Advanced Semi-supervised Possibilistic Fuzzy C-means Clustering using Spatial-Spectral distance for Land-cover Classification

Dinh-Sinh Mai*, Long Thanh Ngo and Le-Hung Trinh
Le Quy Don Technical University, Hanoi, Vietnam
*maidinhsinh@gmail.com, ngotlong@gmail.com, trinhlehung125@gmail.com

***Abstract*—With the explosion of information, characteristics of increasingly complex data, the use of traditional methods in data processing has proved ineffective. Computer applications are increasingly becoming important and essential in many areas such as biology, medicine, psychology, economics, image processing and many other disciplines. A variety of multi-spectral satellite image classification, clustering algorithms have been developed and applied to analyze the surface of the earth. In this paper, we propose a novel semi-supervised possibilistic fuzzy c-means clustering on spatial-spectral distance (SPFCM-SS) for multi-spectral image land-cover classification by the extension of the possibilistic fuzzy C-means (PFCM) algorithm, in which spectral information and spatial information of the pixels are used coupled with labelled data to increase the accuracy of clustering results when the data structure of input patterns is non-spherical and complex. Experiments were performed for multi-spectral satellite image data and clustering efficiency indexes were used to compare the performance of the proposed algorithm with other similar algorithms.**

***Index Terms*—semi-supervised, PFCM, spatial information, multi-spectral image.**

I. INTRODUCTION

In general, fuzzy memberships in fuzzy c-means clustering algorthm (FCM) achieved by computing the relative distance among the patterns and cluster centroids [1]. Hence, to define the primary membership of a pattern, FCM algorithm define the membership using value of m. However, FCM algorithm uses a membership function, so it does not describe all the characteristics of the data.

A variant of fuzzy clustering is based on possibilistic approach which was first proposed in [2]. The algorithm determines a possibilistic partition in which a possibilistic membership is used to define the absolute degree of typicality of a point in any particular clusters. The larger the distance between an object to a centroid, the lower the possibilistic membership grade and the lower the object affects on clustering of the centroid because the algorithm easily falls into identical clusters. To overcome this difficulty making algorithm more robust by inheriting the characteristics of PCM, a novel semi-supervised clustering technique titled semi-supervised possibilistic clustering (sPCM) is proposed in [3].

Zhang et al. [4] proposed the fuzzy possibilistic c-means (FPCM) model and algorithm that generates both membership and typicality values when clustering unlabeled data. Therefore, methods of outlier detection or noise removal may be applied. However, the possibilistic approach still has some drawbacks such as identical clusters and choosing its parameters. Therefore, Nikhil et al. [5] proposed a model called possibilistic fuzzy c-means (PFCM) model. PFCM is a hybridization of possibilistic c-means (PCM) and fuzzy c-means (FCM) that often avoids various problems of PCM, FCM and FPCM. PFCM solves the noise sensitivity defect of FCM, overcomes the coincident clusters problem of PCM and eliminates the row sum constraints of FPCM.

Unsupervised classification can handle a lot of data but the method is complex and sometimes results look a bit strange. Therefore recently, semi-supervised classification has been studied [6]. This classification has advantages of both unsupervised and supervised method of classification. Sinh et al. [7] proposed a semi-supervised fuzzy c-means clustering (SFCM) for change detection from multi-spectral satellite image by using the criteria to move the prototype of clusters to the expected centroids which are pre-determined on the basis of samples. The proposed algorithm is used for a model of change detection on multi-spectral satellite imagery at multiple temporals.

By adding a spatial regularizer into the fuzzy hyperplane based objective function, the proposed method can take into account additional important information of inherently spatial data. Sinh et al. [9] proposed satellite image classification method by using spatial information for spectral fuzzy clustering algorithm. Another approach to spatial information in [8], exploits local spatial information between the pixel and its neighbors to compute the membership degree by using an interval type-2 fuzzy clustering algorithm. The proposed algorithm is applied to the problems of satellite image analysis consisting of land-cover classification and change detection. Liu et al. [10] have come up with a novel fuzzy spectral clustering algorithm with robust spatial information for image segmentation. Vargas et al. [11] introduced two enhanced fuzzy c-means clustering algorithms with spatial constraints for noisy color image segmentation. Zhao et al. [12] also included spatial information in the objective function of a certain generalized fuzzy C-means clustering algorithm, and then the kernel induced distance is adopted to substitute the Euclidean distance in the new objective function. Liu and Pham [13] presented a fuzzy clustering algorithm which can handle spatial constraints, which is based on the notions of

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00739

hyperplanes, fuzzy C-means, and spatial constraints.

In this paper, we have proposed novel semi-supervised possibilistic fuzzy c-means clustering on spatial-spectral distance (SPFCM-SS) for multi-spectral image land-cover classification by the extension of the possibilistic fuzzy C-means (PFCM) algorithm, in which spectral information and spatial information of the pixels are used coupled with labelled data to increase the accuracy of clustering results when the data structure of input patterns is non-spherical and complex.

The paper is organized as follows: Section II briefly introduces some backgrounds about fuzzy clustering, possibilistic fuzzy c-means clustering; Section III proposes the semi-supervised possibilistic fuzzy c-means clustering based on spatial-spectral distance; Section IV offers some experimental results and section V covers a conclusion and proposes future research directions.

II. Background

Possibilistic Fuzzy C-Means Clustering Algorithm (PFCM) [5] was proposed by Nikhil et al. PFCM algorithm is built based on the hybridization of two algorithms FCM and PCM, which has two types of memberships: 1) A possibilistic membership that measures the absolute degree of typicality of a point in any particular cluster, and 2) a fuzzy membership that measures the relative degree of sharing of a point among the clusters. The objective function for PFCM was built as follows:

$$\begin{aligned} J_{m,\eta}(U,T,V,X,\beta) = \sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik}^{m}+\tau_{ik}^{\eta})\|v_i - x_k\|^2 \\ + \sum_{i=1}^{c}\beta_i\sum_{k=1}^{n}(1-\tau_{ik})^{\eta} \end{aligned} \tag{1}$$

Subject to the constraints:

$$m > 1, \eta > 1; 0 \leq \mu_{ik}, \tau_{ik} \leq 1; \sum_{i=1}^{c}\mu_{ik} = 1; \sum_{k=1}^{n}\tau_{ik} = 1 \tag{2}$$

In which, n is the total number of patterns in a given data set, c is the number of clusters, X are data characteristics, V are the centers of the clusters, U is a fuzzy partition matrix, which contains the membership degree, T is a typicality partition matrix, which contains the membership degree, m is the weighting exponent for fuzzy partition matrix and η is the weighting exponent for typicality partition matrix. β_i are user defined constants. The constants defines the relative importance of fuzzy membership and typicality values in the objective function.

The corresponding centers of the clusters and membership degree to each respective data to solve the optimization problem with the constraints in 2 are given by Eqs. 3, 4 and 5, which provide an iterative procedure. The aim is to improve a sequence of fuzzy clusters until no further improvement in $J_{m,\eta}$ can be obtained.

$$v_i = \frac{\sum_{k=1}^{n}(\mu_{ik}^{m}+\tau_{ik}^{\eta})x_i}{\sum_{k=1}^{n}(\mu_{ik}^{m}+\tau_{ik}^{\eta})} \tag{3}$$

$$\mu_{ik} = 1/\sum_{j=1}^{c}\left(\frac{\|v_i - x_k\|}{\|v_j - x_k\|}\right)^{2/(m-1)} \tag{4}$$

$$\tau_{ik} = 1/\left(1 + (\frac{1}{\beta_i}\|v_i - x_k\|)^{1/(\eta-1)}\right) \tag{5}$$

Equations 3, 4 and 5 are an iterative optimization procedure. The aim is to improve a sequence of fuzzy clusters until no further improvement in $J_{m,\eta}$ can be made. The PFCM algorithm consists of the following steps:

Algorithm 1: The PFCM algorithm

1. Given a preselected number of clusters c and a chosen value for m and η, initialize the fuzzy partition matrix and the typically partition matrix with constraint in 2.
2. Calculate the center of the fuzzy cluster, v_i for $i = 1, 2, , c$ using Eq. 3.
3. Use Eq. 4 to update the fuzzy membership U.
4. Use Eq. 5 to update the typically membership T.
5. If the improvement in $Max(|\mu_{ik}^{(t+1)} - \mu_{ik}^{(t)}|)$ is less than a certain threshold (ϵ), then stop; otherwise, go to step 2.

III. Semi-supervised possibilistic fuzzy c-means clustering on spatial-spectral distance

A. Spatial-spectral distance

A multi-spectral image is a collection of several monochrome images of the same scene, each of them taken with a different sensor, each image is referred to as a band. To fully exploit the additional information which is contained in the multiple bands, we should consider the images as one multi-spectral image rather than as a set of monochrome graylevel images. For an image with k bands, we can then describe the brightness of each pixel as a point in a k-dimensional space represented by a vector of length k is $x_i = (b_{i1}, b_{i2}, ..., b_{ik})$ and then multi-spectral image data can be described as a set $X = [x_1, x_2, ..., x_n]$. The spectral distance (SD) is calculated as the Euclidean distance in the spectral space:

$$SD_{ik} = \|x_i - x_k\| \tag{6}$$

The next content will introduce the spatial relationship between the pixels in the image. Image data is made up of pixels, where each pixel can contain different types of information, including its position in the image, color, etc. The position, or coordinate of a pixel in an image is information related to the spatial relationship, which is important in high-level image processing. This can be explained by the fact that, processing algorithms are independent of pixel level, there are many algorithms, processors based on regions or uses information on the whole image.

In image segmentation, the key to determine a pixel belonging to certain area is based on the similarity of these colors. However, the shape and structure of the cluster also has a certain influence on the data clustering. Which means that together with information about the color of the pixel, the spatial information of pixels also need to be considered when clustering data. We use a mask of size SxS to slip on the image, the center pixel of mask is the considered pixel. The number of neighboring pixels is determined corresponding to the selected type of mask i.e., 8 pixels for the 8-directional mask or 4 pixels for the 4-directional mask. To determine the degree of influence of a neighboring pixels for the center pixels, a measure spatial information (SI) is defined on the basic of the degree μ_{ij} and the attraction distance as follows:

$$SI_{ik} = \frac{\sum_{j=1}^{N} \mu_{\mathrm{ij}} \left(\sqrt{(a_k - a_j)^2 + (b_k - b_j)^2} \right)^{-1}}{\sum_{j=1}^{N} \left(\sqrt{(a_k - a_j)^2 + (b_k - b_j)^2} \right)^{-1}} \quad (7)$$

In which, $SI_{ik} \in (0,1)$, μ_{ij} is the membership degree of the neighboring element x_j to the cluster i. The distance attraction $d_{kj} = \sqrt{(a_k - a_j)^2 + (b_k - b_j)^2}$ is the squared Euclidean distance between elements (a_k, b_k) and (a_j, b_j). According to this formula, the value of spatial information is at greater pixel on the mask while many neighboring pixels they have similar color them and the opposite. We use the inverse distance $\left(\sqrt{(a_k - a_j)^2 + (b_k - b_j)^2} \right)^{-1}$ because the closer the neighbors x_j is to the center x_k the more influence it has on the result.

The centers are moved to seed locations corresponding to the lowest gradient position in a SxS neighborhood. This is done to avoid centering a pixel on an edge and to reduce the chance of seeding a pixel with a noisy pixel. Since the expected spatial extent of a pixel is a region of approximate size SxS, the search for similar pixels is done in a region SxS around the pixel center. Distance function based on spectral space and spatial information is used to measure the similarity between pixels, a spatial-spectral distance (SSD) measure is defined as follows:

$$SSD_{ik} = SD_{ik}(1 - \alpha \frac{SI_{ik}}{S}) = \|x_i - x_k\| (1 - \alpha \frac{SI_{ik}}{S}) \quad (8)$$

In which, SD_{ik} and SI_{ik} is the distance in the spectral space and the spatial information of the two pixels, respectively and α is the control parameter. By defining SSD_{ik} in this manner, α also allows us to weigh the relative importance between spectral similarity and spatial proximity. $\alpha \in [0, S]$, when α is large, spatial proximity is more important, when $\alpha = 0$, the new distance is the distance in the spectral space.

B. Semi-supervised method

Because of physical properties of electromagnetic spectrum when reflect from land cover surface, the centroid of clusters are fixed for all region of imagery. Let c be the number of clusters, calculation c centroids, $v_1^*, v_2^*, ..., v_c^*$ from the labeled pixel dataset and $V^* = [v_1^*, v_2^*, ..., v_c^*]$ is the set of expected cluster centroids, $\|v_i - v_i^*\|$ is a measure of the difference between the computing clusters and sampling cluster. Using the set of pre-determined centroids to a just new centroids to move closer to the expected centroids.

The objective function $J_{m,\eta}$ of the PFCM algorithm is changed as follows:

$$J_{m,\eta}(U,T,V,X,\beta) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik}^m + \tau_{ik}^\eta)(SSD_{ik}^2 + \|v_i - v_i^*\|^2) + \sum_{i=1}^{c} \beta_i \sum_{k=1}^{n} (1 - \tau_{ik})^\eta \quad (9)$$

To minimize the objective function, method of Lagrange is used to find the solution by function:

$$J_{m,\eta}(U,T,V,X,\lambda,\beta) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik}^m + \tau_{ik}^\eta)(SD_{ik}^2(1 - \alpha \tfrac{SI_{ik}}{S})^2 + \|v_i - v_i^*\|^2) + \sum_{k=1}^{n} \lambda_k \sum_{i=1}^{c} (1 - \mu_{ik})^m + \sum_{i=1}^{c} \beta_i \sum_{k=1}^{n} (1 - \tau_{ik})^\eta \quad (10)$$

With $0 < \lambda_k, \beta_i < 1$ are user defined constants. Based on the Lagrange method, by taking derivative of $J_{m,\eta}(U,T,V,X,\lambda,\beta)$ over μ_{ik}, τ_{ik}, v_i and assuming the result zero.

$$\frac{\partial J_{m,\eta}(U,T,V,X,\lambda,\beta)}{\partial \mu_{ik}} = 0 \quad (11)$$

$$\frac{\partial J_{m,\eta}(U,T,V,X,\lambda,\beta)}{\partial \tau_{ik}} = 0 \quad (12)$$

$$\frac{\partial J_{m,\eta}(U,T,V,X,\lambda,\beta)}{\partial v_i} = 0 \quad (13)$$

Subject to $0 < \sum_{k=1}^{n} \mu_{\mathrm{ik}} < n$; $0 \le \mu_{\mathrm{ik}} \le 1$; $\sum_{i=1}^{c} \mu_{\mathrm{ik}} = 1$; $0 < \sum_{i=1}^{c} \tau_{\mathrm{ik}} < c$; $0 \le \tau_{\mathrm{ik}} \le 1$; $\sum_{k=1}^{n} \tau_{\mathrm{ik}} = 1$; $1 \le k \le n$; $1 \le i \le c$.

We will obtain the value of the membership functions μ_{ik} and τ_{ik}, cluster centroid v_i as following:

$$\mu_{ik} = 1 / \sum_{j=1}^{c} \left(\frac{{SD_{ik}}^2 (1 - \alpha \frac{SI_{ik}}{S})^2 + \|v_i - v_i^*\|^2}{{SD_{ik}}^2 (1 - \alpha \frac{SI_{jk}}{S})^2 + \|v_j - v_i^*\|^2} \right)^{1/(m-1)} \quad (14)$$

$$\tau_{ik} = 1 / \left(1 + \left({SD_{ik}}^2 (1 - \alpha \frac{SI_{ik}}{S})^2 + \|v_i - v_i^*\|^2 \right)^{1/(\eta-1)} \right) \quad (15)$$

$$v_i = \frac{\sum_{k=1}^{n} (\mu_{ik}^m + \tau_{ik}^\eta)(x_k + v_i^*)}{\sum_{k=1}^{n} (\mu_{ik}^m + \tau_{ik}^\eta)(1 + (1 - \alpha \frac{SI_{ik}}{S})^2)} \quad (16)$$

Algorithm 2: The SPFCM-SS algorithm

Input: Multi-spectral image data, number of clusters c and a chosen value for $m, \eta > 1$, stop condition ϵ, mask size S and $\alpha \in [0, S]$.

Output: The fuzzy membership $\mathrm{U} = \{\mu_{ik}\}$, the center of the fuzzy cluster $\mathrm{V} = \{\mathrm{v}_i\}$.

Step 1. Initialization values of $\mathrm{V} = \{\mathrm{v}_i\}$.

Step 2. Calculate the spatial-spectral distance SSD_{ik} by SD_{ik} and SI_{ik} using Eq. 6, 7, 8.

Step 3. Use Eq. 14 to update the fuzzy membership $\mathrm{U} = \{\mu_{ik}\}$.

Step 4. Use Eq. 15 to update the typically membership $\mathrm{T} = \{\tau_{ik}\}$.

Step 5. Use Eq. 16 to update the center of the fuzzy cluster $\mathrm{V} = \{\mathrm{v}_i\}$.

Step 6. If $Max(|J^{(t+1)}_{m,\eta} - J^{(t)}_{m,\eta}|)$ is less than a certain threshold (ϵ), then stop; otherwise, go to step 3.

Next, defuzzification for SPFCM-SS is made as if $u_i(x_k) > u_j(x_k)$ for $j = 1, ..., c$ and $i \neq j$ then x_k is assigned to cluster i.

It is easy to see that, when $\alpha = 0$, the new distance is the distance in the spectral space $SSD_{ik} = SD_{ik} = \|x_i - x_k\|$ and the SPFCM-SS algorithm becomes the semi-supervised possibilistic fuzzy c-means clustering algorithm (SPFCM).

IV. Experiment

PFCM, SFCM, SPFCM and SPFCM-SS algorithms are executed for a maximum of 100 iterations, $\epsilon = 10^{-5}$, mask size $S = 5$, with $m = 2, \eta = 2$ are selected according [5]. The experiment was tried with many different values α and choose the best classification results, $\alpha = 0.5$. Multi-spectral satellite imagery is Landsat and Spot satellite imagery. The satellite data are clustered to 6 classes as follows: Class 1: Rivers, ponds, lakes ; Class 2: Rocks, bare soil ; Class 3: Fields, grass ; Class 4: Planted forests, low woods ; Class 5: Perennial tree crops ; Class6: Jungles . For an image with k bands, each pixel will be characterized by k components on k gray bands. Multi-spectral image data can be described as a set $X = [x_1, x_2, ..., x_n]$, with $x_i = (b_{i1}, b_{i2}, ..., b_{ik})$.

The clustering results have been evaluated by measuring the goodness of the clusters. For this purpose, some validity indexes are used such as Bezdeks partition coefficient (PC-I), the Dunns separation index (D-I), the Separation index (S-I) and Classification Entropy index (CE-I), Xie-Beni (XB-I) index (see in [14], [15]). Note that the validity indexes are proposed to evaluate the quality of clustering. Algorithms producing better results come with smaller values of D-I, XB-I, S-I, CE-I and the larger value of PC-I.

Besides, from the data that has been labeled, the performance of the classification was evaluated with the True Positive Rate (TPR) and False Positive Rate (FPR) which are defined by the following equations:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{17}$$

where TP is the number of correctly classified data and FN is the number of incorrectly misclassified data.

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}} \tag{18}$$

where FP is the number of incorrectly classified data and TN is the number of correctly misclassified data.

A. Experiment 1

The study dataset from Landsat 7 ETM+ multi-spectral images in 30 September 2009 is region center of Hanoi, Vietnam ($105^0 38' 38.8289" E, 21^0 07' 5.3254" N$ and $105^0 58' 53.5268" E, 20^0 58' 14.9711" N$) in Fig. 1 with 6 image bands was obtained by 6 spectral bands including blue, green, red, near infrared, mid-infrared and thermal infrared. Each image band is of size $512x512$, i.e., the size of the data set to be clustered in all the images is 262,144 pixels. The number of pixels are labeled as 18,271 pixels, distributed equally to the land-cover classes.

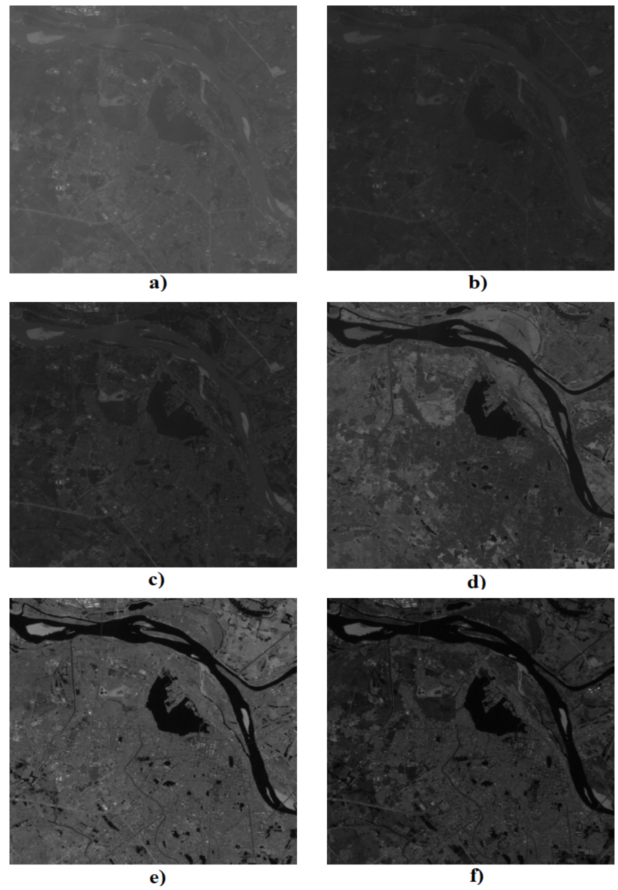

Fig. 1. Hanoi area dataset with 6 spectral bands: a) blue, b) green, c) red, d) near infrared, e) mid-infrared and f) thermal infrared

TABLE I
VALIDY INDICES OBTAINED FOR HANOI AREA

Index / Algorithm	Validy indices				
	XB-I	***PC-I***	***CE-I***	***D-I***	***S-I***
PFCM	0.17523	0.68726	0.56228	0.76258	11.38724
SFCM	0.16782	0.81784	0.38793	0.48723	8.96342
SPFCM	0.15971	0.82878	0.39721	0.38707	6.86293
SPFCM-SS	**0.14874**	**0.85819**	**0.32871**	**0.18943**	**5.89753**

The results are shown in Fig. 2 in which (a), (b), (c) and (d) are the classification results of PFCM, SFCM, SPFCM and SPFCM-SS algorithms, respectively. The results show that SPFCM-SS algorithm noise reduction quite good, while PFCM algorithm is much the most noise.

The results summarized in Tab. I show that SPFCM-SS algorithm produce better quality clustering than those obtained

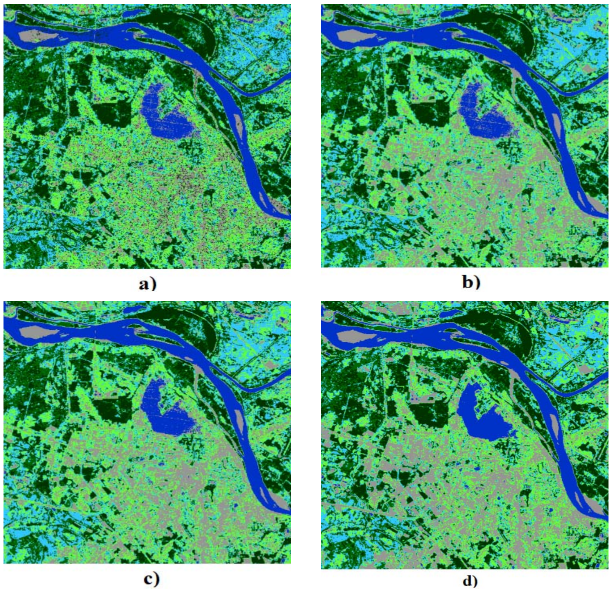

Fig. 2. Hanoi area land-cover classification result: a) PFCM classification; b) SFCM classification; c) SFCM classification; d) SPFCM-SS classification

TABLE II
CLASSIFICATION RESULTS OF PFCM, SFCM, SPFCM AND SPFCM-SS

Methods	PFCM	SFCM	SPFCM	SPFCM-SS
TPR	89.73%	92.84%	95.32%	**97.06%**
FPR	3.81%	3.15%	1.69%	**1.37%**

when running other commonly encountered algorithms such as PFCM, SFCM, SPFCM. Visibly, the indexes obtained from SPFCM-SS are significantly better than those for the PFCM, SFCM, SPFCM.

Table II shows the evaluation results of the different algorithms in the indicators TPR and FPR. The efficient algorithms have larger TPR value and smaller FTR value.

B. Experiment 2

The second experiment is selected in area of Chu Prong district, Gia Lai province (Central highlands of Vietnam). Remote sensing data used in the classification is SPOT multi-spectral image in 2014 in Fig. 3 with 3 image bands was obtained by 3 spectral bands including blue, green and red. The number of pixels are labeled as 15,831 pixels, distributed equally to the land-cover classes.

TABLE III
VALIDY INDICES OBTAINED FOR CHU PRONG AREA

Index	**Validy indices**				
Algorithm	***XB-I***	***PC-I***	***CE-I***	***D-I***	***S-I***
PFCM	0.56385	0.33871	0.37619	0.28936	5.76233
SFCM	0.38723	0.51832	0.29825	0.22843	5.09723
SPFCM	0.32871	0.56591	0.17092	0.19527	4.56188
SPFCM-SS	**0.20672**	**0.56593**	**0.13975**	**0.18723**	**3.87289**

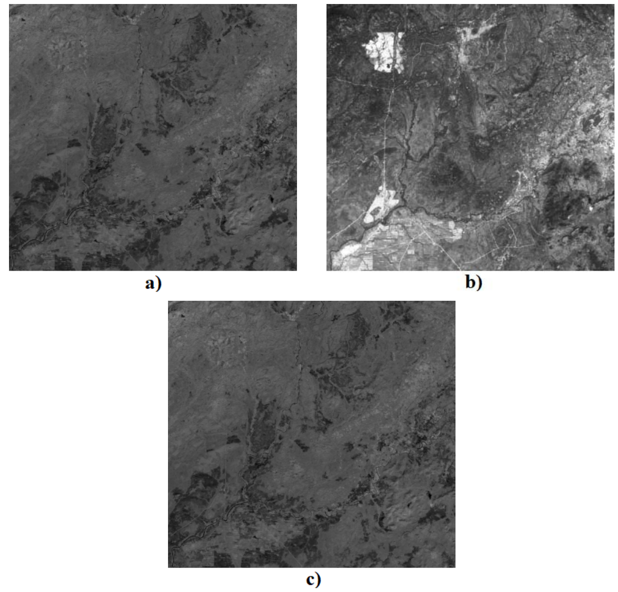

Fig. 3. Chu Prong area dataset with 3 spectral bands: a) blue; b) green 2; c) red

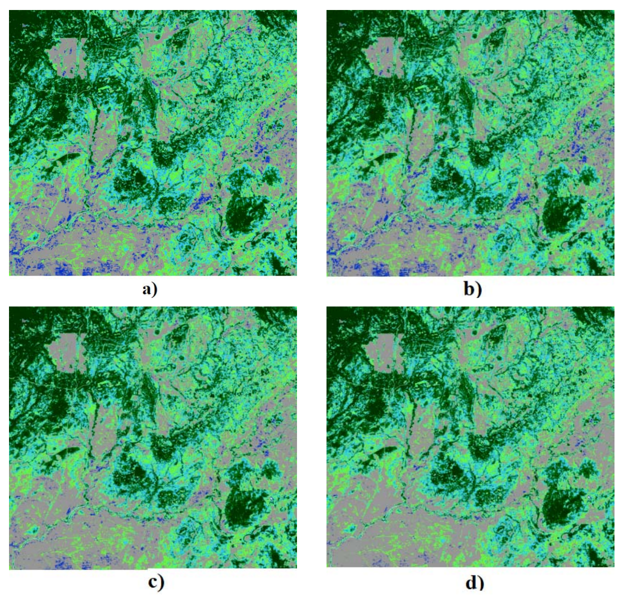

Fig. 4. Chu Prong area land-cover classification result: a)PFCM classification; b) SFCM classification; c) SPFCM classification; d) SPFCM-SS classification

TABLE IV
CLASSIFICATION RESULTS OF PFCM, SFCM, SPFCM AND SPFCM-SS

Methods	PFCM	SFCM	SPFCM	SPFCM-SS
TPR	86.15%	91.27%	94.06%	**96.21%**
FPR	3.67%	2.78%	1.89%	**1.68%**

Fig. 4 shows classificatin results, in which (a), (b), (c) and (d) are the classification results of PFCM, SFCM, SPFCM and SPFCM-SS algorithms, respectively. Tab. III show that

SPFCM-SS algorithm produce better quality clustering than those obtained when running other commonly encountered algorithms such as PFCM, SFCM and SPFCM.

From Tab. II and Tab. IV, the TPR values obtained by running SPFCM-SS on two datasets are greater 96% and obviously higher than the ones obtained from other methods. In addition, the FPR values are also smaller than the ones reached from other algorithms.

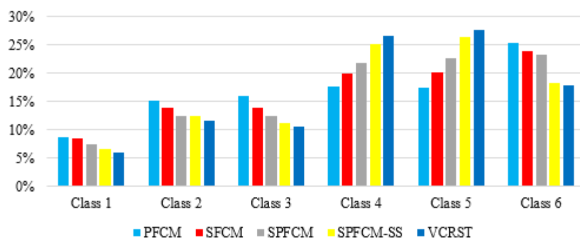

Fig. 5. Study data 1: Comparisons between the result of PFCM, SFCM, SPFCM, SPFCM-SS and the result of VCRST

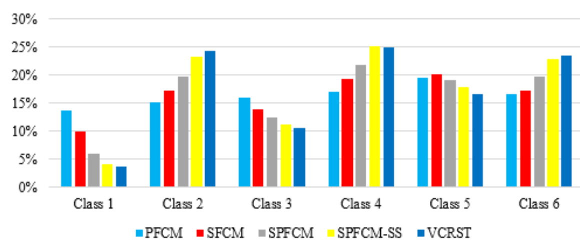

Fig. 6. Study data 2: Comparisons between the result of PFCM, SFCM, SPFCM, SPFCM-SS and the result of VCRST

Fig.5 and Fig.6 compare results between PFCM, SFCM, SPFCM and SPFCM-SS algoritms and data of the Vietnamese Center of Remote Sensing Technology (VCRST) which is considered as the survey data on each class (in percentage %). The significant difference between the algorithms PFCM, SFCM, SPFCM and SPFCM-SS in determining the area of regions. Compare these experimental results with the result of VCRST, with the result of PFCM algorithm, the largest difference is 9%, SFCM algorithm is 8% and SPFCM algorithm is 6%. Meanwhile, the result of SPFCM-SS algorithm does not exceed 5% difference.

Therefore, we can conclude that as forming the spatial-spectral distance from spectral space and spatial information with semi-supervised method for experimental datasets for handling the uncertainties and noises, the quality of the clustering results has been improved.

V. CONCLUSSIONS

This paper presents an advanced possibilistic fuzzy c-means clustering method based on semi-supervised method and spatial-spectral distance (SPFCM-SS), which can reduce the noise and increase the accuracy of clustering results. In addition, the proposed method being endowed with spatial-spectral distance becomes beneficial when it comes to handle the uncertainties. The experiments completed for several multi-spectral image datasets show that the proposed method generates better results than those produced by some other existing clustering methods.

Some next studies may be focused on the use of evolutionary methods (such as particle swarm optimization, genetic algorithms) to optimize parameters of the clustering method.

REFERENCES

[1] Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: the fuzzy C-means clustering algorithm Comput. Geosci. 10 (23), 191 - 203.
[2] R. Krishnapuram and J. Keller, A possibilistic approach to clustering, IEEE Trans. Fuzzy Syst., vol. 1, pp. 98-110, 1993.
[3] Li Liu, Xiao-Jun Wu. Semi-Supervised Possibilistic Fuzzy c-Means Clustering Algorithm on Maximized Central Distance, the 2nd International Conference on Computer Science and Electronics Engineering, pp.1366-1370, 2013.
[4] J.-S Zhang and Y.-W. Leung, Improved Possibilistic C-Means Clustering Algorithms, IEEE Trans. on Fuzzy Systems, vol. 12(2), pp.209-217, 2004
[5] Nikhil R. Pal, Kuhu Pal, James M. Keller, and James C. Bezdek. A Possibilistic Fuzzy c-Means Clustering Algorithm, IEEE Transactions on Fuzzy Systems, Vol. 13, No. 4 (2005), pp.517-530.
[6] Y. Endo, Y. Hamasuna, M. Yamashiro, and S. Miyamoto, "On semisupervised fuzzy c-means clustering", IEEE International Conference on Fuzzy Systems, pp.1119-1124, 2009.
[7] Sinh Dinh Mai and Long Thanh Ngo, Semi-Supervised Fuzzy C-Means Clustering for Change Detection from Multi-spectral Satellite Image, IEEE International Conference on Fuzzy Systems, pp.1-8, 2015.
[8] Sinh Dinh Mai, Long Thanh Ngo, Interval Type-2 Fuzzy C-Means Clustering with Spatial Information for Land-Cover Classification, ACIIDS 2015, Springer LNAI 9011, pp.387-39, 2015.
[9] Sinh Dinh Mai, Long Thanh Ngo and Hung Le Trinh. Satellite Image Classification based Spatial-Spectral Fuzzy Clustering Algorithm. ACIIDS 2018, Springer LNAI 10752, pp. 505 - 518, 2018.
[10] Liu, H., Zhao, F., Jiao, L., 2012. Fuzzy spectral clustering with robust spatial information for image segmentation. Appl. Soft Comput. 12, 3636 - 3647.
[11] Vargas, D.M., Funes, F.J.G., Silva, A.J.R., 2013. A fuzzy clustering algorithm with spatial robust estimation constraint for noisy color image segmentation. Pattern Recognit. Lett. 34, 400 - 413.
[12] Zhao, F., Jiao, L., Liu, H., 2013. Kernel generalized fuzzy C-means clustering with spatial information for image segmentation. Digit. Signal Process. 23, 184 - 199.
[13] Liu, J., Pham, T.D., 2012. A spatially constrained fuzzy hyper-prototype clustering algorithm. Pattern Recognit. 45, 1759 - 1771.
[14] W. Wang,Y. Zhang (2007). On fuzzy cluster validity indices, Fuzzy Sets and Systems, Vol. 158, 2095-2117, 2007.
[15] Ujjwal Maulik and Sanghamitra Bandyopadhyay, Performance Evaluation of Some Clustering Algorithms and Validity Indices, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 12, 2002, pp.1650-1654.