate the positions of the K centroids

The K means algorithm repeat the steps 2-4 until convergence i.e., centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated. Generally, if a problem or model has many objects and each object have several attributes and want to classify the objects based on the attributes, then we can apply this algorithm. That's why K-mean clustering can be applied for many problems such as pattern recognition, classification analysis, artificial intelligence, image processing, machine vision, etc. The main disadvantage of K-Means clustering is not producing the same result for every run because the resulting clusters depends on the initial random assignments. So we ensure the same result on recurrent runs of the K-Means algorithm. This problem can be solved by the cluster centroids were determined using a fixed seed based randomization algorithm. As a result, every time the process starts the same centroids will be generated and the same outcome is obtained from the K-Means Clustering. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. This algorithm can be run number times to reduce this effect.

III. PREVIOUS WORK ON MODIFICATIONS ON K-MEANS CLUSTERING ALGORITHM FOR IMAGE SEGMENTATION

There are many ways to modify and improve the traditional K-Means clustering algorithm 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, in the initialization stage, the most of the modifications and improvement is performed on the K-Means clustering. The detailed survey of the previous work on modifications on the K-Means clustering algorithm is shown in table 1.

Table 1. Previous work on modifications on the K-Means clustering algorithm

Author - Year – Reference	*Modifications on the traditional K-Means Clustering Algorithm*
J.B.McQueen, 1967, 9	Suggested a learning strategy to determine a set of cluster seeds for segmentation.
J.Tou and R.Gonzales , 1974, 10	Proposed simple cluster seeking method.
Y. Linde, A. Buzo, R. M. Gray, 1980,11	Proposed binary splitting method for segmentation.
G. P. Babu, M. N. Murty, 1993, 15	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
C.Huang, R. Harris , 1993, 13	Based on principle component analysis, proposed the direct search binary splitting method
I. Katsavounidis, C. C. J. Kuo, Z. Zhen, 1994, 14	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.
M. B. A. Daoud, S. A. Roberts, 1996, 16	Introduced a method to divide the entire data into two different groups and the points are randomly distributed in the group.
P. S. Bradley, U. M. Fayyad, 1998,12	Proposed a method for choosing the most centrally located instance as the first seed.
A. Likas, N. Vlassis, J. J. Verbeek, 2003, 17	Proposed a global K means algorithm in which gradually increase the number of seeds till K is found that is till convergence
S. S. Khan, A. Ahmad, 2004 ,18	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.
Bo Zhao, Zhongxiang Zhu, Enrong Mao and Zhenghe Song , 2007, 19	Suggested a method for the image segmentation based on the ant colony optimization and the K means clustering.
Nor Ashidi Mat Isa, Samy A. Salamah, Umi Kalthum Ngah, 2009, 3	Proposed various modified version of the moving k-means (MKM) clustering algorithm. This proposed algorithm constantly checks the fitness of each centroid during the clustering stage. If the centroid cannot satisfy the criteria, the centroid moved to the group of data with the most closest or active center.

© 2010, IJARCS All Rights Reserved

IV. MODIFIED KMEANS CLUSTERING ALGORITHM FOR IMAGE SEGMENTATION

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 X_i *be* the i-th data and C_j 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 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)

$$m_{ik} = \frac{1}{\sum_{j=1}^{c}\left(d_{ik}/d_{jk} \right)^{\frac{2}{q-1}}} \qquad (2)$$

where d_{ik} is the distance from point k to the current centroid d_{jk} 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)

$$F(c_j) = \sum_{i \in c_j} (x_i - c_j)^2 \qquad (3)$$

The new location for all the centroid is calculated using (4)

$$c_j = {}^{1}/_{n_j} \sum_{i \in c_j} x_i \qquad (4)$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

The database which consists of 25 satellite images was created to test the proposed algorithm. Due to the time constraint all images are resized to 120 * 80 pixels. The proposed algorithms 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. Fig 1 shows the segmentation result of five test images by using the proposed method. For this execution, all images are processed with five clusters and the window size is five. The result is tabulated in table 1.

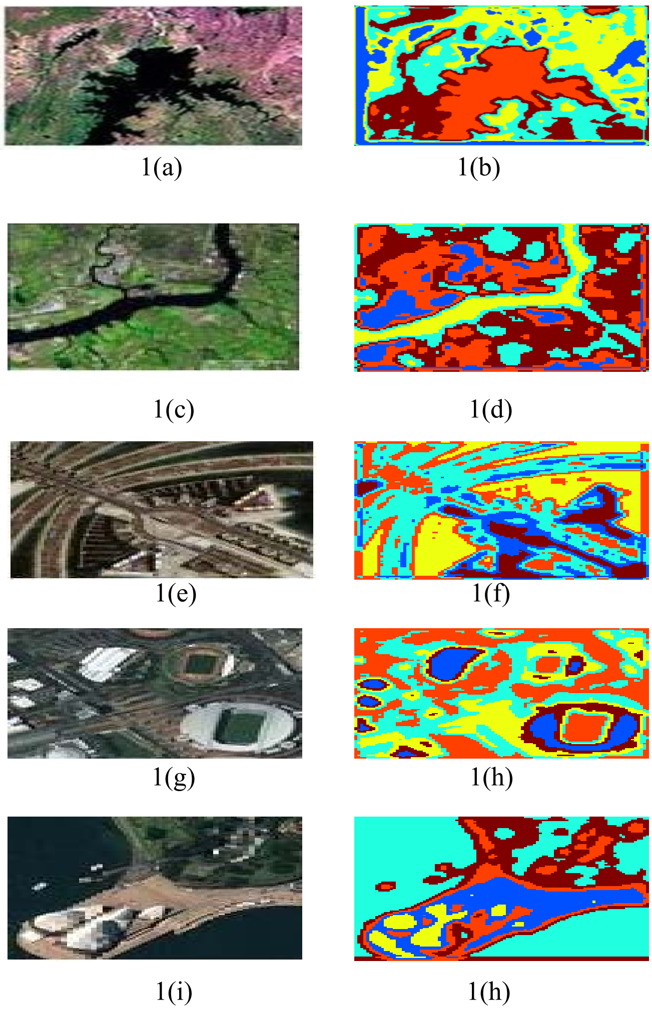
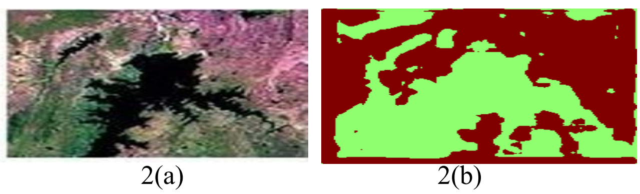


Figure 1. (First Column) Test Image (Second Column) Segmentation result of the proposed method with five clusters

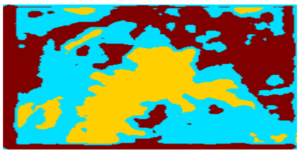
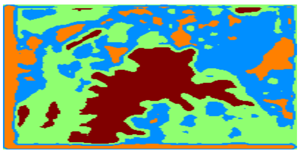
Table 2.Segmentation results for the test image with number of cluster is five and window size is six.

Sl. No	*Test Image*	*No. of cluster*	*Window size*	*Execution time in seconds*
1	2(a)	5	6	0.5928
2	2(c)	5	6	0.5748
3	2(e)	5	6	0.6204
4	2(g)	5	6	0.6328
5	2(i)	5	6	0.5390

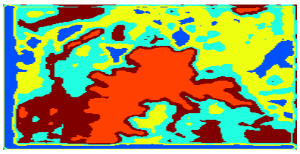
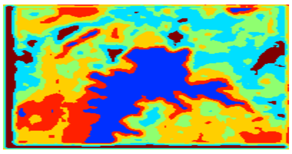
Fig 2 shows the segmentation result of a test image for various numbers of clusters ranging from two to six. Since this clustering is performed on the basis of color information, the number of cluster in the segmented image is defined by number of colors. The result is tabulated in the table 2.



© 2010, IJARCS All Rights Reserved

2(c) 2(d)

2(e) 2(f)

Figure 2. (a) Test Image (b)-(f) the segmentation result of the proposed method with two-to-six clusters respectively

Table 3.Segmentation result for the test image with number of cluster is 2, 3, 4, 5, and 6 and window size is six.

Sl. No	*No. of cluster*	*Window size*	*Execution time in seconds*
1	2	6	0.5148
2	3	6	0.5304
3	4	6	0.5817
4	5	6	0.5928
5	6	6	0.6240

VI. CONCLUSION

The modified K-Means clustering algorithm for satellite image segmentation was proposed in this paper. The characteristics 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 accurately, reduce the segmentation time, and improve the segmentation effect.

VII. REFERENCES

[1] 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 New Convergence Condition", Gervasi and M. Gavrilova (Eds.): ICCSA 2007, LNCS 4707, Part III, Springer-Verlag Berlin Heidelberg 2007, pp. 674–682.

[2] S. J. Redmond, C. Heneghan, "A method for initialising the K-means clustering algorithm using kd-trees. Science direct", Pattern Recognition Letters 28 (2007), 965–973.

[3] Nor Ashidi Mat Isa,Samy A. Salamah,, Umi Kalthum Ngah "Adaptive Fuzzy Moving K-means Clustering Algorithm for Image Segmentation " , IEEE Transactions on Consumer Electronics, Vol. 55, No. 4, Nov 2009

[4] D. Małyszko, S. T. Wierzchoń "Standard and Genetic K-means Clustering Techniques in Image Segmentation", (CISIM'07) 0-7695-2894-5/07 IEEE 2007

[5] Zhengjian Ding, Jin Sun, and Yang Zang, "FCM Image Segmentation algorithm based on color space and spatial information", International journal on computer and communication,Vol 2,No 1,2013

[6] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu "An efficient k-means clustering algorithm: analysis and implementation," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, No. 7, 2002.

[7] S. J. Redmond, C. Heneghan, "A method for initialising the K-means clustering algorithm using kd-trees. Science direct", Pattern Recognition Letters 28 (2007), 965–973

[8] Ganesan P. and Rajini,V, "A method to segment color images based on modified fuzzy-possibilistic-c-means clustering algorithm", In Proceedings of the IEEEI Conference on recent advances in space technology services and climate change., 157-163.

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

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

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

[12] P. S. Bradley, U. M. Fayyad, "Refining initial points for K-means clustering", *In: Proc.* 15th Internat. Conf. on Machine Learning. Morgan Kaufmann, San Francisco, CA, 1998, pp. 91–99.

[13] C. Huang, R. Harris "A comparison of several codebook generation approaches", IEEE Trans. Image Process. 2 (1), 1993, 108–112.

[14] 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.

[15] G. P. Babu, M. N. Murty, "A near-optimal initial seed value selection in K-means algorithm using a genetic algorithm", Pattern Recognition Lett. 14 (10), 1993, 763–769.

[16] M. B. A. Daoud, S. A. Roberts, "New methods for the initialization of clusters", Pattern Recognition Lett.17 (5), 1996, 451–45.

[17] A. Likas, N. Vlassis, J. J. Verbeek, 2003. "The global K-means clustering algorithm", Pattern Recognition 36, 451–461.

[18] S. S. Khan, A. Ahmad, 2004 "Cluster center initialization algorithm for k means clustering", Pattern Recognition Lett. 25 (11), 1293–1302.

[19] 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

© 2010, IJARCS All Rights Reserved