ISSN 2229-5518

Clustering Algorithm Combined with Empirical Mode Decomposition for Classification of Remote Sensing Image

B.Sai Chandana, Dr.K.Srinivas, Dr.R.Kiran Kumar

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, the estimation of number of clusters and initial centroids using Empirical Mode Decomposition algorithm (ECEMD) 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.

Index terms: Image Clustering, Remote Sensing, Image Processing, Image Classification

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]. 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.

At present, there is different image classification methods used for different purposes by various researches. These techniques are distinguished in two main ways as supervised and unsupervised classifications [3]. Supervised classification has different sub-classification methods which are named as parallelepiped, maximum likelihood, minimum distances and Fisher classifier methods. Unsupervised classification has evolved in two basic strategies [4], 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 [5] [10].

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. Instead of randomly initializing the parameters in the clustering algorithms, the ECEMD algorithm on the histogram of input image will automatically determine the cluster centers and the number of clusters in the image. The ECEMD algorithm is extension of EEMD algorithm presented in [16]. Using ECEMD 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 2 presents the Empirical Mode Decomposition (EMD) Algorithm and estimation of number of number of clusters and initial centroids using EMD (ECEMD), Section 3 presents the K-means clustering algorithm, Section 4 presents Moving K-means clustering algorithm, Section 5 presents Fuzzy C-means clustering algorithm, Section 6 presents Fuzzy Moving K-means Clustering algorithm, Section 7 presents Experimental results and finally Section 8 report conclusions.

2 EMPIRICAL MODE DECOMPOSITION

ISSN 2229-5518

The Empirical Mode Decomposition (EMD) proposed by Dr. Norden Huang [12] , was a technique for analyzing nonlinear and non-stationary signals. It serves as an alternative to methods such as wavelet analysis and short-time Fourier transform. It decomposes any complicated signal into a finite and often small number of Intrinsic Mode Functions (IMF). The IMF is symmetric with respect to local zero mean and satisfies the following two conditions.

1. The number of extrema and the number of zero crossings must either be equal or differ by one.
2. At any point, the mean value of the envelope defined by local maxima and local minima is zero, indicating the function is locally symmetric.

The decomposition method in EMD is called Shifting Process [13]. The shifting process of the 1- dimensional signal can be adapted as follows.

1. Let Ioriginal be the original signal to be decomposed. Let j=1 (index number of IMF), Initially, I= Ioriginal .
2. Identify the local maxima and local minima points in I.
3. By using interpolation, create the upper envelope Eup of local maxima and the lower envelope Elw of local minima.
4. Compute the mean of the upper envelope and lower envelope.
   Emean= [Eup + Elw]/2
5. Iimf = I- Emean.
6. Repeat steps 2-5 until Iimf can be considered as an IMF.
7. IMF(j)= Iimf, j=j+1, I = I- Iimf ,
8. Repeat steps 2-7 until, the standard deviation of two consecutive IMFs is less than a predefined threshold or the number of extrema in I is less than two.

The first few IMFs obtained from EMD contain the high frequency components which correspond to salient features in original image and the residue represents low frequency component in the image. The original image can be recovered by inverse EMD as follows:

$$I = RES + \sum_{j} IMF(j) \qquad (1)$$

ESTIMATION OF NUMBER OF CLUSTERS & INITIAL CENTROIDS USING EMD (ECEMD)

1. Let h(k) be the histogram for the input image I with k=0 ,…., G and G being the maximum intensity value in the image

2. Calculate the probability mass function p(k) for the input image histogram.

   $$p(k) = h(k)/M \qquad (2)$$

   where M is the total number of image pixels.

3. Divide the normalized histogram p(k) into IMFs using empirical mode decomposition. The first IMF carries the histogram noise, irregularities and sharp details of the histogram, while the last IMF and residue describe the trend of the histogram. On the other hand, the intermediate IMFs describe the initial histogram with simple and uniform pulses.

4. Consider the summation of intermediate IMFs as follows:

   $$I_{INT} = \sum_{j=2}^{n-2} IMF_j \qquad (3)$$

   where n is the number of IMFs.

5. Determine all local minima in I_{INT} .

   $$I^* = \left\{ \min_{0 \le T \le G} I_{INT}(T) \right\} \qquad (4)$$

   I* is the vector carrying all local minima. All those local minima could express image clusters, but most of them are very close to each other and some of them lie too high to be a cluster. So, truncate the local minima to the important ones that could express an image cluster.

6. The truncation process is carried out in two steps. In the first step, the algorithm truncates all local minima that have a value larger than the threshold, where threshold is equal to average of the values of local minima. The truncation step is expressed as follows:

   $$thr = \frac{1}{2N_{I^*}} \sum_{I_i^* \in I^*} I_i^* , \qquad (5)$$

   where N_{I^*} is the number of local minima belonging to I^* and I_i^* is the local minima belonging to vector I^*.

$$I^t = \{I_i^*\}, \text{ if } I_i^* < \text{thr and } I_i^* \in I^*. \quad (6)$$

Where I^t consists of all local minima which are less than the estimated threshold thr.

7. In the second truncation step, the algorithm applies an iterative procedure that calculates the number of image pixels belonging to each candidate image cluster and prunes the cluster with smallest number of image pixels (less than 2 percent of total number of image pixels). The pruned candidate clusters are merged with their closest image clusters.
8. The number of elements in final vector I^t represents the number of clusters denoted by NC.
9. Determine the local maxima in I_{INT}.

$$I_M^* = \left\{ \min_{0 \le T \le G} I_{INT}(T) \right\} \quad (7)$$

IM* is the vector carrying all local maxima.

10. Sort the vector IM* in descending order. The corresponding values (X-coordinate values) for the first NC values in vector IM*, represent the initial centriods.

3 K-MEANS CLUSTERING ALGORITHM

K-means is one of the basic clustering methods introduced by Hartigan [6]. This method is applied to clasiffy the remote sensing image in recent years. The K-means clustering algorithm for classification of remote sensing image is summarized as follows [14]:

Algorithm K-means(x, N, c)

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

Begin

Step 1: cluster centroids and number of clusters are determined by ECEMD 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:

$$\Delta_{ij} = || x_i - c_j ||. \quad \arg\min \sum_{i=1}^{N} \sum_{j=1}^{C} \Delta_{ij}^2 \quad (8)$$

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

$$c_j = \frac{1}{N_j} \sum x_i \text{ where } x_i \text{ belongs to } c_j . \quad (9)$$

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

End.

4 MOVING K-MEANS CLUSTERING ALGORITHM

The Moving K-means clustering algorithm is the modified version of K-means proposed in [7]. 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:

Algorithm Moving K-means(x, N, c)

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

Begin

Step 1: cluster centroids and number of clusters are determined by ECEMD 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:

$$\Delta_{ij} = || x_i - c_j ||. \quad \arg\min \sum_{i=1}^{N} \sum_{j=1}^{C} \Delta_{ij}^2 \quad (10)$$

Step 3: The fitness for each cluster is calculated using

$$f(c_k) = \sum_{t \in c_k} (|| x_t - c_k ||)^2 \quad (11)$$

All centers must satisfy the following condition:

$$f(c_s) \ge \alpha_a f(c_l) \quad (12)$$

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

ISSN 2229-5518

rest are maintained as the members of c_l. The positions of c_s and c_l are recalculated according to:

$$C_s = 1/n_{cs} \left(\sum_{t \in c_s} x_t \right) \quad (13)$$

$$C_l = 1/n_{cl} \left(\sum_{t \in c_l} x_t \right) \quad (14)$$

The value of α_a is then updated according to:

$$\alpha_a = \alpha_a - \alpha_a / n_c \quad (15)$$

The above process are repeated until (12) is fulfilled. Next all data are reassigned to their nearest center and the new center positions are recalculated using (9).

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

$$f(c_s) \geq \alpha_a f(c_l) \quad (16)$$

End

5 FUZZY C-MEANS CLUSTERING ALGORITHM

The Fuzzy C-means [8] [9] 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 ECEMD algorithm. The Fuzzy C-means algorithm is summarized as follows:

Algorithm Fuzzy C-Means (x,N,c,m)
Input:
N=number of pixels to be clustered;
$x = \{x_1, x_2, ..., x_N\}$: pixels of remote sensing image;
$c = \{c_1, c_2, c_3, ..., c_j\}$: clusters respectively.
m=2: the fuzziness parameter;
Output:
u: membership values of pixels and clustered Image
Begin
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.

$$\sum_{j=1}^{c} u_{ij} = 1 \quad (17)$$

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:

$$F = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^{m} \, || x_j - c_i ||^2 \quad (18)$$

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

$$v_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} x_j}{\sum_{j=1}^{N} u_{ij}^{m}}. \quad (19)$$

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

End.

6 FUZZY MOVING K-MEANS CLUSTERING ALGORITHM

In the Fuzzy Moving K-means clustering algorithm [9], 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 ECEMD algorithm 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;
$x = \{x_1, x_2, x_3, ..., x_N\}$: pixels of remote sensing image

ISSN 2229-5518

$c = \{c_1, c_2, c_3, \ldots, c_j\}$: clusters respectively

m=2: the fuzziness parameter;

Output:

u: membership values of pixels and clustered Image

Begin

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.

$$\sum_{j=1}^{c} u_{ij} = 1 \qquad (20)$$

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:

$$F = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^{m} \, || x_j - c_i ||^2 \qquad (21)$$

where u_{ij} represents the membership of pixel xj in the ith cluster and m is the fuzziness parameter.

Step 3: The fitness for each cluster is calculated using

$$f(c_k) = \sum_{t \in c_k} (|| x_t - c_k ||)^2 \qquad (22)$$

All centers must satisfy the following condition:

$$f(c_s) \geq \alpha_a \, f(c_l) \text{ and } m(c_{sk}) > m(c_{lk}) \qquad (23)$$

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 (5) is not fulfilled, the members of cl 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:

$$C_s = 1/n_{cs} \left(\sum_{t \in c_s} x_t \right) \qquad (24)$$

$$C_l = 1/n_{cl} \left(\sum_{t \in c_l} x_t \right) \qquad (25)$$

The value of α_a is then updated according to:

$$\alpha_a = \alpha_a - \alpha_a / n_c \qquad (26)$$

The above process are repeated until (23) is fulfilled. Next all data are reassigned to their nearst center and the new center positions are recalculated using (9).

Compute the updated membership values uij 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

$$v_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} x_j}{\sum_{j=1}^{N} u_{ij}^{m}}. \qquad (27)$$

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

$$f(c_s) \geq \alpha_a \, f(c_l) \quad \text{and} \quad m(c_{sk}) > m(c_{lk}) \qquad (28)$$

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 [15]. Clustering algorithms with and without ECEMD are executed on the two remote sensing images. The classification result of fuzzy moving k-means clustering algorithm with ECEMD on two images is shown in figure 2. The ECEMD algorithm is executed on the histogram of two images. The IMFs, Local Minima and Local Maxima for the histogram of input image 1 is shown in figure 1. By using ECEMD, for the first image the number of clusters =4 and for the second image, the number of clusters =5. 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. The Mean Square Error (MSE) is significant metric to validate the quality of image. It measures the square error between pixels of the original and the resultant images. The MSE is mathematically defined as

$$MSE = \frac{1}{N} \sum_{j=1}^{k} \sum_{i \in c_j} ||v_i - c_j||^2 \qquad (29)$$

Where N is the total number of pixels in an image and xi is the pixel which belongs to the jth cluster. The lower

difference between the resultant and the original image reflects that all the data in the region are located near to its centre. Table 1 shows the quantitative evaluations of four clustering algorithms. The results confirm that Fuzzy Moving K-means algorithm produces the lowest MSE value for classifying the remote sensing image. As the initial centroids required for clustering algorithms are determined by ECEMD, the number of iterative steps required for classifying the objects is reduced. While the initial centroids obtained by ECEMD 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 ECEMD.

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 ECEMD 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 ECEMD overcomes the problem of random selection of number of clusters and initialization of centroids. The proposed method reduces the number of iterations for classifying an 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", 2008 International Conference on Computer Science and Software Engineering, 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" , 2007 IEEE.

[3] Aykut AKGÜN, A.Hüsnü ERONAT and Necdet TÜRKa, "Comparing different satellite image classification methods: an application in ayvalik district,western, Turkey.

[4] Josef Cihlar, Rasim Latifovic, Jean Beaubien, "A Comparison of Clustering Strategies for Unsupervised Classification, Canadian Journal of Remote Sensing, February 2000.

[5] D. LU and 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, 823–870.

[6] Volkan Uslan, Omur Bucak, "Clustering based spot segmentation of microarray cDNA Microarray Images ", International Conference of the IEE EMBS , 2010.

[7] 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.

[8] LJun-Hao Zhang, Ming Hu HA , Jing Wu," Implementation of Rough Fuzzy K-means Clustering Algorithm in Matlab", Proceedings of Ninth International Conference on Machine Learning and Cybernetics", July 2010.

[9] Nor Ashidi Mat Isa, Samy A.Salamah, Umi Kalthum Ngah.," Adaptive Fuzzy Moving K-means Clustering Algorithm for Image Segmentation", IEEE 2009.

[10] B.Saichandana," De-noising based clustering Algorithm for Classification of Remote Sensing Image", Journal of Computing, Vol 4, Issue 11, November 2012.

[11] 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.

[12] N.E.Huang, Z.Shen, S.R.Long, "The empirical mode decomposition and the Hilbert Spectrum for non-linear and non-stationary time series analysis". Proc. Roy. Soc, London.A, Vol.454, pp.903-995, 1998.

[13] Harikiran, J.; RamaKrishna, D.; Phanendra, M. L.; Lakshmi, P. V. and Kumar, R. Kiran. "Fuzzy C- means with Bi-dimensional Empirical Mode Decomposition for Segmentation of Microarray Image", International Journal of
Computer Science Issues (IJCSI), 2012.

[14] Harikiran.j.,M.Phanendra,N.N.Swamy, P.V.Lakshmi and R.Kiran Kumar. "K-means with Bi-dimensional Empirical Mode Decomposition for Segmentation of Microarray Image", Proceedings of the Second International Conference on Computational Science engineering and Information Technology, 2012.

[15] www.bhuvan.nrsc.gov.in.

[16] S.Krindis, M.Krindis, V.Chatzis," Workspace for image clustering based on empirical mode decomposition", IET image processing, doi:10.1049/iet-ipr.2010.0592.

TABLE 2: COMPARISON OF ITERATIVE STEP NUMBERS

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

TABLE 1: MSE VALUES

Method	MSE Values (IMAGE 1)	MSE Values (IMAGE 2)
K-means	282.781	346.47
Moving K-means	216.392	298.69
Fuzzy K-means	138.327	198.76
Fuzzy Moving K-means	96.322	142.83

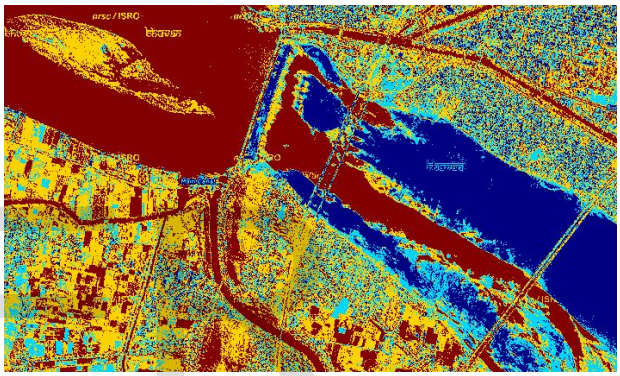
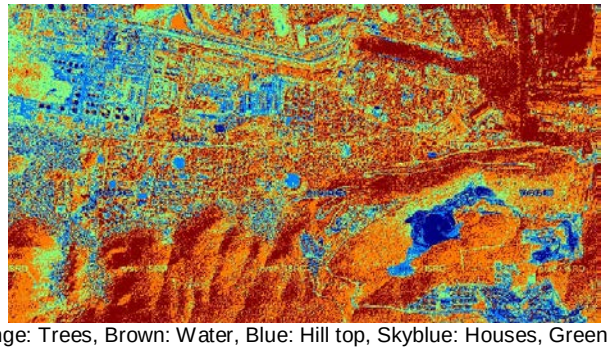
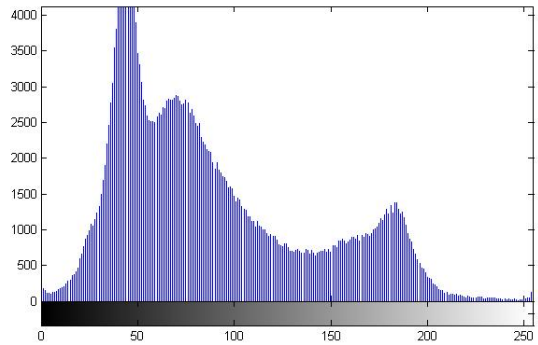
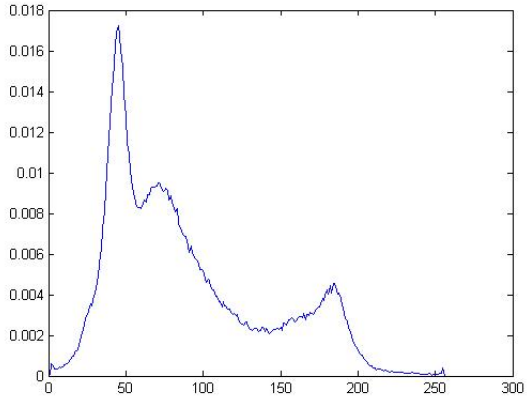
Image 1	Fuzzy Moving K-means (Number of clusters =4)
	 Brown: Water, Blue: Sand, Orange: Green area, Sky blue: Houses
Image 2	Fuzzy Moving K-means (Number of clusters =5)
	 Orange: Trees, Brown: Water, Blue: Hill top, Skyblue: Houses, Green: Roadway

Figure 2: classification results

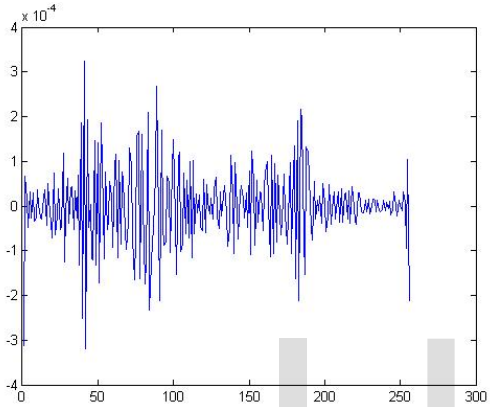
ISSN 2229-5518



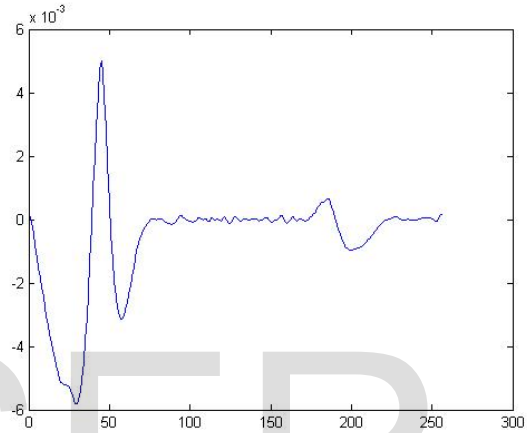
Histogram



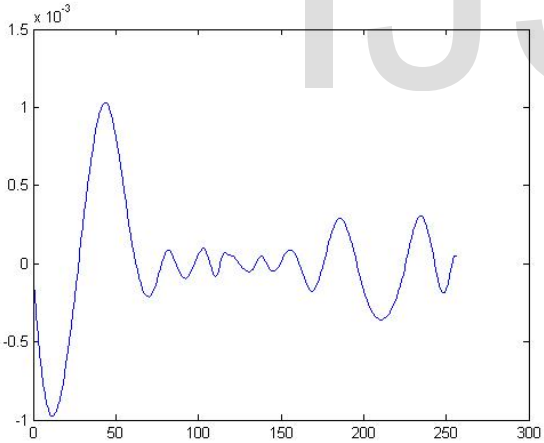
Probability Density Function



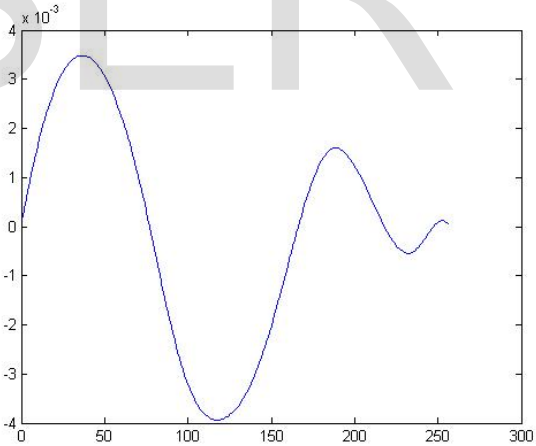
IMF1



IMF2



IMF3



IMF4

IJSER

ISSN 2229-5518

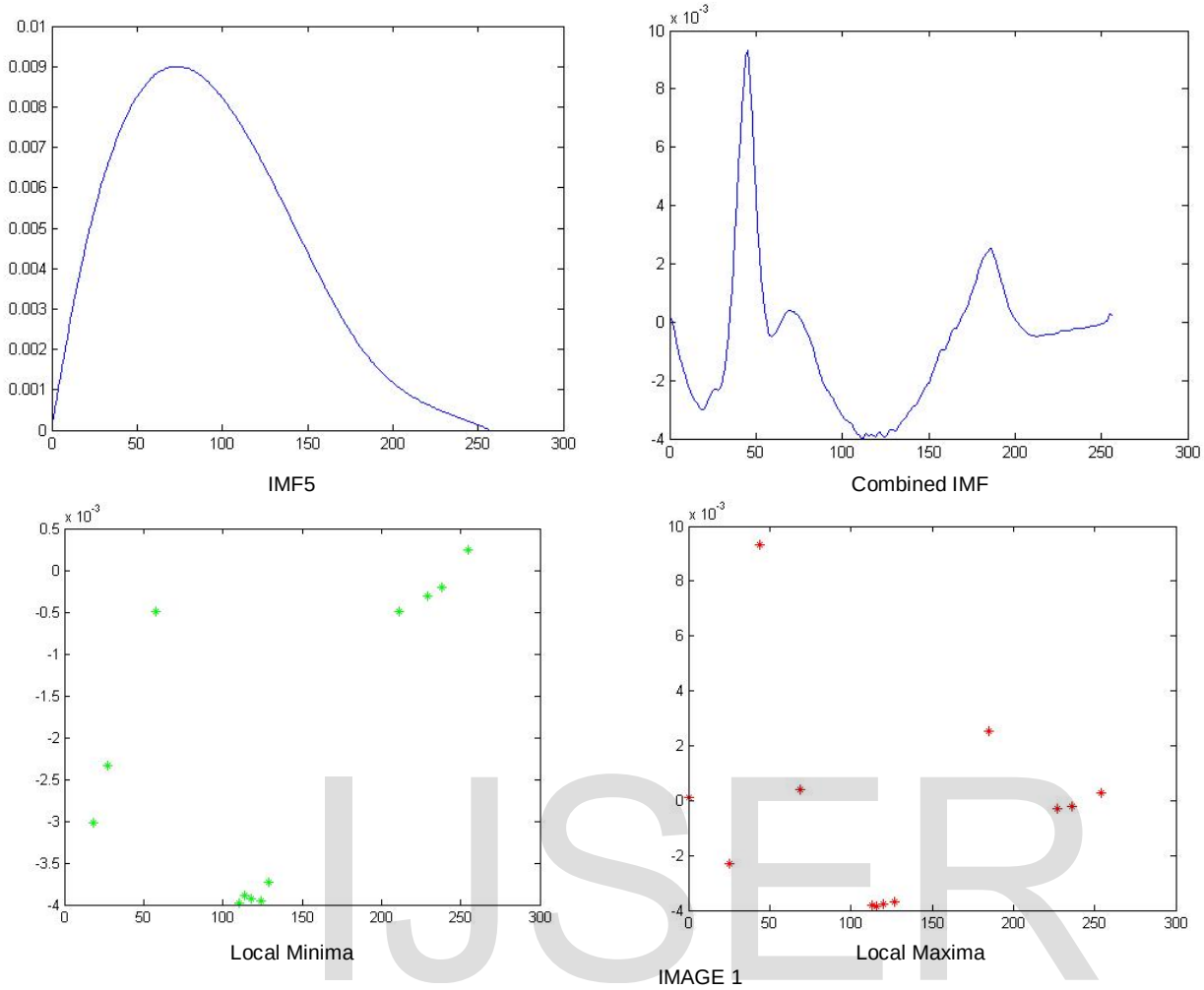


Figure 1: Results of ECEMD on image 1

IJSER