



SPATIAL-SPECTRAL FUZZY K-MEANS CLUSTERING FOR REMOTE SENSING IMAGE SEGMENTATION

Mai Dinh Sinh^{1,*}, Ngo Thanh Long², Trinh Le Hung¹

¹*Institute of Technique and Special Engineering, Le Quy Don Technical University,
236 Hoang Quoc Viet, Ha Noi, Viet Nam*

²*Institute of Simulation Technology, Le Quy Don Technical University,
236 Hoang Quoc Viet, Ha Noi, Viet Nam*

*Email: maidinhsinh@gmail.com

Received: 3 October 2017; Accepted for publication: 09 March 2018

Abstract. Spectral clustering is a clustering method based on algebraic graph theory. The clustering effect by using spectral method depends heavily on the description of similarity between instances of the datasets. Although, spectral clustering has been of significant interest in recent times, the raw spectral clustering is often based on Euclidean distance, and it is impossible to accurately reflect the complexity of the data. Despite having a well-defined mathematical framework, good performance and simplicity, it suffers from several drawbacks, such as it is unable to determine a reasonable cluster number and is sensitive to complex shape objects and noise elements. In this paper, we present a new approach named spatial-spectral fuzzy clustering which combines spectral clustering and fuzzy clustering with spatial information of the data samples with their neighbors into a unified framework to solve these problems. The study consists of three main steps: Step 1, calculate the spatial information value of the pixels, step 2 applies the spectral clustering algorithm to change the data space from the color space to the new space, and step 3 clusters the data in new data space by fuzzy clustering algorithm. Experimental results on the remote sensing image were evaluated based on a number of indicators, such as IQI, MSE, DI and CSI, showing that it can improve the clustering accuracy and avoid falling into local optimum.

Keywords: spectral clustering, spatial information, remote sensing image, fuzzy clustering.

Subject Classification: 4.7.4; 4.8.2; 4.8.3

1. INTRODUCTION

Spectral clustering is a technique for finding group structure in data. This method was first introduced by Donath and Hoffman in [1] who suggested to use eigenvectors and eigenvalues of a graph-related matrix called the graph Laplacian to solve the problem of detecting clusters in an undirected graph. It is based on viewing the data points as nodes of a connected graph and clusters are found by partitioning this graph [2], based on its spectral decomposition, into subgraphs that possess some desirable properties [3]. Spectral clustering applies in various fields

such as medical image classification [4], satellite image [5, 6], bioinformatics [7]. The widely used similarity measure for spectral clustering is Gaussian kernel function which measures the similarity between data points. However, it is difficult for spectral clustering to choose a suitable scaling parameter in Gaussian kernel similarity measure [8].

So much research has improved to overcome the drawbacks of the spectral clustering method, the research focus based on distribution characteristics, the structure of the cluster or data density. Some research as, Yue Li et al. [9] constructed a new matrix S based on density as the similarity matrix, and proposed k-Means based on density to converge the global optimization. Making it to find the spatial distribution characteristics of complex data. As we know it, k-means clustering algorithm is hard, with overlapping data, the application of k-Means technique do not often result in high precision. Spectral clustering has shown to be a powerful technique for grouping and/or ranking data as well as a proper alternative for unlabeled problems, so, D.H. Peluffo-Ordóñez et al. [10] has an overview of clustering techniques based on spectral clustering problem of dynamic data analysis. Chang-an Liu and his team [11] improved spectral clustering algorithm (ISC) by applying the hill-climbing techniques to find the cluster centroids and reduce computing costs by merging the pixels have the same gray level. However, merging the pixels have the same gray level will lose the relationship between the spatial pixels, which affect the clustering results.

To increase the accuracy of image clustering, a new Hierarchical Iterative Clustering Algorithm using Spatial and Spectral information was introduced by Sayyed Bagher Fatemi et al. [5] This algorithm separates pixels into uncertain and certain categories based on decision distances in the feature space. The algorithm labels the certain pixels using the k-means clustering, and the uncertain ones with the help of information in both spatial and spectral domains of the image. The difficulty of this algorithm is knowing exactly the certainty of a pixel belonging to a certain class.

The classic spectral clustering algorithm has a superior performance in the category in any shape of data collection, but the computational complexity of the classic spectral clustering algorithm is very high. In the case of limited computer memory and computing speed, only can process low-dimensional image. To solve these problems, Hua Bo et al. [6] Improved Spectral Clustering Algorithm adopts a grid method that the original image was divided into several sub-images first and then processed separately. This is to reduce the computational complexity, but disadvantages of this approach is that on each sub-image, segmentation results easy to fall into local convergence. Feng Zhao et al. [12] developed a fuzzy similarity measure for spectral clustering to replace the traditional measurements of spectral clustering algorithm used is based on Gaussian kernel function. Peng Yang et al. [13] defined the density sensitive similarity measure which can adjust the distance in regions with different density to replace conventional Euclidean distance based and Gaussian kernel function based spectral clustering.

Moreover, H.Q. Liu et al. [14] proposed a method, novel non-local spatial spectral clustering algorithm for image segmentation. In the presented method, the objective function of weighted kernel k-means algorithm is firstly modified by incorporating the non-local spatial constraint term. Then the equivalence between the objective functions of normalized cut and weighted kernel k-means with non-local spatial constraints is given and a novel non-local spatial matrix is constructed to replace the normalized Laplacian matrix. Finally, spectral clustering techniques are applied to this matrix to obtain the final segmentation result. Jun Yan and et al. [15] is inspired by density sensitive similarity measure. Making it increase the distance of the pairs of data in the high density areas, which are located in different spaces. And this can reduce

the similarity degree among the pairs of data in the same density region, so as to find the spatial distribution characteristics complex data.

Besides, spectral clustering algorithms are sensitive to noise and other imaging artifacts because of not taking into account the spatial information of the pixels in the image. Owing to the limitations of the feature space in multispectral images and spectral overlap of the clusters, it is required to use some additional information such as the spatial information of pixel in image. So the need for improvements in order to overcome the drawbacks mentioned above with spectral clustering method. We present a new approach named spatial-spectral clustering which combines spectral clustering with spatial information into a unified framework to solve these problems, this paper constructed a new matrix S based on spatial information as the similarity matrix, and proposed using fuzzy k-Means algorithm to converge the global optimization (SSFKM).

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Spectral clustering algorithm

Spectral clustering techniques [1, 2, 3] make use of the spectrum (eigenvalues) of the similarity matrix of the data to perform dimensionality reduction before clustering in fewer dimensions. The similarity matrix is provided as an input and consists of a quantitative assessment of the relative similarity of each pair of points in the dataset.

Let $X = \{x_1, x_2, \dots, x_n\}$ be the set of n points to be clustered, and S be the $n \times n$ similarity matrix with its elements, s_{ij} , showing pairwise similarities between n points. Let $G = (V, S, X)$ be a weighted, undirected graph with V representing n nodes ($x_i \in X$ to be clustered), and S defining the edges. When constructing similarity graphs the goal is to model the local neighborhood relationships between the data points. There are other several popular constructions to transform the data points with pairwise similarities s_{ij} into a graph. S is usually constructed as a Gaussian function based on (often Euclidean) distances, $d(x_i, x_j)$, between samples x_i, x_j :

$$s_{ij} = \exp\left(-\frac{d^2(x_i, x_j)}{\sigma^2}\right) \quad (1)$$

with a global parameter σ determining the decay of the similarity. This definition requires either a user-set σ value or a selection among many σ values to find the optimal value.

D is a diagonal matrix and its elements are the degrees of the nodes of G . The degree of each node, d_i , is computed with:

$$d_i = \sum_j s(i, j) \quad (2)$$

The Laplacian matrix L , is constructed using the similarity matrix S and degree matrix D , depending on the approach for graph-cut optimization. Ng et al. [8] define a normalized Laplacian matrix, L_{norm} , as:

$$L_{norm} = D^{-1/2} S D^{-1/2} \quad (3)$$

then L_{norm} is used for extraction of c clustering by finding its c eigenvectors with the c highest eigenvalues. The spectral clustering algorithm can be summarized as follows:

1. Calculate a similarity matrix S (1), diagonal degree matrix D (2), and L_{norm} (3).
2. Find the c eigenvectors $\{e_1, e_2, \dots, e_c\}$ of L_{norm} , associated with the c highest eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_c\}$.
3. Construct the $n \times c$ matrix $E = [e_1, e_2, \dots, e_c]$ and obtain $n \times c$ matrix U by normalizing

the rows of E to have norm 1, i.e. $u_{ij} = \frac{e_{ij}}{\sqrt{\sum_c e_{ic}^2}}$.

4. Cluster the n rows of U with the clustering algorithm into c clusters.

2.1.2. Fuzzy k-Means clustering algorithm

The fuzzy k-means procedure by Bezdek [16] allows each data pattern to have a degree of membership in cluster i . In general, degree of memberships in fuzzy k-means achieved by computing the relative distance among the patterns and cluster centroids. Initialize the initial centroids $v_i, i = \overline{1, c}$ and degree of membership u_{ik} is determined as follow:

$$u_{ik} = \frac{1}{d_{ik}} / \sum_{j=1}^c \left(\frac{1}{d_{jk}} \right) \quad (4)$$

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.

Cluster centroids is computed as follows with the fuzzy mean of all of the examples for cluster i :

$$v_i = \sum_{k=1}^N u_{ik} x_k / \sum_{k=1}^N u_{ik} \quad (5)$$

In which $i = 1, \dots, C$; $k = 1, \dots, N$. Repeat the formulas (4) and (5) until there is no significant change about centroid of clusters.

Next, defuzzification for FKM 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 .

2.2. Analytical methods

2.2.1. Spatial information

In image segmentation, the key to determine a pixel belonging to certain area is based on the similarity of these colors, which is calculated through a distance function in the color space $d_{ij} = \|x_i - x_j\|$ e.g. Euclidean distance between the pattern x_i and x_j . However in fact, 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. The main idea of the method is to add information about the location of the pixels into the clustering process, with the pixels have colors close to the color of the center pixel, in terms of location, the pixel is closer to the center, it more likely belongs to the same cluster with central pixels than other pixels.

To calculate the degree of influence of the neighboring pixels to the center pixel, we use a mask of size $n \times n$ to position on the image, the center pixel of the mask is the considered pixel. The number of neighboring pixels P is determined corresponding to the selected type of mask size i.e. 8 pixels for mask 3×3 , 24 pixels for mask 5×5 , 48 pixels for mask 7×7 , so on.

To determine the degree of influence of the neighboring pixels for the center pixels, a spatial information measure M_i is defined on the basis of the distance $\|x_i - x_j\|$ and the attraction distance r_{ij} :

$$M_i = \frac{\sum_{j=1}^P (\|x_i - x_j\| r_{ij})^{-1}}{\sum_{j=1}^P r_{ij}^{-1}} \quad (6)$$

in which $\|x_i - x_j\|$ is the distance of the all neighboring element x_j on the mask to the cluster x_i . The distance attraction r_{ij} is the squared Euclidean distance between elements (x_i, y_i) and (x_j, y_j) about position on the mask. According to the above expression, spatial information of each pixel comes with a higher value if its color is similar to the color of neighboring pixels and vice versa. We use the inverse distance r_{ij}^{-1} because the closer the neighbors x_j of the center x_i are the more influence they exert on the result and vice versa.

The idea behind the use of this spatial relationship information can be outlined as follows: consider the local $n \times n$ mask, for sliding the mask on the image and calculating the spatial information of the center pixel x_i based on the location of the center pixel x_i with the pixels x_j in the mask and the distance in color space $\|x_i - x_j\|$. This aims to reduce the effect of noise on the image. From above description, this method of similarity measure fully considers the spatial information and can avoid the influence of the image noise.

Set $r = \max(r_{ij})_{\forall i, j}$ is the radius of the largest circle in which pixels that affect the central pixel. Next, without loss of generality we standardized similar measurements on the following formula:

$$\overline{M}_i = \frac{M_i - \min(M_i)_{\forall i}}{\max(M_i)_{\forall i} - \min(M_i)_{\forall i}} \quad (7)$$

2.2.2. Spatial-spectral fuzzy clustering algorithm

From the above description, a new similarity measure is defined as follows:

$$s_{ij} = \exp\left(-\frac{d^2(x_i, x_j)}{\alpha r^2}\right) \quad (8)$$

where s_{ij} showing pairwise similarities between pixels x_i, x_j ; $d(x_i, x_j) = \|x_i - x_j\|$ is the Euclidean distance between x_i and x_j ; r is the radius of the largest circle in which pixels that affect the central pixel; α is a constant belonging to $[0,1]$. Similarity Matrix S is usually constructed by s_{ij} according to the formula (8).

With degree matrix D , we build by adding spatial-spatial information of each pixel, the degree of each pixel, d_i , is computed with:

$$d_i = \overline{M}_i * \sum_j s(i, j) \quad (9)$$

From the above description, the new Laplacian matrix L_{new} , is constructed using the new similarity matrix S and new degree matrix D :

$$L_{new} = D^{-1/2} S D^{-1/2} \quad (10)$$

Find the c eigenvectors $\{e_1, e_2, \dots, e_c\}$ of L_{new} , associated with the c highest eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_c\}$ and construct the $n \times c$ matrix $E = [e_1, e_2, \dots, e_c]$ and obtain $n \times c$ matrix U by normalizing the rows of E to have norm 1, i.e. $u_{ij} = \frac{e_{ij}}{\sqrt{\sum_c e_{ic}^2}}$. Finally, cluster the n rows of U

with the fuzzy k-means clustering algorithm into c clusters.

The main steps of the proposed method SSFKM are given as follows:

Input: 1. Remote sensing image data.

2. Matrix size used to calculate spatial information, number of clusters c .

Step 1. Calculate spatial information measure M_i by (6), (7).

Step 2. Applies the spectral clustering algorithm to change the data space

2.1 Calculate a new similarity matrix S by (8).

2.2 Calculate a diagonal degree matrix D by (9).

2.3 Calculate a new matrix L_{new} by (10).

2.4 Find the c eigenvectors $\{e_1, e_2, \dots, e_c\}$ of L_{new} , associated with the c highest eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_c\}$ and define the c dimensional space $Y = (y_i)_{i=1, \dots, n} \in R^c$.

Step 3. Clusters the data in new data space

3.1 Calculates the function value u_{ij} by (4).

3.2 Update centroids $v_i, i = 1, \dots, c$ by (5).

3.3 Checks the stop condition $\max\{\|v^{(t+1)} - v^{(t)}\|\} \leq \varepsilon$, If satisfied, go to output, otherwise return to step 3.1.

Output: C_1, C_2, \dots, C_c with $C_i = \{x_j | u_{ij} \in c_i\}$.

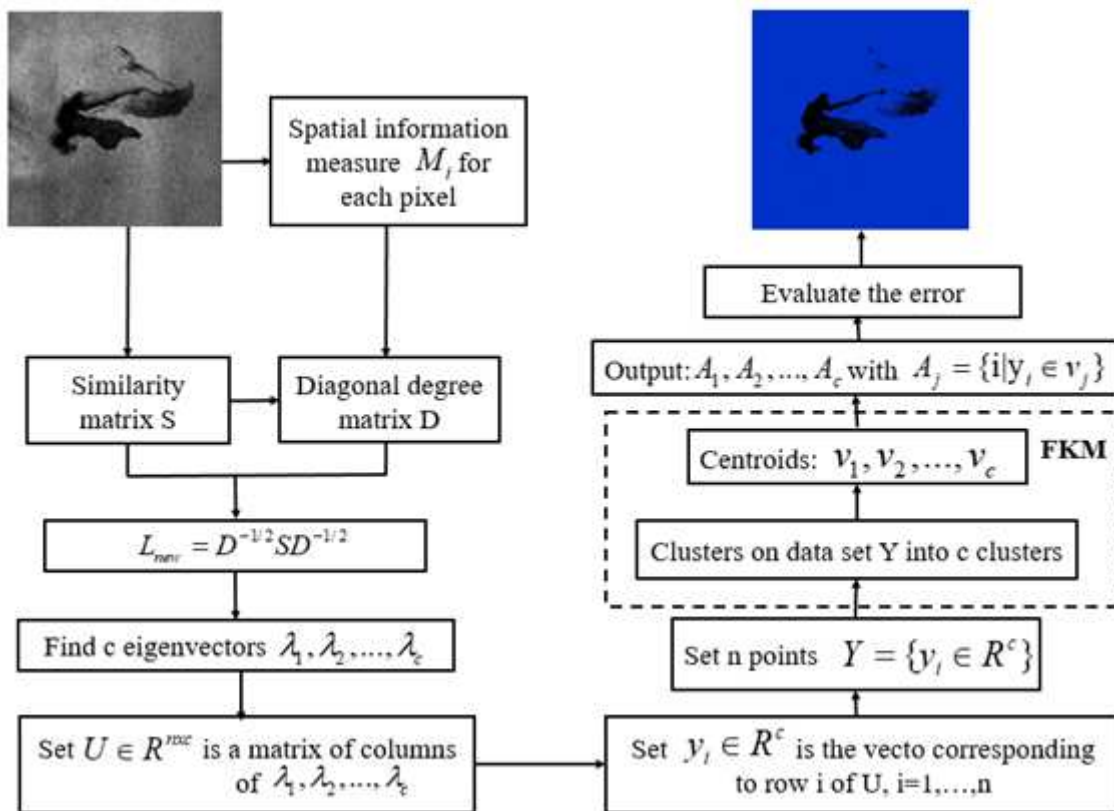


Figure 1. Diagram the calculation steps of the SSFKM algorithm.

Figure 1 shows a detailed diagram of the calculation steps of SSFKM algorithm. With this approach, the new algorithm will overcome the difficulties in the selection of parameters σ and pepper salt noise reduction in the image. This can increase the accuracy of the image clustering with spectral clustering algorithm.

2.2.3. Spatial-spectral fuzzy clustering algorithm

To assess the quality of image segmentation, we use some indexes:

- Mean Squared Error (MSE) index [17]:

$$MSE(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (11)$$

where, $X = \{x_i\} = \{x_1, x_2, \dots, x_N\}$ and $Y = \{y_i\} = \{y_1, y_2, \dots, y_N\}$ corresponding to the original image and segment results image. Smaller MSE value is better quality clusters.

- Image Quality Index (IQI) [18]:

$$IQI = \frac{4\sigma_{xy} \bar{x} \bar{y}}{(\sigma_x^2 + \sigma_y^2)(\bar{x}^2 + \bar{y}^2)} \quad (12)$$

with $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$, $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$, $\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$, $\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2$ and

$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$. Where, $X = \{x_i\} = \{x_1, x_2, \dots, x_N\}$ and $Y = \{y_i\} = \{y_1, y_2, \dots, y_N\}$

corresponding to the original image and segment results image. The best value 1 is achieved if and only if $y_i = x_i$, the lowest value of -1 occurs when $y_i = 2\bar{x} - x_i$ with $i = 1, N$.

- The Dunn's index (DI) [19] is defined as

$$DI = \min_{i \in c} \left\{ \min_{j \in c, j \neq i} \left\{ \frac{\delta(A_i, A_j)}{\max_{k \in c} \{\Delta(A_k)\}} \right\} \right\} \quad (13)$$

in which,

$$\delta(A_i, A_j) = \min \{d(x_i, x_j) \mid x_i \in A_i; x_j \in A_j\}$$

$$\Delta(A_k) = \max \{d(x_i, x_j) \mid x_i, x_j \in A_k\}$$

and d is a distance function and A_i is the set whose elements are the data points assigned to the i^{th} cluster. The main drawback with direct implementation of Dunn's index is its computation since calculation becomes computationally very expensive as c and n increase. If a data set contains well-separated clusters, the distances among the clusters are usually large and the diameters of the clusters are expected to be small. Therefore, large values of Dunn's index correspond to good clustering solution.

- The CS measure is proposed to evaluate clusters with different densities and/or sizes [20]. It is computed as:

$$CSI = \frac{\frac{1}{c} \sum_{i=1}^c \left\{ \frac{1}{|A_i|} \sum_{x_j, x_k \in A_i} \max \{d(x_j, x_k)\} \right\}}{\frac{1}{c} \sum_{i=1}^c \left\{ \min_{j \in c, j \neq i} \{d(v_i, v_j)\} \right\}} \quad (14)$$

where $|A_i|$ is the number of elements in cluster A_i and $d(x_j, x_k)$ a distance function. The smallest CS measure indicates a valid optimal clustering.

3. RESULTS AND DISCUSSION

3.1. SAR image segmentation

Synthetic Aperture Radar (SAR) is used to obtain high-resolution images from broad areas of terrain. SAR is capable of operating under inclement weather conditions, day or night. SAR images have wide applications in remote sensing and mapping of the surfaces of both the Earth and other planets. There are many other applications for this technology such as environmental monitoring, earth-resource mapping, land-cover classification, oil spill classification, so on. However, SAR image segmentation is a difficult task in remote sensing applications due to the influence of the speckle noise (see Figure 2 and Figure 4), therefore, using conventional methods will not be efficient with speckle noise. Using combinations of techniques with spatial information can help increase the accuracy of classification results (The test data is taken from <https://earth.esa.int/web/guest/data-access/sample-data>).

3.1.1. Experiment 1

To test the SSFKM algorithm that we proposed, SAR image segmentation with oil spill on the sea. The Envisat ASAR image data were taken in Matera coastal areas, Italy on 17 November 2002 (Figure 2). The oil slick released by the tanker is visible as a black streak across the images. The tanker itself is visible as a white dot at the bottom-left of the main slick (bottom-left of the image). To perform SAR image segmentation into 2 layers: seawater layer and oil spill layer. To compare the proposed algorithm with the methods that have been studied before, we empirically tested four methods: FKM, SC, ISC and SSFKM is proposed in this paper.

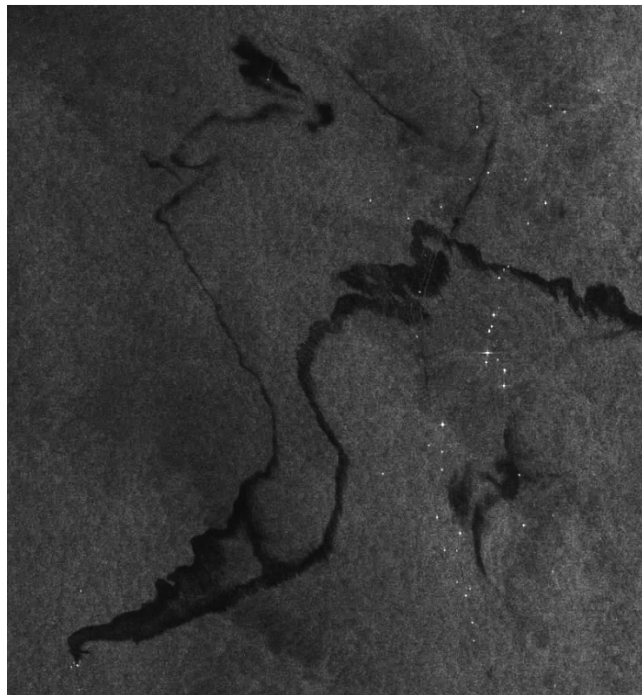


Figure 2. Spill oil area on Envisat ASAR image in Matera coastal areas, Italy on 17/11/2002.

The output images is shown in Figure 3 below. Easy to see in Figure 3a, oil stain classification result is unclear, there are many noises appear in the result image. In Figure 3b and Figure 3c noise significantly reduced, particularly on the amount of noise in Figure 3d has fallen strongly and oil stains is clearer.

To assess the performance of the algorithms on the experimental images, we analyzed the results on the basis of several validity indexes. We considered several validity indexes such as the MSE, IQI, DI and CSI.

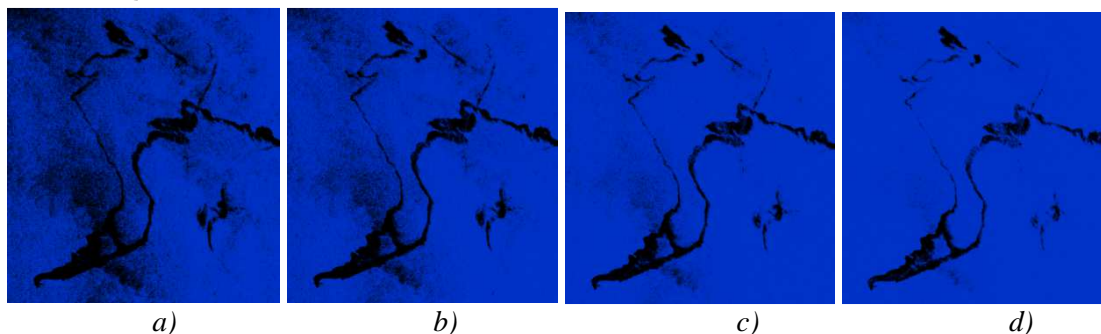


Figure 3. Results of oil stain classification on the Envisat ASAR image in Matera coastal areas, Italy on 17/11/2002.

Table 1. Performance of the FKM, SC, ISC and the SSFKM algorithms.

Index	FKM	SC	ISC	SSFKM
MSE	0.1986	0.1452	0.1075	0.0918
IQI	0.4982	0.6124	0.6732	0.8124
DI	0.0198	0.0109	0.0365	0.0425
CSI	1.3017	0.9683	0.7750	0.7751

Table 1 shows that IQI index and DI index are the largest with 0.8124 and 0.0425 on SSFKM algorithm, while ISC algorithm is 0.6732 and 0.0365; These indicators decrease with 0.6124 and 0.4982; 0.0109 and 0.0198, respectively on algorithms SC and FKM. With MSE index, the largest is 0.1986 when tested on FKM algorithm, and descending on algorithms SC, ISC and SSFKM, while, the lowest value is 0.0918 on SSFKM algorithm. The smallest value of CSI index is 0.7750 with ISC algorithm, while, SSFKM algorithm is 0.7751 and we easily see that this deviation is insignificant. While CSI index with SC algorithm is 0.9683 and FKM algorithm is 1.3017. Thus, based on the value of the index clustering quality evaluation, the most of the cases showed SSFKM algorithm for clustering results better than the algorithm ISC, SC and FKM.

3.1.2. Experiment 2

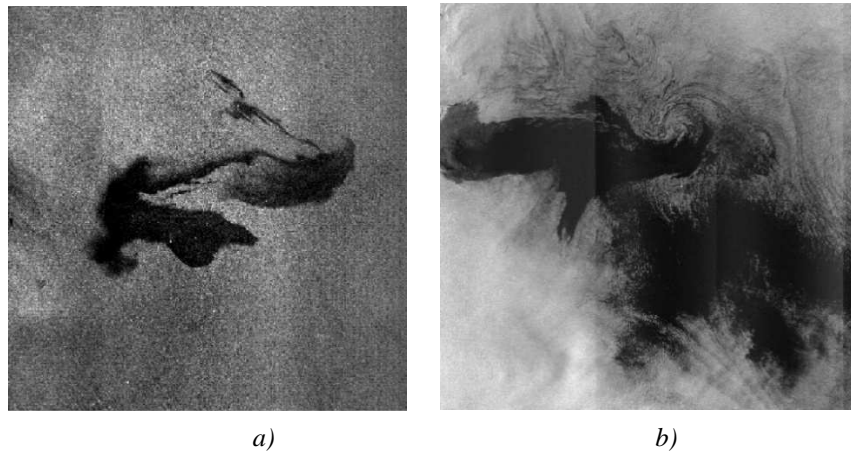


Figure 4. Spill oil area on Envisat ASAR image in Gulf of Mexico (a) 26/04/2010, (b) 29/04/2010.

Test data that Asar envisat images was taken from a spill oil area in Gulf of Mexico on 26/04/2010 (4a) and 29/04/2010 (4b), with coordinates ($0^{\circ}14'02.75''N$, $0^{\circ}03'56.39''E$ to $0^{\circ}04'27.33''N$, $0^{\circ}22'13.94''E$) with area of 23.32 hectare. Easily recognized oil stains on Figure 4a with clearer boundaries, whereas in Figure 4b oil stains have long existed in the sea, so the contrast with the surrounding waters as well as the boundaries of oil stains is not clarity, many parts mixed with water and oil stains area has spread.

Classification results shown in Figure 5 in Gulf of Mexico on 26/04/2010, the FKM, SC, ISC and the SSFKM algorithm with Figure 5a, Figure 5b, Figure 5c and Figure 5d, respectively. In Figure 5, still quite a lot of noise on the Figure 5a, 5b and 5c, especially, in Figure 5a. Classification results in Figure 5d shows the noise almost nonexistent on water layer spill area is also clear than other results.

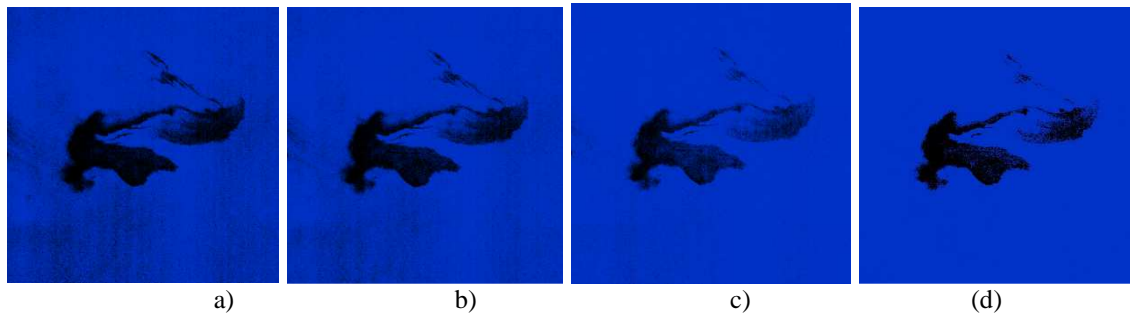


Figure 5. Results of oil stain classification on the Envisat ASAR image in Gulf of Mexico on 26/4/2010.

Figure 6 shows classification results in Gulf of Mexico on 29/04/2010, the FKM, SC, ISC and the SSFKM algorithm with Figure 6a, Figure 6b, Figure 6c and Figure 6d, respectively. Easy to see, the noise is reduced quite good on all the results in Figure 6 because SAR image data has quite little pepper salt noise (see Figure 4b), however a large amount of information about the oil stains on the Figure 6a, 6b and 6c is confused with noise, therefore undetectable. Test result with our algorithm on Figure 6d shows, not only is the noise reduction possible but that oil stains classification result is also complete and clearer.

Table 2. Performance of the FKM, SC, ISC and the SSFKM algorithms on 26/4/2010.

Index	FKM	SC	ISC	SSFKM
MSE	0.2081	0.1642	0.1212	0.1063
IQI	0.4267	0.5638	0.7851	0.8542
DI	0.0169	0.0317	0.0561	0.0618
CSI	1.3516	0.9981	0.8725	0.6068

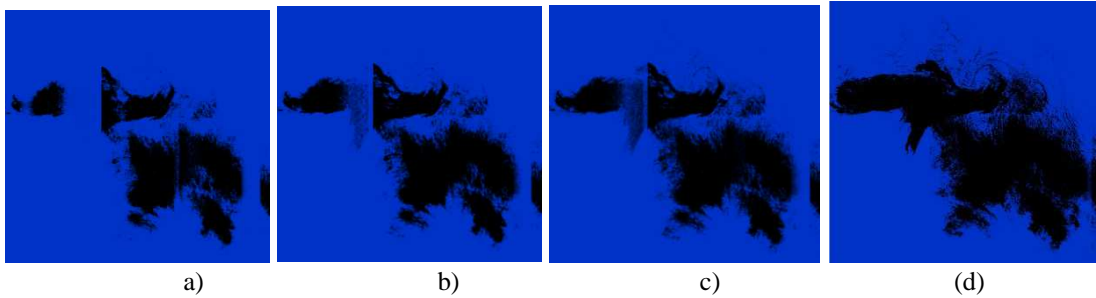


Figure 6. Results of oil stain classification on the Envisat ASAR image in Gulf of Mexico on 29/4/2010.

Table 3. Performance of the FKM, SC, ISC and the SSFKM algorithms on 29/4/2010.

Index	FKM	SC	ISC	SSFKM
MSE	0.1916	0.1578	0.1082	0.0098
IQI	0.4372	0.5892	0.6823	0.9071
DI	0.0318	0.0461	0.0598	0.0783
CSI	1.877	1.2319	0.8873	0.7652

Tables 2 and 3 indicate the value of the index assessing the quality of clustering results. Overall, the results classified according algorithm that we propose for better results than the algorithm FKM, SC and ISC. Based on the value of this index, the FKM algorithm for clustering result is the worst, followed by SC and ISC algorithm.

3.2. Land-cover classification



Figure 7. Remote sensing image in Hanoi center on 30/9/2009.

The test data is Landsat-7 ETM+ remote sensing image taken in Hanoi central area on 30/9/2009 (<https://earthexplorer.usgs.gov/>), coordinates from (1050 38' 38.8289"E, 210 07' 5.3254"N) to (1050 58' 53.5268"E, 200 58' 14.9711"N) with 564.13 sq km (see Figure 7). These Landsat-7 image will be classified into six classes representing six types of land covers: ■ Class 1: Rivers, ponds, lakes. ■ Class 2: Rocks, bare soil. ■ Class 3: Fields, grass. ■ Class 4: Planted forests, low woods. ■ Class 5: Perennial tree cops. ■ Class 6: Jungles.

Table 4. Results of land cover classification in Hanoi area.

Class	FKM	SC	ISC	SSFKM
1	8.581 %	8.034 %	7.126 %	6.920 %
2	14.164 %	15.601 %	18.629 %	19.345 %
3	12.959 %	15.369 %	19.404 %	21.103 %
4	25.156 %	21.882 %	15.828 %	14.598 %
5	25.148 %	22.680 %	19.854 %	18.564 %
6	13.992 %	16.434 %	19.158 %	19.469 %

The results of land cover classification were shown in Figure 8, in which figure 8a, 8b, 8c and 8d are classification results of FKM, SC, ISC and SSFKM proposed algorithm, respectively. Table 4 showed the comparison of classification results obtained by using FKM, SC, ISC and SSFKM algorithm. As can be seen, there was a significant difference on the area of regions classified by the aforementioned algorithms.

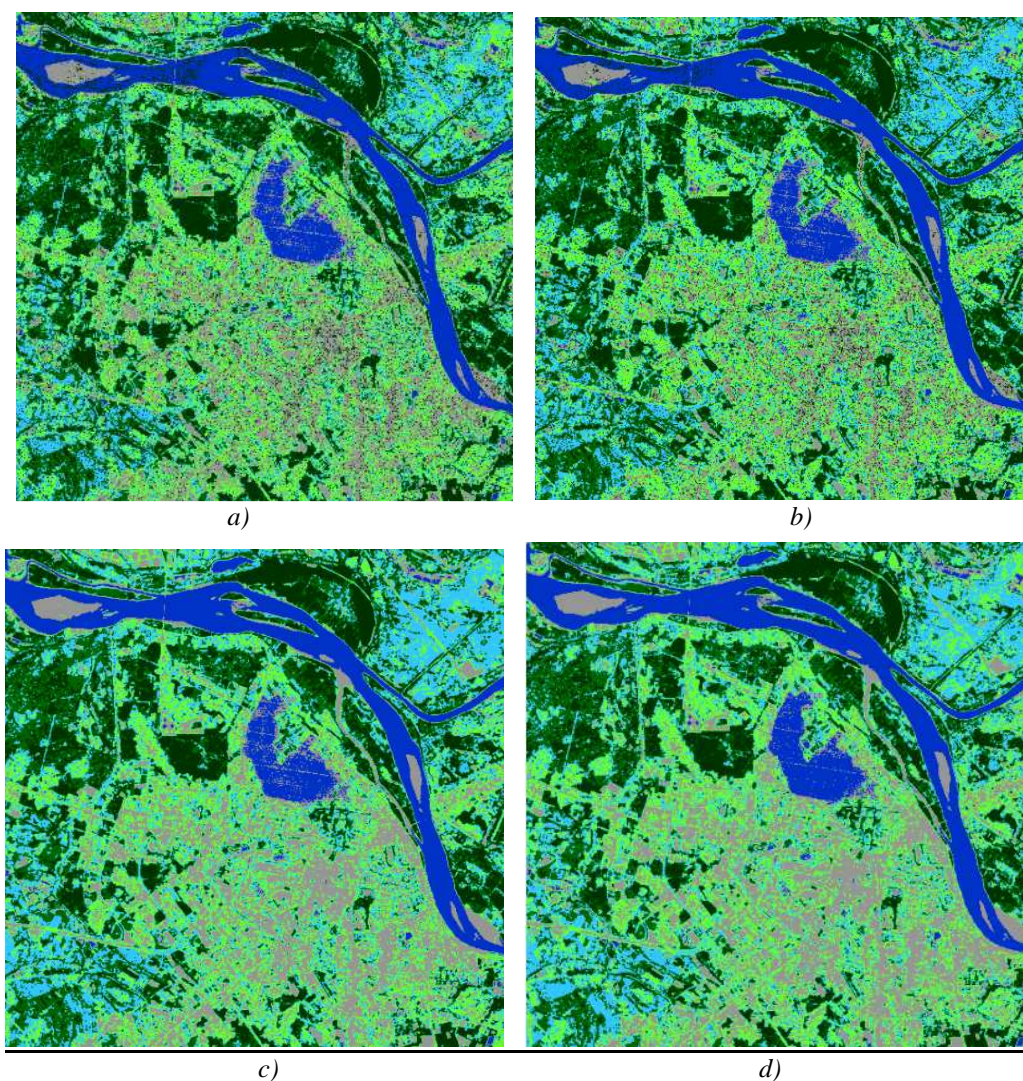


Figure 8. Result land-cover classification in Hanoi area, FKM (a), SC (b), ISC (c) and the SSFKM (d).

Table 5. Performance of the FKM, SC, ISC and the SSFKM algorithms.

Index	FKM	SC	ISC	SSFKM
MSE	0.3872	0.2091	0.1762	0.1483
IQI	0.2837	0.6523	0.7323	0.7879
DI	0.0476	0.0568	0.1086	0.1085
CSI	1.4729	1.1734	0.7678	0.7139

In this study, to evaluate the quality of clusters, we considered the different validity indices, such as MSE, IQI, DI and CSI. It can be seen that the accuracy of land cover classification using FKM algorithm and SC algorithm was very low. Many objects, such as bare soil and water, bare soil and sparse vegetation were misclassified. The accuracy of land cover classification was improved when using ISC algorithm, however, it was not so high. I can be seen that the SSFKM

provided better classification result than the other algorithms, such as FKM, SC and ISC in MSE, IQI, CSI. With DI index, classification result of SSFKM is better than FKM and SC. The results of calculation of IQI, MSE, DI and CSI indices by 4 algorithms FKM, SC, ISC and SSFKM were shown in Table 5.

4. CONCLUSIONS

In this study, we have presented a spatial-spectral fuzzy k-mean clustering algorithm (SSFKM). The first one is based on spatial relationships between the pixels and their neighbors in circle to similarity matrix. The second one uses the set of spatial information values to degree matrix. The advantages of the proposed algorithms are pointed out in reducing noise on image. Test results show that proposed algorithm has high segmentation accuracy and significantly reduces the computational complexity of classical spectral clustering algorithm and through experimental results, according to the visual and validity indexes MSE, IQI, DI and CSI, basically SSFKM for sharper image quality, better noise reduction.

There are several further research directions including the use and refinement of the proposed clustering to processing on other satellite image types and application to land-cover change detection, environmental classification and assessment of land surface temperature changes. The issues of speed-ups of the proposed methods based on GPU platforms form another important research direction.

Acknowledgements. This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.05-2016.09.

REFERENCES

1. Donath W. E., Hoffman A. J. - Lower Bounds for the Partitioning of Graphs, IBM Journal of Research and Development **17**(5) (1973) 420-425.
2. Shi J. and Malik J. - Normalized cuts and image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(8) (2000) 888–905.
3. Fowlkes C., Belongie S., Chung F., and Malik J. - Spectral grouping using the Nystrom method, IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2) (2004) 214–225.
4. Chia-Tung Kuo, Peter B. Walker, Owen Carmichael, Ian Davidson - Spectral Clustering for Medical Imaging, 2014 IEEE International Conference on Data Mining (2014) 887-892.
5. Sayyed Bagher Fatemi, Mohammad Reza Mobasheri, and Ali Akbar Abkar - Clustering Multispectral Images Using Spatial–Spectral Information, IEEE Geoscience and Remote Sensing Letters **12** (7) (2015) 1521 – 1525.
6. Hua Bo, Jun Zhang and Xiaofeng Wang - Improving Spectral Clustering Algorithm based SAR Spill Oil Image Segmentation, 2011 International Conference on Network Computing and Information Security, IEEE (2011).
7. Desmond J. Higham, Gabriela Kalna, Milla Kibble - Spectral clustering and its use in bioinformatics, Journal of Computational and Applied Mathematics **204** (2007) 25-37.

8. Ng A., Jordan M., and Weiss Y. - On spectral clustering: analysis and an algorithm” in T. Dietterich, S. Becker, and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems* **14**, MIT Press, (2002).
9. Yue Li, Xiyu Liu, and Xuebin Yan - A Modified Spectral Clustering Algorithm Based on Density, *HCC 2016, LNCS 9567* (2016) 901–906.
10. Peluffo-Ordóñez D.H., Alvarado-Pérez J.C., and Castro-Ospina A.E. - On the Spectral Clustering for Dynamic Data, *IWINAC 2015, Part II, LNCS 9108* (2015) 148–155.
11. Chang-an Liu, Zhen Guo, Chunyang Liu, and Hong Zhou - An Image-Segmentation Method Based on Improved Spectral Clustering Algorithm, *ISIA 2010, CCIS 86* (2010) 178–184.
12. Feng Zhao, Hanqiang Liu, Licheng Jiao - Spectral clustering with fuzzy similarity measure, *Digital Signal Processing* **21** (2011) 701–709.
13. Peng Yang, Qingsheng Zhu, Biao Huang - Spectral clustering with density sensitive similarity function, *Knowledge-Based Systems* **24** (2011) 621–628.
14. Liu H.Q., Jiao L.C., Zhao F. - Non-local spatial spectral clustering for image segmentation, *Neurocomputing* **74** (2010) 461–471.
15. Jun Yan, Debo Cheng, MingZong, and Zhenyun Deng - Improved Spectral Clustering Algorithm Based on Similarity Measure, *ADMA 2014, LNAI 8933* (2014) 641–654.
16. Bezdek J. C. - *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
17. Wang Z. and Bovik A. C. - Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE signal processing magazine*, (2009) 98-117.
18. Wang Z. and Bovik A. C. - A universal image quality index. *IEEE signal processing letters* **9** (3) (2002) 81-84.
19. Eréndira Rendón, Itzel Abundez, Alejandra Arizmendi and Elvia M. Quiroz - Internal versus External cluster validation indexes. *International journal of computers and communications* **1** (5) (2011) 27-34.
20. Chou C.H., Su M.C., Lai E. - A new cluster validity measure and its application to image compression, *Pattern Analysis Applications* **7** (2004) 205–220.