

# COMBINING FUZZY PROBABILITY AND FUZZY CLUSTERING FOR MULTISPECTRAL SATELLITE IMAGERY CLASSIFICATION

**Dinh-Sinh Mai<sup>\*</sup>, Le-Hung Trinh, Long Thanh Ngo**

*Le Quy Don Technical University, No.236 Hoang Quoc Viet Road, Bac Tu Liem, Hanoi*

\*Email: [maidinhsinh@gmail.com](mailto:maidinhsinh@gmail.com)

Received: 23 June 2015; Accepted for publication: 2 March 2016

## ABSTRACT

In practice, the classification algorithms and the initialization of the clusters and the initial centroid of clusters have great influence on the stability of the algorithms, dealing time and classification results. Some algorithms are used commonly in data classification, but their disadvantages are low accuracy and unstability such as k-Means algorithm, c-Means algorithm, Iso-data algorithm. This paper proposes a method of combining fuzzy probability and fuzzy clustering algorithm to overcome these disadvantages. The method consists of two steps, first to calculate the number of clusters and the centroid of clusters based fuzzy probability, then to use fuzzy clustering algorithm to land-cover classification. The results showed that, the accuracy of the land cover classification using multispectral satellite images according to the developed method significantly increases compared with various algorithms such as k-Means, Iso-data.

*Keyword:* satellite imagery, probability, fuzzy c-means clustering.

## 1. INTRODUCTION

The algorithms applied to image segmentation such as k-Means, c-Means, Iso-data show the same way based on the euclidean distance to determine the degree of similarity between the considered objects and cluster centroids. In problems of land cover classification, methods based on statistical parameters have been widely used because they are easy to implement and highly accurate [1 - 3]. However, these methods are quite expensive, time consuming and unsuitable.

Fuzzy logic has been widely applied in most of scientific and technical fields [4 - 7]. Typically in the clustering algorithms, it is fuzzy c-means algorithm (FCM) [8], which is quite common in many fields such as image processing, data mining etc.. With FCM algorithm - a loop is done to minimize the objective function by updating the membership function values, which have function as the weight values that exhibit degree of influence of a data sample on clusters. However, this algorithm does not perform well and is unstable when centroids initializing is far different from the real centroids.

There have been many improvements based on FCM algorithm to overcome the disadvantages of FCM algorithm. Despotovic et al. [9] used a mask with center considered pixel, the other pixels of the mask use the information about position on the mask to calculate the degree of similarity between the mask center with the neighboring pixels, then calibrate the value of the membership function of the FCM algorithm for image segmentation problem. Zhao Li et al. [10] also used spatial information to improve FCM algorithm, and the authors had to use FCM clustering as step initialization, then to use spatial information to eliminate noise and final used to FCM algorithm after noise reduction based on the value of the membership function. Zhengjian Ding et al. [11] improved FCM algorithm for land cover classification based on combination of spatial information and pixel values. These methods have certain limitations such as only applying on satellite image processing with high resolution, a method using multi-spectral satellite imagery, but the accuracy is not high or unsuitable. Hamed Shamsi et al. [12] improved the FCM algorithm by combining the spatial information of pixels surrounding areas to calculate the weights of the membership function, that having relevance to all data clusters.

In Vietnam, the studies related to satellite images have recently been conducted by several groups. Long Thanh Ngo et al. [13], Sinh Dinh Mai et al. [14] have researched on fuzzy logic and fuzzy logic type 2 applications in satellite image classification. Trinh Le Hung et al. [15] showed results in detection and classification of oil spills in envisat asar imagery using adaptive filter and fuzzy logic.

Currently, there are many methods to classify the satellite imagery, in which using fuzzy logic method has been interested and is widely studied because of their advantages [16 - 21].

In the present report, the authors proposed a new method, in which combining with fuzzy probability theory as the initial step for fuzzy clustering algorithm to classifying land-cover on satellite image. Experiments of the methods are implemented and compared with previous algorithms like Iso-data, k-Means to show the advantage of the proposed approach.

The paper is organized as follows: Section II shows background; Section III Proposed method, Section IV demonstrates how to apply the PFCM to land cover classification with some experiments; Section V is conclusion and future works.

## 2. BACKGROUND

### 2.1. Fuzzy Probability

Let us notice that the probability of a fuzzy event  $A \in F_B(R^n)$  could be expressed also in another way as a fuzzy set  $P_F(A)$  on  $[0,1]$  [1], [3]. Its membership function would be defined for any  $\tilde{p} \in [0,1]$  by the following formula:

$$P_F(A)(\tilde{p}) = \begin{cases} \sup \alpha \in (0,1) | \tilde{p} = p(A_\alpha) & \text{if } \alpha \in (0,1) | \tilde{p} = p(A_\alpha) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

It means, the fuzzy probability  $P_F(A)$  is uniquely determined by the probabilities of  $\alpha$  - cuts of  $A$ ,  $p(A_\alpha), \alpha \in (0,1]$ . The following relation between  $P(A)$  holds for any fuzzy event

$$A \in F_B(R^n) : P(A) = \int_a^b p(A_\alpha) d_\alpha .$$

As the fuzzy probability  $P_F$  seems to be too complicated to be used in practice, the crisp probability  $P$  will be preferred in this paper. Now, it will be shown how the fuzzy probability space can be applied to perform fuzzy discretization of continuous risk factors in decision making under risk. First, let us suppose that consequences of alternatives are affected by only one continuous risk factor  $Z$  whose probability distribution is given by a density function  $f(Z)$ . Consider a fuzzy scale  $A_1, A_2, \dots, A_n$  on the domain of the risk factor. As elements of the fuzzy scale are fuzzy random events, their probabilities  $P(A_i), i=1, \dots, n$ , are given by:

$$P(A_i) = \int_{SupA_i} A_i(z) f(z) dz$$

It is easy to check that  $\sum_a^b P(A_i) = 1$  and  $P(A_i) \geq 0, i=1, \dots, n$ . So,

a discrete probability distribution is defined on the given fuzzy scale. If the density function of the risk factor  $Z$  is not known, a similar probability distribution on the given fuzzy scale can be derived directly from measured data. If measurements  $z_1, z_2, \dots, z_m$  of  $Z$  are given,  $m \gg n$ , then probabilities of the fuzzy scale elements can be set by the formula:

$$P(A_i) = \frac{1}{m} \sum_{j=1}^m A_i(z_j), i=1, \dots, n \tag{2}$$

The fuzzy expected value and the fuzzy standard deviation of such a fuzzy random variable  $Z$  that takes on values  $A_i$  of the given fuzzy scale with probabilities  $P(A_i), i=1, \dots, n$  [2], are defined by the following formulas:

$$FEZ = \sum_{i=1}^n P(A_i) A_i \tag{3}$$

$$F\sigma Z = \sqrt{\sum_{i=1}^n P(A_i) (A_i - FEZ)^2} . \tag{4}$$

## 2.2. Fuzzy c-means clustering

In general, fuzzy memberships in FCM [8] achieved by computing the relative distance among the patterns and cluster centroids. Hence, to define the primary membership for a pattern, we define the membership using value of  $m$ . The use of fuzzifier gives different objective function as follows:

$$J_m(U, v) = \sum_{k=1}^N \sum_{i=1}^C (u_{ik})^m d_{ik}^2 \tag{5}$$

in which  $d_{ik} = \|x_k - v_i\|$  is Euclidean distance between the pattern  $x_k$  and the centroid  $v_i$ ,  $C$  is number of clusters and  $N$  is number of patterns. Degree of membership  $u_{ik}$  is determined as follow:

$$u_{ik} = \frac{1}{\sum_{j=1}^C \left( \frac{d_{ik}}{d_{jk}} \right)^{2/(m-1)}} \tag{6}$$

in which  $i = 1, \dots, C$ ;  $k = 1, \dots, N$ . Cluster centroids is computed as follows:

$$v_i = \frac{\sum_{k=1}^N (u_{ik})^m x_k}{\sum_{k=1}^N (u_{ik})^m} \quad (7)$$

in which  $i = 1, \dots, C$ . Next, defuzzification for FCM is made as if  $u_i(x_k) > u_j(x_k)$  for  $j = 1, \dots, C$  and  $I \neq j$  then  $x_k$  is assigned to cluster  $i$ .

### 3 COMBINING PROBABILITY THEORY AND FUZZY CLUSTERING SATELLITE IMAGE CLASSIFICATION

In fact, the image information is stored as numeric values so the problem of image partitions is usually based on the degree of similarity among these values to decide whether an object belongs to any region in the image. Therefore the key to determine a pixel will belong to certain area is based on the similarity in these colours, which is calculated through a function of the distance in the color space  $d_{ik}$  between the pattern  $x_k$  and the centroid  $v_i$ .

In that, the centroid will be in the samples that the density surrounding the sample data are large. The concept of statistical variance mathematical model is used to solve the problem of selecting a surrounding data points. To beginning we compute the expected pattern  $FEZ_i$  by the following equation:

$$FEZ_i = \sum_{i=1}^n P(x_i)x_i \quad (8)$$

and standard deviation  $F\sigma Z_i$ :

$$F\sigma Z_i = \sqrt{\sum_{i=1}^n P(x_i)(x_i - FEZ)^2} \quad (9)$$

with  $I = 1, 2, \dots, d$ ;  $X = (x_1, x_2, \dots, x_n)$ ,  $x \in R^d$ .

Consider the surround of each data point is  $m$ -dimensional box with radius defined by the standard deviation is  $R = \min_{1 \leq i \leq d} F\sigma Z_i$ . Compute density  $D_i$  of pattern  $x_i$ :

$$D_i = \sum_{j=1}^N T(R - |x_j - x_i|) \quad (10)$$

in which  $T = 1$  if  $z \geq 0$  otherwise  $T = 0$ .

Find pattern  $x_i$  with  $D_i = \max_{1 \leq j \leq N} D_j$  then  $V_k = V_k \cup x_i$  and  $X = X \setminus x_i$ . If  $X = \square$  given a set of candidate points  $V_k$ , else back to finding  $D_i$ .

If  $V_k$  is large then we can proceed with this algorithm to reduce the number of candidate clusters. We can speed up calculations by dividing the input data set into subsets, then proceed to apply the algorithm for that subset, we have candidates set  $V_i$ . Then we proceed with the candidate set  $\cup V_i = V$ , then apply this algorithm to the set  $V$ . The centroid matrix  $V$  can be initialized by choosing the patterns in  $V_k$  according to the density of candidates. The detailed algorithm consists of the following four main steps:

**Algorithm 1:** Find centroids using fuzzy probability.

**Step 1:** Initialization

- 1.1 Number of cluster  $C$ , ( $C > 1$ ).
- 1.2 Compute the  $FEZ_i$  by the formula (8).
- 1.3 Compute the  $F\sigma Z_i$  by the formula (9).

**Step 2:** Finding candidate

- 2.1. Compute density  $D_i$  by the formula (10).
- 2.2. Find pattern  $x_i$  with  $D_i = \max_{1 \leq j \leq N} D_j$  then  $V_k = V_k \cup x_i$  and  $X = X \setminus x_i$

**Step 3:** Check the stop condition:

If  $X = \emptyset$  or  $i > C$ , go to Step 5 else back to Step 2.

**Step 4:** Given a set of candidate points  $V_k$ .

Overall diagram of finding centroids using fuzzy probability is shown in Figure 1.

**Algorithm 2:** Probability Fuzzy C-means Clustering (PFCM)

**Step 1:** Initialization

- 1.1 The parameter of fuzzy  $m$ , ( $1 < m$ ), error  $e$ .
- 1.2 Initialization centroid  $V = [v_i], v_i \in R^n$  by algorithm 1.

**Step 2:** Compute the fuzzy partition matrix  $U$  and update centroid  $V$ :

- 2.1. Fuzzy partition matrix  $U_{ik}$  by the formula (6).
- 2.2. Update the cluster centroid  $V^j$  by the formula (7).

**Step 3:** Check the stop condition: If true, go to step 4, otherwise go to step 2.

**Step 4:** Given the clustering results.

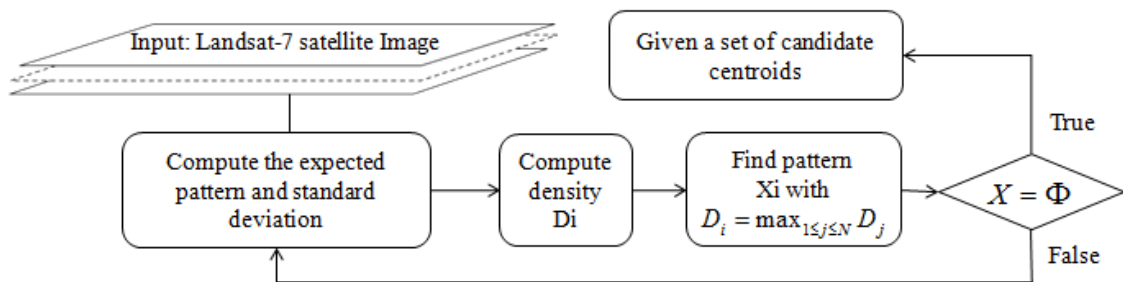


Figure 1. Diagram of finding centroids using fuzzy probability.

#### 4. LAND-COVER CLASSIFICATION USING PFCM

In the experiments, authors have selected the problem of classification on satellite imagery to test the proposed algorithm. The detailed algorithm of PFCM for land cover classification from multi-spectral satellite images consists of the following three main steps:

**Algorithm 3:** The PFCM algorithm

**Step 1:** Multi-spectral satellite imagery preprocessing.

**Step 2:** Apply PFCM on the n-bands of images. These n-bands will be classified into six classes representing six types of land covers:

1.  Class1: Rivers, ponds, lakes.
2.  Class2: Rocks, bare soil.
3.  Class3: Fields, grass.
4.  Class4: Planted forests, low woods.
5.  Class5: Perennial tree crops.
6.  Class6: Jungles.

**Step 3:** Compute percentage of the identical region:

$$S_i = n_i / N \quad (11)$$

where  $S_i$  be area of  $i^{th}$  region,  $n_i$  be the number of points of the  $i^{th}$  region,  $N$  be the total samples of n-bands imagery.

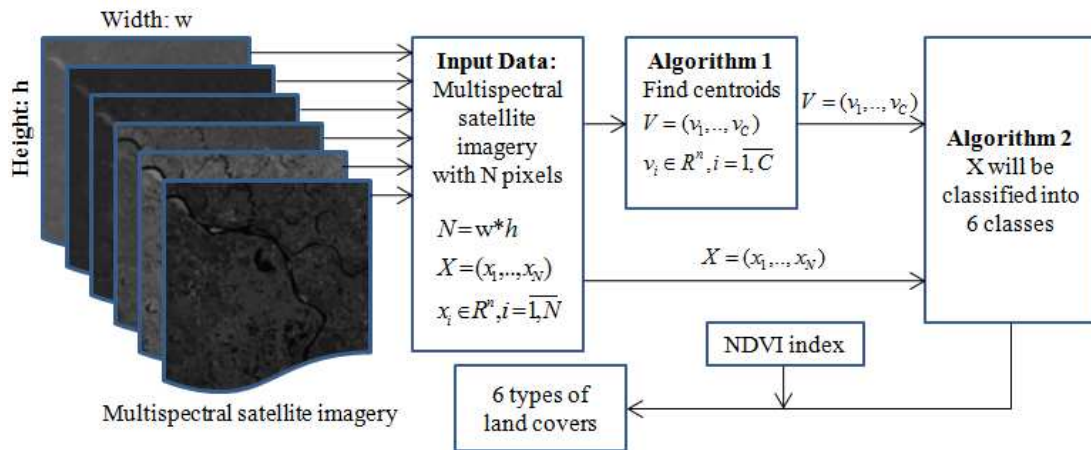


Figure 2. Overall diagram of classification problem.

Overall diagram of classification problem is illustrated in Figure 2, the multispectral satellite images are read into X array. Algorithm 1 will be made to find 6 approximate centroids corresponding to 6 layers of data, these centroids and X array will be input data to the algorithm 2, the algorithm 2 will conduct classify the pixels into 6 layers and based on Normalized Difference Vegetation Index (NDVI) [24] to determine six classes representing six types of land covers.

#### 4.1. Experiments 1

The study dataset from Landsat-7<sup>TM</sup> imagery is region center of Hanoi, Vietnam ( $21^{\circ} 10' 15.304''\text{N}$ ,  $105^{\circ} 29' 28.173''\text{E}$  to  $20^{\circ} 52' 34.401''\text{N}$ ,  $106^{\circ} 09' 57.317''\text{E}$ ) in Figure 3, its area:  $871.24 \text{ km}^2$ .

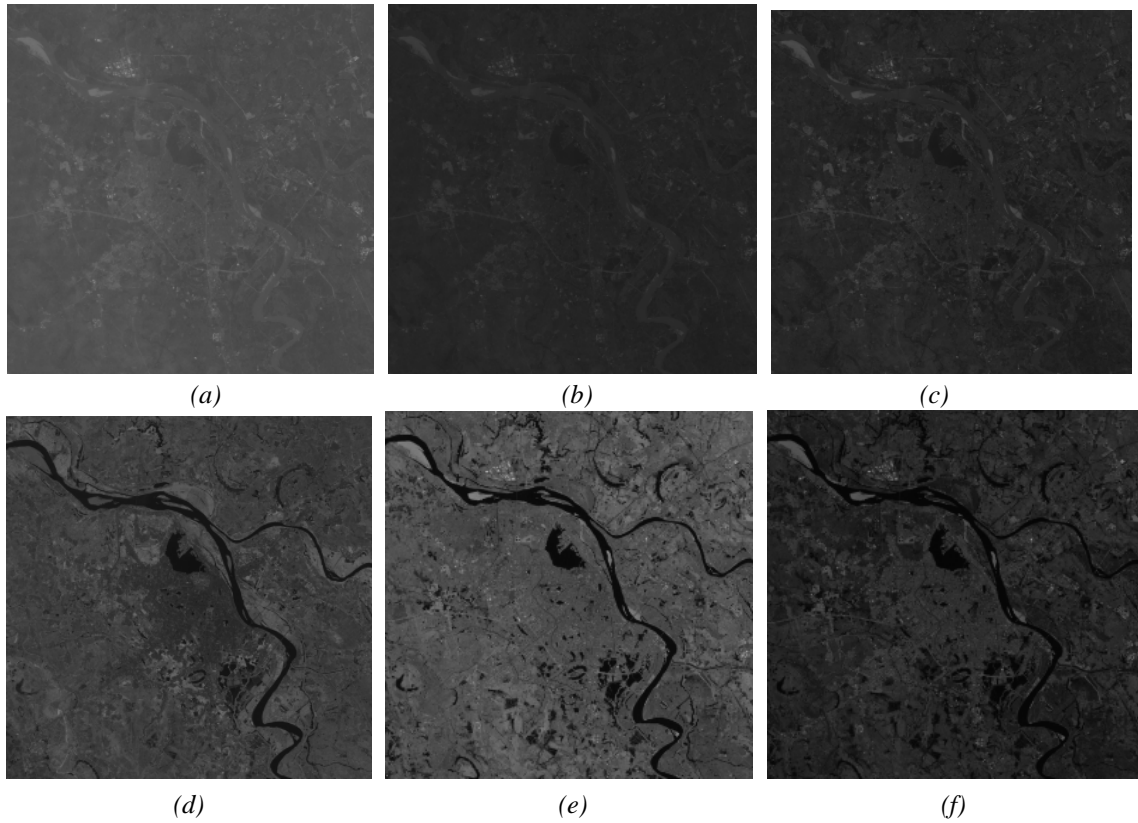


Figure 3. Study data of Hanoi: a) Band 1; b) Band 2; c) Band 3; d) Band 4; e) Band 5; f) Band 7.

Table 1. Results of land cover classification in Hanoi.

Class	PFCM (%)	FCM (%)	Iso-data (%)	K-means (%)
1	4.5263	4.8804	5.5866	9.9213
2	13.3306	14.3340	15.7601	16.9050
3	22.8785	22.0145	19.3886	16.7701
4	26.1571	25.6635	24.3211	21.2313
5	21.4329	20.3387	20.2183	19.0090
6	11.6747	12.7688	14.7253	16.1633

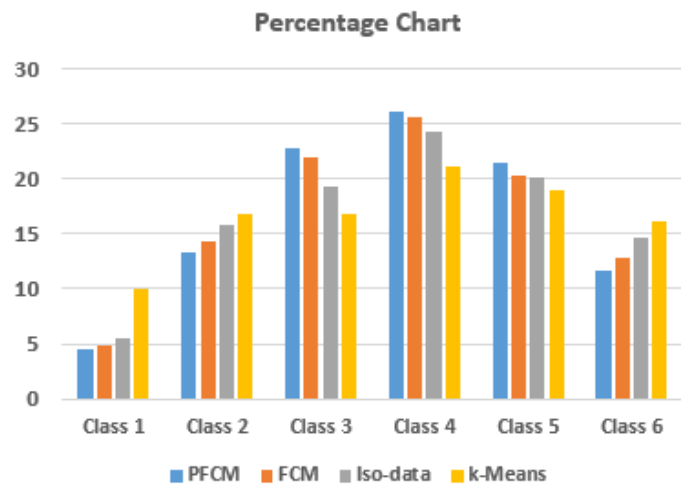


Figure 4. The result of algorithms: PFCM, FCM, Iso-data and K-Means.

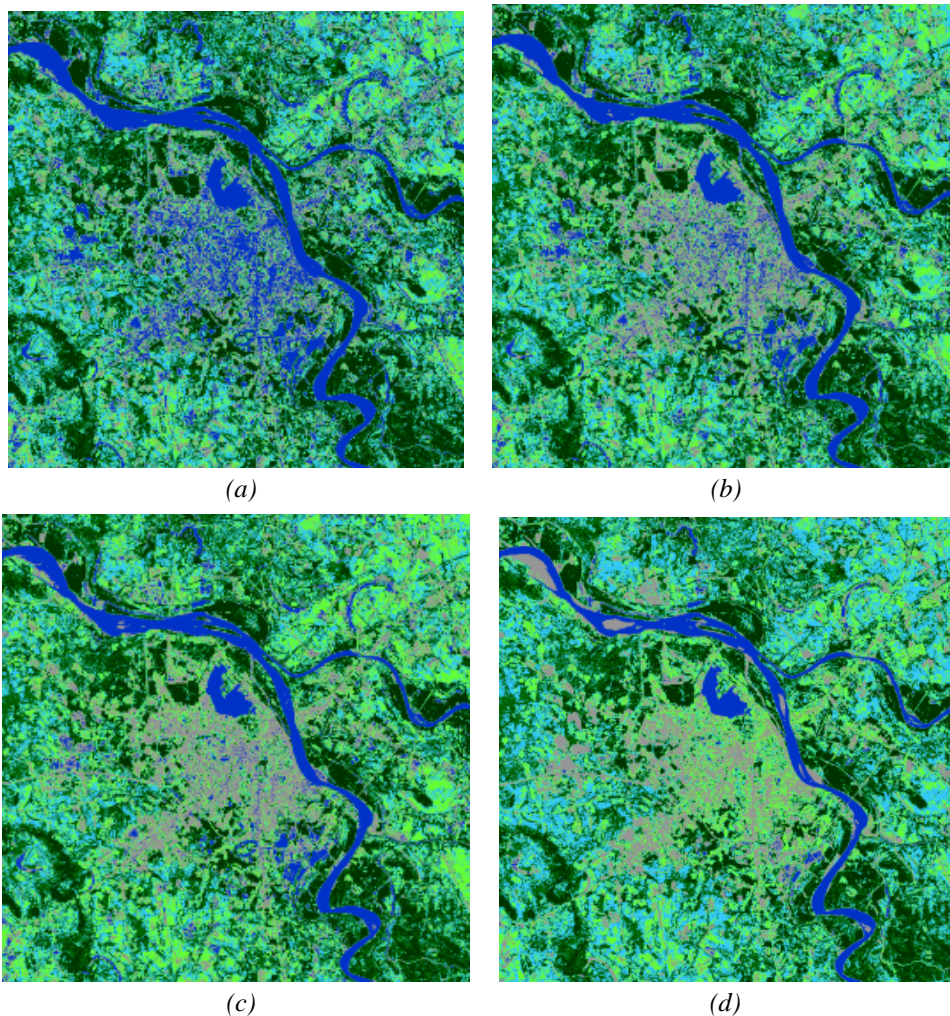


Figure 5. Result of land cover classification. a) K-Means; b) Iso-data; c) FCM; d) PFCM.



The results are shown in Figure 5 for 6 bands in which (a), (b), (c) and (d) are classification results of PFCM, FCM, Iso-data and k-Means algorithms, respectively. Figure 4 and Table 1 compare classification results between PFCM, FCM, Iso-data and k-Means. There is a significant difference between the algorithms of PFCM, FCM, k-Means and Iso-data in classifying based on estimating the area of regions. This result showed that the area of the layers is different, the biggest difference between k-Means and PFCM.

To assessing the performance of the algorithms on the experimental images we analyzed the results on the basis of several validity indexes. We considered the different validity indexes such as the Bezdeks partition coefficient (PC-I) [22], Classification Entropy index (CE-I) [23,24] and Kappa index. The values of these validity indexes are shown in the Table 2.

Table 2. The various validity indexes on the LANDSAT-7 images of Hanoi area.

Validity Index	K-means	Iso-data	FCM	<b>PFCM</b>
CE-I	0.9869	0.5872	0.1972	<b>0.1317</b>
PC-I	0.6982	0.7282	0.8628	<b>0.8893</b>
Kappa	0.4182	0.4882	0.7628	<b>0.9156</b>

Note that the validity indexes are proposed to evaluate the quality of clustering. The better algorithms have smaller values of CE-I and larger value of PC-I, Kappa. The results in Table 2 show that the PFCM have better quality clustering than the other typical algorithm such as FCM, K-means and Iso-data.

## 4.2. Experiments 2

The authors using Landsat-7 satellite image data, which taken Lamdong area on 12/02/2010, 12° 13'01.88"N, 107° 33'27.511"E to 11° 37'40.927"N, 108° 49'49.252"E and square of area: 3393.7 hectares, see in Figure 6.

The results are shown in Figure 7 in which (a), (b), (c) and (d) are the classification results of PFCM, FCM, Iso-data and K-means algorithms, respectively. Figure 8 and Table 3 compare classification results between PFCM, FCM, Iso-data and k-Means. There is a significant difference between the algorithms of PFCM, FCM, k-Means and Iso-data in classifying based on estimating the area of regions. In Figure 7, the results show that PFCM algorithm noise reduction quite good, while K-means algorithm is much the most noise. Table 4 show that the PFCM have better quality clustering than the other typical algorithm such as FCM, K-means and Iso-data.

In summary, from two test areas, these deviations can be explained that the boundary of water and soil classes are usually quite clear, while the vegetation classes are often confused between grasses and trees. With satellite imagery resolution 30mx30m, the differences of classification results can be acceptable in assessment of land cover on a large area, reducing costs compared to other methods. This result not only makes predictions about the land cover fluctuations but also supports urban planning, natural resources management and so on.

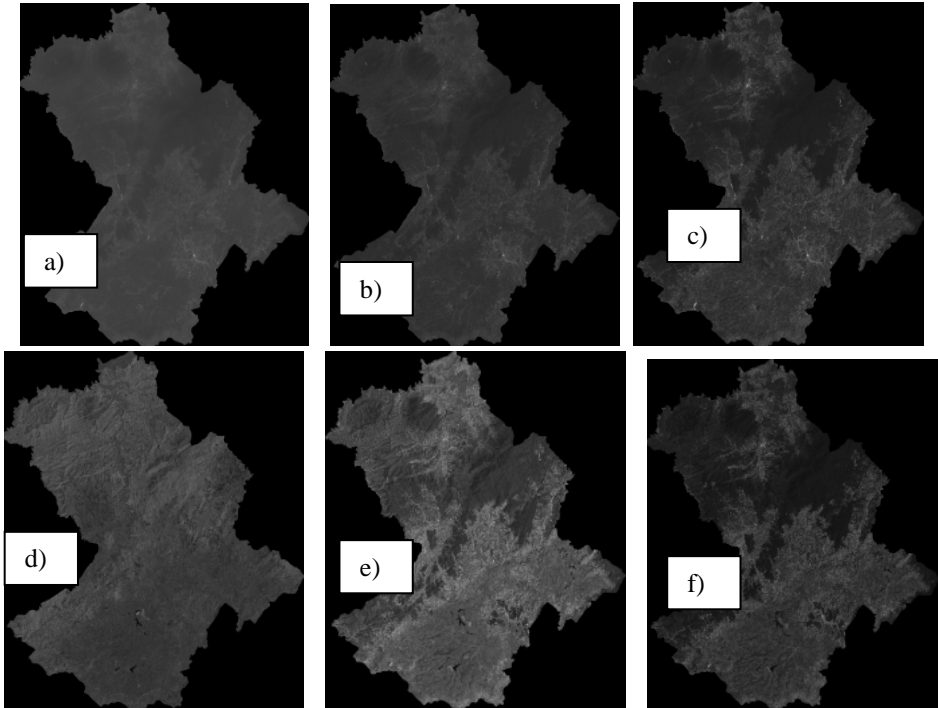


Figure 6. Study data of Lamdong: a) Band 1; b) Band 2; c) Band 3; d) Band 4; e) Band 5; f) Band 7.

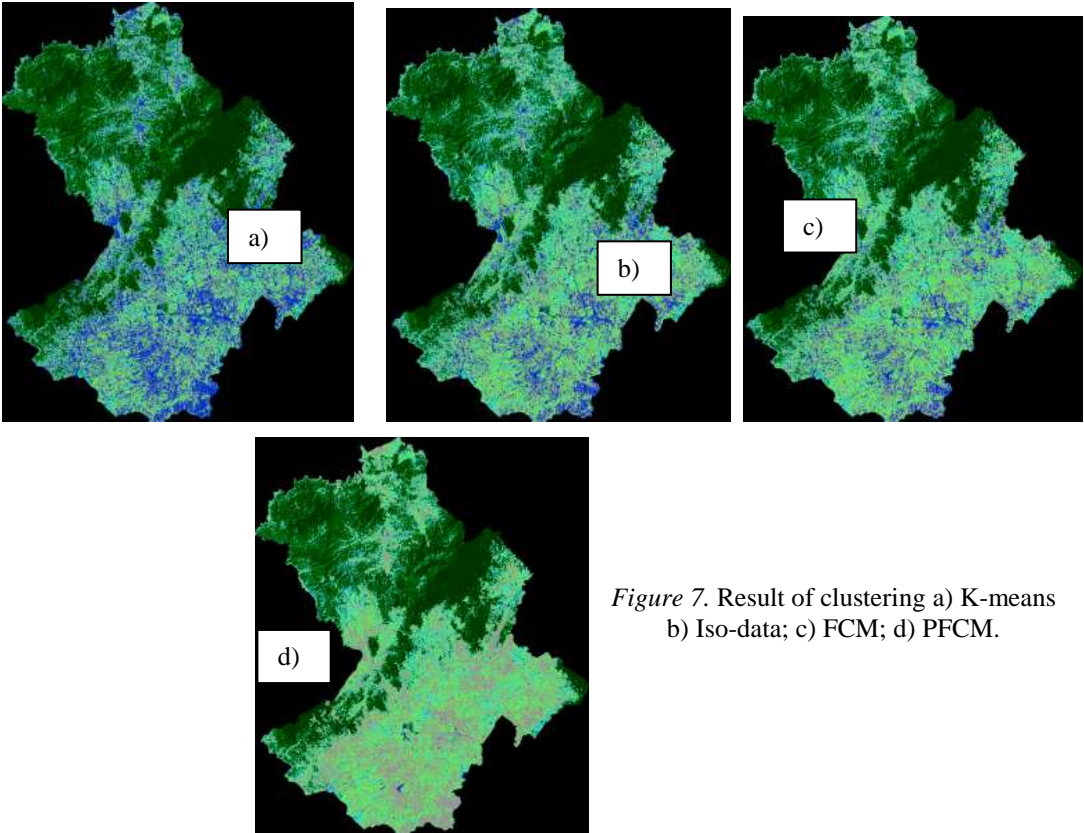


Figure 7. Result of clustering a) K-means  
b) Iso-data; c) FCM; d) PFCM.

Table 3. Results of land cover classification in Lamdong (%).

Class	PFCM (%)	FCM (%)	Iso-data (%)	K-means (%)
1	8.3890	9.2619	12.3099	17.1510
2	20.0359	19.4947	17.8593	15.5828
3	19.8786	18.5279	15.9935	13.4288
4	15.7184	15.0599	13.9328	11.4685
5	19.0240	20.1407	21.2678	21.3346
6	16.9540	17.5149	18.6368	21.0342

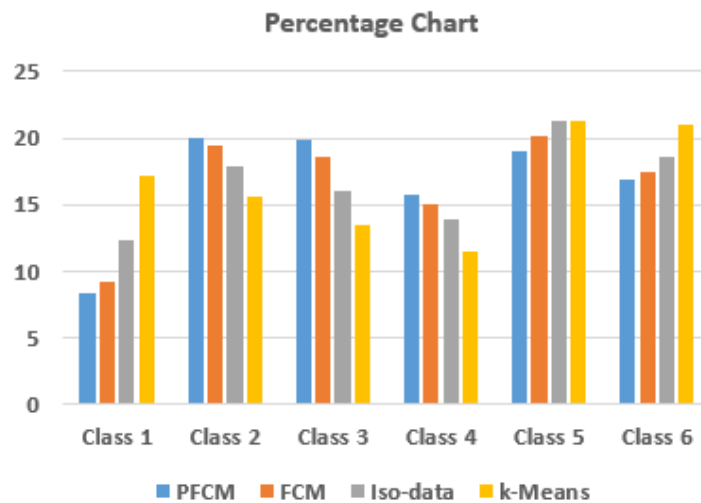


Figure 8. The result of algorithms: PFCM, FCM, Iso-data and K-Means.

Table 4. The various validity indexes on the LANDSAT-7 images of Lamdong area

Validity Index	K-means	Iso-data	FCM	<b>PFCM</b>
CE-I	0.9629	0.6581	0.2271	<b>0.1429</b>
PC-I	0.5817	0.7022	0.7843	<b>0.8891</b>
Kappa	0.2989	0.3986	0.7982	<b>0.8599</b>

### 5. CONCLUSION

This paper presents a new method by combined fuzzy probability theory as the initial step for fuzzy clustering algorithm to classify land-cover on satellite image. The results showed that the proposed algorithm has improved the quality of clusters for a problem class of land cover classification. Based on the Landsat-7 satellite images many experiments of land cover classification were done. Besides, the proposed approach can be applied to other types of satellite images, which saves costs and time compared to other ways of land cover change detection.

The next goal is to implement further research on the Landsat-8 satellite images, hyper-spectral satellite imagery for environmental classification, assessment of land surface temperature changes; speed-up the proposed methods based on GPUs platforms.

## REFERENCES

1. Lotfi Zadeh - Fuzzy Set Theory and Probability Theory: What is the Relationship? International Encyclopedia of Statistical Science (2011) 563-566.
2. Jana Talasova and Ondrej Pavlacka - Fuzzy Probability Spaces and Their Applications in Decision Making, Austrian Journal of Statistics **35** (2006) 347-356.
3. Denis de Brucq, Olivier Colot, Arnaud Sombo - Identical Foundation of Probability theory and Fuzzy Set Theory, Proceedings of the Fifth International Conference on Information Fusion **II**, USA, 2002, pp. 1442 - 1449.
4. Naoko Iino, Kisei Kinoshita and Chikara Kanagaki - Satellite Images of air pollutants and Land Cover for Environmental education and disaster prevention, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science **XXXVI**, Part 6, Tokyo Japan, 2006.
5. Cihlar J. - Land cover mapping of large areas from satellites: status and research priorities, Int. j. remote sensing **21** (6 & 7) (2000) 1093-1114.
6. Stavrakoudis, D. G., Galidaki, G. N., Gitas, I. Z., and Theocharis, J. B. - Enhancing the Interpretability of Genetic Fuzzy Classifiers in Land Cover Classification from Hyperspectral Satellite Imagery. IEEE World Congress on Computational Intelligence, 2010, pp. 1277-1284.
7. Ali Asghar Torahi, Suresh Chand Rai - Land Cover Classification and Forest Change Analysis, Using Satellite Imagery - A Case Study in Dehdez Area of Zagros Mountain in Iran, Journal of Geographic Information System **3** (2011) 1-11.
8. James C. Bezdek, Robert Ehrlich, William Full - FCM: The Fuzzy C-Means clustering algorithm, Computers and Geosciences **10** (2) (1984) 191-203.
9. Despotovic I., Goossens B., Vansteenkiste E., Philips W. - An improved fuzzy clustering approach for image segmentation, 17th IEEE International Conference on Image Processing, 2010, pp. 249-252.
10. Zhao Li, Xiaoming Zhou - Classifications Modification Based FCM with Spatial Information for Image Segmentation, 3rd International Conference on Multimedia Technology, 2013, pp. 731-738.
11. Zhengjian Ding, Jin Sun, and Yang Zhang - FCM Image Segmentation Algorithm Based on Color Space and Spatial Information, International Journal of Computer and Communication Engineering **2** (2013) 48-51.
12. Hamed Shamsi and Hadi Seyedarabi - A Modified Fuzzy C-Means Clustering with Spatial Information for Image Segmentation, International Journal of Computer Theory and Engineering **4** (5) (2012) 762-766.

13. Long Thanh Ngo and Dinh Dung Nguyen - Land cover classification using interval type-2 fuzzy clustering for multi-spectral satellite imagery, IEEE Conference on Systems, Man, and Cybernetics, 2012, pp. 2371 - 2376.
14. Sinh Dinh Mai, Long Thanh Ngo - Interval Type-2 Fuzzy C-Means Clustering with Spatial Information for Land-Cover Classification, The 7th Asian Conference on Intelligent Information and Database Systems, part I, Springer LNAI 9011, 2015, pp.387-397.
15. Trinh Le Hung, Mai Dinh Sinh - Detection and classification of oil spills in envisat asar imagery using adaptive filter and fuzzy logic, Journal of Petrovietnam **5** (2014) 49-55.
16. Rauf K. S., Valentin V. G., Leonid P. P. - Fuzzy clustering methods in Multi-spectral Satellite Image Segmentation, International Journal of Computing **8** (2009) 87-94.
17. Han J. G., Chi K. H., and Yeon Y. K. - Land Cover Classification of IKONOS Multispectral Satellite Data: Neuro-fuzzy, Neural Network and Maximum Likelihood Methods, Lecture Notes in Computer Science, (3642), 2005, pp. 251-262.
18. Gordo O., Martinez E., Gonzalo C., Arquero A. - Classification of Satellite Images by means of Fuzzy Rules generated by a Genetic Algorithm. Latin America Transactions, Revista IEEE America Latina **9** (1) (2013) 743-748.
19. Genitha C. H. and Vani K. - Classification of satellite images using new Fuzzy cluster centroid for unsupervised classification algorithm, IEEE Conference on Information and Communication Technologies, 2013, pp. 203-207.
20. Shackelford A. K. and Davis C. H. - A fuzzy classification approach for high-resolution multispectral data over urban areas, IEEE International Geoscience and Remote Sensing Symposium **3** (2002) 1621-1623.
21. Eric K. Forkuo, Adubofour Frimpong - Analysis of Forest Cover Change Detection, International Journal of Remote Sensing Applications **2** (4) (2012) 82-92.
22. Wang W. and Zhang Y. - On fuzzy cluster validity indices, Fuzzy Sets and Systems **158** (2007) 2095-2117.
23. Bezdek J., Pal N. - Some new indexes of cluster validity. IEEE Transactions on Systems, Man and Cybernetics **3** (1998) 301-315.
24. Ray D. Jackson and Alfredo R. Huete - Interpreting vegetation indices, Preventive veterinary Medicine, Elsevier **11** (1991) 185-200.

## **TÓM TẮT**

### **KẾT HỢP XÁC SUẤT MỜ VÀ PHÂN CỤM MỜ PHÂN LOẠI ẢNH VỆ TINH ĐA PHỔ**

Mai Đình Sinh\*, Trịnh Lê Hùng và Ngô Thành Long

*Học viện Kỹ thuật Quân sự, 236 Hoàng Quốc Việt, Bắc Từ Liêm, Hà Nội, Việt Nam*

\*Email: [maidinhsinh@gmail.com](mailto:maidinhsinh@gmail.com)

Trên thực tế đối với các thuật toán phân loại, việc khởi tạo số lượng cụm và trọng tâm các cụm ban đầu có ảnh hưởng lớn đến độ ổn định của thuật toán, thời gian xử lý và kết quả phân loại. Một số thuật toán được sử dụng phổ biến trong phân loại dữ liệu, nhưng nhược điểm của chúng là độ chính xác thấp và không ổn định như thuật toán k-Means, c-Means, Iso-data. Bài báo đề xuất một phương pháp kết hợp xác suất mờ và thuật toán phân cụm mờ để khắc phục một số nhược điểm này. Phương pháp này bao gồm 2 bước, thứ nhất tính toán số cụm và trọng tâm các cụm dựa trên xác suất mờ, sau đó sử dụng thuật toán phân cụm mờ để phân loại lớp phủ. Các kết quả cho thấy rằng, độ chính xác khi phân loại lớp phủ sử dụng ảnh vệ tinh đa phổ theo phương pháp đề xuất tăng đáng kể khi so sánh với một số thuật toán phổ như k-Means, Iso-data.

*Từ khóa:* ảnh vệ tinh, xác suất, phân cụm mờ c-Means.