International Journal of Electrical and Computer Engineering (IJECE)
Vol. 4, No. 6, December 2014, pp. 923~930
ISSN: 2088-8708

Clustering Algorithm Combined with Hill Climbing for Classification of Remote Sensing Image

B. Sai Chandana*, K. Srinivas, R. Kiran Kumar***
*Department of CSE, GIT GITAM University, Visakhapatnam, India
**Department of CSE, VR Siddhartha Engineering College, Vijayawada, India
*Department of CSE, Krishna University, Machilipatnam, India

Article Info

Article history:

Received Aug 27, 2014
Revised Nov 7, 2014
Accepted Nov 22, 2014

Keyword:

Image Classification
Image Clustering
Image Processing
Remote Sensing

ABSTRACT

Clustering is an unsupervised classification method widely used for classification of remote sensing images. As the spatial resolution of remote sensing images getting higher and higher, the complex structure is the simple objects becomes obvious, which makes the classification algorithm based on pixels being losing their advantages. In this paper, four different clustering algorithms such as K-means, Moving K-means, Fuzzy K-means and Fuzzy Moving K-means are used for classification of remote sensing images. In all the traditional clustering algorithms, number of clusters and initial centroids are randomly selected and often specified by the user. In this paper, a hill climbing algorithm for the histogram of the input image will generate the number of clusters and initial centroids required for clustering. It overcomes the shortage of random initialization in traditional clustering and achieves high computational speed by reducing the number of iterations. The experimental results show that Fuzzy Moving K-means has classified the remote sensing image more accurately than other three algorithms.

Corresponding Author:

B. Saichandana
Member IEEE, CSI, IAENG and Assistant Professor
Department of Computer science and Engineering,
GITAM Institute of Technology, GITAM University,
Visakhapatnam-530045, India
Email: bschandana@gmail.com

1. INTRODUCTION

Remote sensing can be defined as any process whereby information is gathered about an object, area or phenomenon without making physical contact with the object [1]. The remote sensing technology (aerial sensor technology) is used to classify objects on the Earth (both on the surface, and in the atmosphere and oceans) by means of propagated signals. New opportunities to use remote sensing data have arisen, with the increase of spatial and spectral resolution of recently launched satellites. Remote sensing image classification is a key technology in remote sensing applications [2]. High resolution remote sensing has received much attention because of the detailed information it provides of the earth surface. The problem of image classification is to assign label to each image pixel [3]. Rapid and high accuracy remote sensing image classification algorithm is the precondition of kinds of practical applications. In remote sensing, sensors are available that can generate multispectral data, involving five to more than hundred bands.

Based on the criteria whether training samples are used or not, image classification methods are classified into two categories. These categories are distinguished in two main ways as supervised and unsupervised classification approaches [4]. In supervised classification approaches land cover classes are defined. Sufficient reference data are available and used as training samples. The signatures generated from the training samples are then used to train the classifier to classify the spectral data into thematic map [5].

Examples of supervised classification approaches are Maximum likelihood, minimum distance, artificial neural network, decision tree classifier etc. In Unsupervised classification approaches, clustering based algorithms are used to partition the spectral image into a number of spectral classes based on the statistical information inherent in the image. No prior definitions of the classes are used. The analyst is responsible for labeling and merging the spectral classes into meaningful classes. Examples such as ISODATA, K-means Clustering algorithm etc belong to unsupervised classification approaches. Unsupervised classification has evolved in two basic strategies [6], Iterative and Sequential. In an iterative procedure such as K-Means or ISODATA, an initial number of desired clusters are selected, and the centroid locations are then moved around until a statistically optimal fit is obtained. In a sequential algorithm such as Classification by Progressive Generalization, the large number of spectral combinations is gradually reduced through a series of steps using various proximity measures.

In this paper, we used four different clustering algorithms for classification of remote sensing image. In the clustering algorithms, parameters such as cluster number and initial centroid positions are chosen randomly and often specified by the user. In the proposed clustering algorithms, instead of randomly initializing the parameters in the clustering algorithms, the hill climbing algorithm on the histogram of input image will automatically determine the cluster centers and the number of clusters in the image. The Hill climbing algorithm adaptively determines the initial cluster centers and the number of clusters according to the characteristics of the image. Using hill climbing algorithm as a preliminary stage with clustering algorithms reduces the number of iterations for classification and costs less execution time. The qualitative and quantitative results show that Fuzzy Moving K-means clustering algorithm has classified the image better than other clustering algorithms.

The paper is organized as follows: Section 1 presents the Hill Climbing Algorithm, Section 2 presents the K-means clustering algorithm, Section 3 presents Moving K-means clustering algorithm, Section 4 presents Fuzzy C-means clustering algorithm, Section 5 presents Fuzzy Moving K-means Clustering algorithm, Section 6 presents Experimental results and finally Section 7 report conclusions.

2. HILL CLIMBING ALGORITHM

Traditional hill-climbing segmentation [7] [8] is a nonparametric algorithm that clusters the colors of an image. The idea is that each cluster is represented by a hill in the histogram, where the hill consists of adjacent colors. In this paper, an extended version of hill climbing algorithm is presented. This Hill climbing algorithm which is used in the preliminary stage for a clustering algorithm is stated as follows:

Input: Global three dimensional color histogram of the image by the three channels of selected suitable color space.

Output: The number and values of peaks= Number of clusters and initial centroids respectively

Step 1: Start at a non-zero bin of the histogram and make uphill moves until reaching a peak as follows:

1. Compare the number of pixels of the current histogram bin with the number of pixels of the neighboring bins.
2. If the neighboring bins have different number of pixels, the algorithm make an uphill move towards the neighboring bin with large number of pixels.
3. If the immediate neighboring bins have the same number of pixels, the algorithm checks the next neighboring bins, and so on, until two neighboring bins with different number of pixels are found. Then, an uphill move is made towards the bin with larger number of pixels.
4. The uphill is continued until reaching a bin from where there is no possible hill movement. That is the case when the neighboring bins have smaller number of pixels than the current bin. Hence the current bin is identified as peak representing local maximum.
5. If no uphill move is done, the stopping bin is identified as a peak of a hill, and all bins leading to this peak are associated with it.

Step 2: Select another unclimbed bin as a starting bin and perform step 1 to another peak. This step is continued until all the nonzero bins of the histogram are climbed.

Step 3: Thresholding: Find the peaks whose value is higher than one percent of the maximum peak (which is identified as local maximum in step 1 and 2).

Step 4: Remove the peaks which are very close. This is done by checking the difference between the grey levels of the two individual peaks. If the difference is less than 20, then the peak with lowest value is removed.

Step 5: Neighboring pixels that lead to the same peak are grouped together.

Step 6: The identified peaks represent the initial number of clusters of the input image. Thus the number and values of the peaks are saved.

3. K-MEANS CLUSTERING ALGORITHM

K-means is one of the basic clustering methods introduced by Hartigan [9]. This method is used to group data to the nearest centre. The method is numerical, unsupervised, nondeterministic and iterative. The K-means clustering algorithm for classification of remote sensing image is summarized as follows:
Input: N: number of pixels to be clustered; $x=\{x_1, x_2, x_3, \ldots\ldots, x_N\}$: pixels of remote sensing image; $c = \{c_1, c_2, c_3, \ldots., c_j\}$clusters respectively.
Output: cl: cluster of pixels
Step 1: Number of clusters and cluster centroids are determined by hill climbing algorithm.
Step 2: compute the closest cluster for each pixel and classify it to that cluster, ie: the objective is to minimize the sum of squares of the distances given by the following:

$$\Delta_{ij} = || x_i - c_j ||. \quad arg\ min \sum_{i=1}^{N} \sum_{j=1}^{C} \Delta_{ij}^2 \tag{1}$$

Step 3: Compute new centroids after all the pixels are clustered. The new centroids of a cluster is calculated by the following

$$c_j = \frac{1}{N\,j} \sum x_i \text{ where } x_i \text{ belongs to } c_j \tag{2}$$

Step 4: Repeat steps 2-3 till the sum of squares given in equation is minimized.

4. MOVING K-MEANS CLUSTERING ALGORITHM

The Moving K-means clustering algorithm is the modified version of K-means proposed in [10]. It introduces the concept of fitness to ensure that each cluster should have a significant number of members and final fitness values before the new position of cluster is calculated. The Moving K-means clustering algorithm for classification of remote sensing image is summarized as follows:
Input: N: number of pixels to be clustered; $x=\{x_1, x_2, x_3, \ldots\ldots, x_N\}$: pixels of remote sensing image.
$c = \{c_1, c_2, c_3, \ldots., c_j\}$: clusters respectively.
Output: cl: cluster of pixels
Step 1: Number of clusters and cluster centroids are determined by hill climbing algorithm.
Step 2: compute the closest cluster for each pixel and classify it to that cluster, ie: the objective is to minimize the sum of squares of the distances given by the following:

$$\Delta_{ij} = || x_i - c_j ||. \quad arg\ min \sum_{i=1}^{N} \sum_{j=1}^{C} \Delta_{ij}^2 \tag{3}$$

Step 3: The fitness for each cluster is calculated using

$$f(c_k) = \sum_{t \in c_k} (|| x_t - c_k ||)^2 \tag{4}$$

The relationship among the centers must satisfy the following condition:

$$f(c_s) \geq \alpha_a f(c_l) \tag{5}$$

where α_a is small constant value initially with value in range $0< \alpha_a <1/3$, c_s and c_l are the centers that have the smallest and the largest fitness values. If (5) is not fulfilled, the members of c_l are assigned as members of c_s, while the rest are maintained as the members of c_l. The positions of c_s and c_l are recalculated according to:

$$C_s = 1/n_{cs} \left(\sum_{t \in c_s} x_t \right) \tag{6}$$

$$C_l = 1/n_{cl} \left(\sum_{t \in c_l} x_t \right) \tag{7}$$

The value of α_a is then updated according to:

$$\alpha_a = \alpha_a - \alpha_a / n_c \quad (8)$$

The above process are repeated until (5) is fulfilled. Next all data are reassigned to their nearst center and the new center positions are recalculated using (2).
Step 4: The iteration process is repeated until the following condition is satisfied.

$$f(c_s) \geq \alpha_a f(c_l) \quad (9)$$

5. FUZZY C-MEANS CLUSTERING ALGORITHM

The fuzzy c-means [11] is an unsupervised clustering algorithm. The main idea of introducing fuzzy concept in the fuzzy c-means algorithm is that an object can belong simultaneously to more than one class and does so by varying degrees called memberships. It distributes the membership values in a normalized fashion. It does not require prior knowledge about the data to be classified. It can be used with any number of features and number of classes. The fuzzy c-means is an iterative method which tries to separate the set of data into a number of compact clusters. It improves the partition performance and reveals the classification of objects more reasonable. The predefined parameters such as number of clusters and initial clustering centers are provided by Hill climbing algorithm on the histogram of remote sensing image. The fuzzy c-means algorithm is summarized as follows:
Input: N=number of pixels to be clustered; $x = \{x_1, x_2, ..., x_N\}$: pixels of remote sensing image;
$c = \{c_1, c_2, c_3, \ldots, c_j\}$: clusters respectively;m=2: the fuzziness parameter;
Output: membership values of pixels and clustered Image
Step_1: Initialize the membership matrix u_{ij} is a value in (0,1) and the fuzziness parameter m (m=2). The sum of all membership values of a pixel belonging to clusters should satisfy the constraint expressed in the following.

$$\sum_{j=1}^{c} u_{ij} = 1 \quad (10)$$

for all i= 1, 2, …….N, where c is the number of clusters and N is the number of pixels in remote sensing image.
Step_2: Compute the centroid values for each cluster c_j. Each pixel should have a degree of membership to those designated clusters. So the goal is to find the membership values of pixels belonging to each cluster. The algorithm is an iterative optimization that minimizes the cost function defined as follows:

$$F = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^{m} \, || x_j - c_i ||^2 \quad (11)$$

where u_{ij} represents the membership of pixel x_j in the ith cluster and m is the fuzziness parameter.
Step_3: Compute the updated membership values u_{ij} belonging to clusters for each pixel and cluster centroids according to the given formula.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_j - v_i\|}{\|x_j - v_k\|} \right)^{2/(m-1)}},$$

and

$$v_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} x_j}{\sum_{j=1}^{N} u_{ij}^{m}}. \quad (12)$$

Step_4: Repeat steps 2-3 until the cost function is minimized.

6. FUZZY MOVING K-MEANS CLUSTERING ALGORITHM

In the Fuzzy Moving K-means clustering algorithm [11], the membership function is used in addition to the Euclidian distance to control the assignment of the members to the proper center. The predefined parameters such as number of clusters and initial clustering centers required for clustering algorithm are provided by executing Hill climbing mechanism on the histogram of the remote sensing image. The Fuzzy Moving K-means clustering algorithm is summarized as follows:
Input:N: number of pixels to be clustered;$x=\{x_1,x_2,x_3,\ldots\ldots,x_N\}$: pixels of remote sensing image; $c=\{c_1,c_2,c_3,\ldots.,c_j\}$: clusters respectively; m=2: the fuzziness parameter;
Output: membership values of pixels and clustered Image
Step_1: Initialize the membership matrix u_{ij} is a value in (0,1) and the fuzziness parameter m (m=2). The sum of all membership values of a pixel belonging to clusters should satisfy the constraint expressed in the following.

$$\sum_{j=1}^{c} u_{ij}=1 \tag{13}$$

for all i= 1,2,…….N, where c (=2) is the number of clusters and N is the number of pixels in remote sensing image.
Step_2: Compute the centroid values for each cluster c_j. Each pixel should have a degree of membership to those designated clusters. So the goal is to find the membership values of pixels belonging to each cluster. The algorithm is an iterative optimization that minimizes the cost function defined as follows:

$$F=\sum_{j=1}^{N}\sum_{i=1}^{c} u_{ij}^{m} \,||\, x_j-c_i||^2 \tag{14}$$

where u_{ij} represents the membership of pixel x_j in the i^{th} cluster and m is the fuzziness parameter.
Step 3: The fitness for each cluster is calculated using

$$f\,(c_k) = \sum_{t\in c_k} (\,||\, x_t-c_k\,||\,)^2 \tag{15}$$

All centers must satisfy the following condition:

$$f(c_s) \geq \alpha_a f(c_l) \text{ and } m(c_{sk})>m(c_{lk}) \tag{16}$$

where α_a is small constant value initially with value in range $0< \alpha_a <1/3$, c_s and c_l are the centers that have the smallest and the largest fitness values, $m(c_{sk})$ is the membership value of point k according to the smallest centre and $m(c_{lk})$ is the membership value of point k according to the largest centre. If (16) is not fulfilled, the members of c_l are assigned as members of c_s, while the rest are maintained as the members of c_l. The positions of c_s and c_l are recalculated according to:

$$C_s = 1/n_{cs}\,(\sum_{t\in c_s} x_t) \tag{17}$$

$$C_l=1/n_{cl}\,(\sum_{t\in c_l} x_t) \tag{18}$$

The value of α_a is then updated according to:

$$\alpha_a=\alpha_a- \alpha_a/n_c \tag{19}$$

The above process are repeated until (16) is fulfilled. Next all data are reassigned to their nearest center and the new center positions are recalculated using (2).
Compute the updated membership values u_{ij} belonging to clusters for each pixel according to given formula

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_j - v_i\|}{\|x_j - v_k\|} \right)^{2/(m-1)}},$$

and

$$v_i = \frac{\sum_{j=1}^{N} u_{ij}^m x_j}{\sum_{j=1}^{N} u_{ij}^m}. \quad (20)$$

Step 4: The iteration process is repeated until the following condition is satisfied.

$$f(c_s) \geq \alpha_a f(c_l) \quad and \quad m(c_{sk}) > m(c_{lk}) \quad (21)$$

7. EXPERIMENTAL RESULTS

Qualitative Analysis:

The proposed four clustering algorithms are performed on a two remote sensing images, which are taken from bhuvan.nrsc.gov.in [12] Clustering algorithms with and without hill climbing are executed on the two remote sensing images. The classification result of fuzzy moving k-means clustering algorithm with hill climbing on two images is shown in figure 1. The hill climbing algorithm is executed on the histogram of two images. For the first image the number of peaks =4 and for the second image, the number of peaks =3. Then the clustering algorithm is executed with the identified number of clusters and initial centroids.

Quantitative Analysis:

Quantitative analysis is a numerically oriented procedure to figure out the performance of algorithms without any human error. In the classification map the classes are labeled with different colors. Table 1 shows the classification accuracy of different clustering methods on two remote sensing images. The results confirm that Fuzzy Moving K-means algorithm produces highest classification accuracy in classifying the remote sensing image.

As the initial centroids required for clustering algorithms are determined by Hill climbing algorithm, the number of iterative steps required for classifying the objects is reduced. While the initial centroids obtained by hill climbing are unique, the classified result is more stable compared with traditional algorithms. Table 2 shows the comparison of iterative steps numbers for clustering algorithms with and without Hill climbing.

Table 2. Comparison of iterative step numbers

	Clustering algorithm	Iterative steps (without hill climbing)	Iterative steps (with hill climbing)
IMAGE 1	K-means	33	16
	Moving K-means	40	19
	Fuzzy k-means	55	28
	Fuzzy Moving K-means	62	38

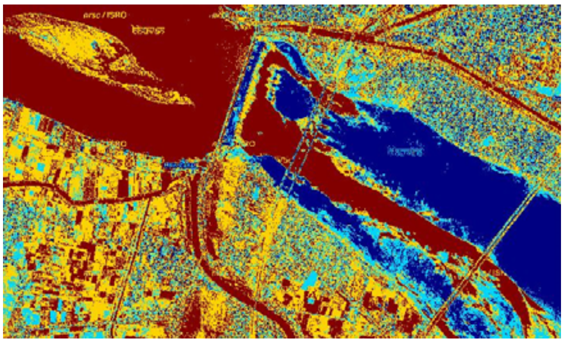
	Clustering algorithm	Iterative steps (without hill climbing)	Iterative steps (with hill climbing)
IMAGE 2	K-means	43	23
	Moving K-means	58	32
	Fuzzy k-means	72	41
	Fuzzy Moving K-means	89	49

Table 1. Classification Accuracy in Percentages

Method	(IMAGE 1)	(IMAGE 2)
K-means	81	84
Moving K-means	86	89
Fuzzy K-means	89	92
Fuzzy Moving K-means	92	94

Image 1

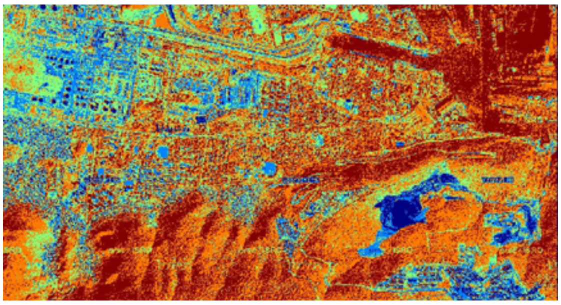
Fuzzy Moving K-means (Number of clusters =4)



Brown: Water, Blue: Sand, Orange: Green area, Sky blue: Houses

Image 2

Fuzzy Moving K-means (Number of clusters =5)



Orange: Trees, Brown: Water, Blue: Hill top, Skyblue: Houses, Green: Roadway

Figure 1. classification results

8. CONCLUSION

This paper has presented four clustering algorithms namely K-means, Moving K-means, Fuzzy K-means and Fuzzy Moving K-means combined with hill climbing for the classification of remote sensing image. The qualitative and quantitative analysis done proved that Fuzzy Moving K-means has higher classification quality than other clustering algorithms. Clustering algorithm combined with hill climbing overcomes the problem of random selection of number of clusters and initialization of centroids. The proposed method reduces the number of iterations for classifying a remote sensing image and costs less execution time.

REFERENCES

[1] Yi Ma, Jie Zhang, Yufeng Gao, "High Resolution Remote Sensing Image Classification of Coastal Zone and its Automatic Realization", *International Conference on Computer Science and Software Engineering,* pp. 827-829 2008 IEEE.

[2] Xiaofang Liu, Xiaowen Li, Ying Zhang, Cunjian Yang, Wenbo Xu, Min Li, Huanmin Luo, "Remote Sensing Image Classification Using Function Weighted FCM Clustering Algorithm", *International Geoscience and Remote Sensing Symposium*, pp. 2010-2013, 2007 IEEE.

[3] Penglin Zhang, Zhiyong Lv, Lipeng Gao and Li Huang, "A New Framework of the unsupervised Classification for High-Resolution Remote Sensing Image", *Telkomnika Indonesian Journal of Electrical Engineering*, Vol 10, No 7, November 2012, pp. 1746-1755.

[4] D. Lu, Q. Weng, "A Survey of image classification methods and techniques for improving classification performance", *International Journal of Remote sensing*, Vol 28, No. 5, 10 March 2007, pp. 823-870.

[5] Josef Cihlar, Rasim Latifovic, Jean Beaubien, "A Comparison of Clustering Strategies for Unsupervised Classification", *Canadian Journal of Remote Sensing*, Vol. 26, issue 5, pp. 446-454.

[6] Aykut AKGÜN, A. Hüsnü ERONAT and Necdet TÜRKa, "Comparing different satellite image classification methods: an application in ayvalik district, western, Turkey", ISPRS Archives- Volume XXXV Part B4, 2004, pp. 1091-1097.

[7] Zhengjian DING, Juanjuan JIA, DIA LI, "Fast Clustering Segmentation Method Combining Hill Climbing for Color Image", Journal of Information and Computational Sciences, Vol 8, pp. 2949-2957.
[8] Takumi OHASHI, Zaher AGHBARI, Akifumi MAKINOUCHI, "Hill Climbing algorithm for Efficient Color bases Image Segmentation", *Signal Processing, Pattern Recognition and Applications*, 01/2003.
[9] J. A. Hartigan and M. A. Wong, "A K-means clustering algorithm", *Applied Statistics*, 28, 1979, pp 100-108.
[10] Siti Naraini Sulaiman, Nor Ashidi Mat Isa, "Denoising based Clutering Algorithms for Segmentation of Low level of Salt and Pepper Noise Corrupted Images", IEEE Transactions on Consumer Electronics, Vol. 56, No.4, November 2010, pp 2702-2710.
[11] 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, Issue 4, pp 2145-2153.
[12] www.bhuvan.nrsc.gov.in

BIOGRAPHIES OF AUTHORS

B. Saichandana received B.Tech and M.Tech degree from JNTU Hyderabad and Andhra University in the year 2005 and 2007 respectively. She is currently working as Assistant professor in the Department of CSE, GIT, Gitam University. Her research interest include image classification, pattern recognition etc. Currently she is persuing phd from JNTU Kakinada.

Dr. K. Srinivas received MCA, M.Tech and Phd degrees from Andhra University, JNTU kakinada and Acharya Nagarjuna University. His research interest include image processing and bioinformatics. Currently he is working as Professor, Department of Computer science and Engineering, VR Siddhartha Engineering College, Vijayawada.

Dr. R. Kiran kumar received MCA, M.Tech and Phd degrees from Andhra University, JNTU kakinada and Acharya Nagarjuna University. His research interest include image processing and bioinformatics. Currently he is working as assistant Professor, department of Computer science, krishna University, Machilipatnam.