Engineering Applications of Artificial Intelligence 85 (2019) 85–98

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

# Multiple kernel collaborative fuzzy clustering algorithm with weighted super-pixels for satellite image land-cover classification ☆

Trong Hop Dang [a,b], Dinh Sinh Mai [a], Long Thanh Ngo [a,*]

[a] Institute of Simulation Technology, Le Quy Don Technical University, 236 Hoang Quoc Viet, Hanoi, Viet Nam
[b] Hanoi University of Industry, Hanoi, Viet Nam

ARTICLE INFO

*Keywords:*
Fuzzy c-means clustering
Super-pixel based segmentation
Collaborative fuzzy clustering
Land-cover classification
Multiple kernels

ABSTRACT

Applications are dealing with processing the large data size, especially in satellite multi-spectral imagery analysis. In fact, many reasons related to security respect, big data or limited bandwidth, the datasets cann't be centrally clustered by the FCM algorithm or its variants, so the collaborative clustering algorithms are efficiently used. The clustering algorithms have been applied to analyze the surface of the earth, specifically a variety of multispectral satellite image (MSI) classification. In this paper, a novel collaborative fuzzy clustering based framework is proposed, in particular, which is Multiple kernel Collaborative Fuzzy C-means Clustering with weighted super-pixel granulation technique (SMKCFCM algorithms) for satellite image classification. There are two phases consisting of (1) MSI are preprocessed by grouping of the similar pixels into super-pixels granules and then (2) the super-pixels granules are clustered using the collaborative fuzzy clustering with multiple kernels and the their weights. The granule's weight is determined based on its size. It means that amount of the considered objects reduces from a large amount of pixels to only a few hundred super-pixels. In the final step, the performance is improved by the collaborative clustering algorithm on multiple sites of image data. This method is combined with multiple kernels that implicitly convert the original feature space into a higher dimensional space via a non-linear map. This transformation leads to greatly increases the linear separability of the non-spherical and complex input patterns. Experiments were performed on the multi-spectral satellite image datasets and the validity indices were summarized with comparison between the SMKCFCM algorithm and some algorithms in the family of collaborative fuzzy clustering.

## 1. Introduction

Patterns that offer high level of similarity within objects in a dataset can be discovered through clustering techniques. These clustering techniques have been used successfully in various applications during their long development in data mining and machine learning. There are various algorithms using unsupervised learning techniques. In particular, Fuzzy C-Means (FCM) and its variants (Bezdek et al., 1984; Dang et al., 0000) are predominant examples of algorithms using the clustering techniques.

The FCM-based collaborative clustering or collaborative fuzzy clustering (CFC) developed by Pedrycz (2002, 2007, 2008) is used to determine patterns among separate datasets. Collaborative data clustering has two important traits. First, the data in each data site cannot be shared. Second, the discovered patterns can only be shared and it gives high level details than the initial data. According to Pedrycz (2002, 2008), the clustering solution from a data site can immensely impact the findings at a different data site.

Consequently, Coletta et al. (2012) improved on Pedrycz's method to enhance the specifications for pairs of data sites and the number of clusters in individual data sites. A work by Mitra et al. (2006) proposed a collaborative clustering algorithm by taking the combination of rough sets and fuzzy sets.

Yu et al. (2007) and Yu and Yu (2009) concentrated on Horizontal CFC with different spatial attributes or the information provided by the prototypes instead of membership matrixes. He also extended algorithms by labeling targets by levels or entropy, then he used these labels while collaborating, which is called semisupervised collaborative clustering.

When datasets illustrated by numerous views (Jiang et al., 2014), each view with individual characterization is clustered then final clustering result is compounded from multiple view clustering by CFCM in order to get benefit of all view clustering.

Using a distributed P2P network, a novel collaborative clustering was introduced by Zhou et al. (2014). The idea was to make an

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.engappai.2019.05.004.

* Corresponding author.
*E-mail addresses:* dangtronghop@gmail.com (T.H. Dang), maidinhsinh@gmail.com (D.S. Mai), ngotlong@mta.edu.vn (L.T. Ngo).

https://doi.org/10.1016/j.engappai.2019.05.004
Received 19 October 2018; Received in revised form 14 April 2019; Accepted 19 May 2019
Available online xxxx

extensive search for an effective cluster at individual data sites and then exchanging information only with the neighboring data sites. Prasad et al. (2014) introduced the preprocessing phase before running the collaborative phase to avoid some of the deficiency of Pedrycz's method.

Ren and Malik (2003) introduced the idea of super-pixel. It is defined as a group of connected pixels with identical colors or gray levels. The common method in image segmentation is working with pixels but Ren et al. uses super-pixels. Super-pixel segmentation divide an image into hundreds of non-overlapping super-pixels. This type of super-pixel-based image segmentation has gained many achievements in computer vision, which is a convenient way in finding features thereby minimizing complication of image processing problems. Generating super-pixels could be achieved by graph-based method or clustering-based and they are the common approaches to do that. It could be wise to use super-pixels in building graph that results in a more regular lattice (Shi and Malik, 2000), while if focusing on image boundaries, the suitable method may be the graph-based method (Felzenszwalb and Huttenlocher, 2004). The accuracy of classification by linking the spatial information can be greatly improved by super-pixel segmentation.

Homogeneous Super-pixels proposed by Perbet and Maki (2011) aims at applying a Markov-based clustering algorithm that was developed on the graph-based method to compute super-pixels by applying stochastic flow circulation where the nodes correspond to the original pixels. In addition, Liu et al. (2011) proposed a new method called entropy rate super-pixel whose idea is to map image to a graph where the vertex connotes the pixels and edging weights denote the pairwise similarity. This super-pixel segmentation chooses a subset of edges in the resulting graph in which the number of the connected sub-graphs equals the number of super-pixels. Achanta et al. (2012) introduced a clustering based method for super-pixel segmentation called SLIC algorithm.

With the complexity of $O(NlogN)$, Vincent and Soille (1991) was relatively fast but the amount of super-pixels or their compactness cannot be controlled. The resulting super-pixels take high irregulation in both size and shape, and also do not demonstrate good boundary adhesion. Veksler et al. (2010) introduced another kind of super-pixel called constant intensity super-pixels by introducing some constraint into the objective function. Weikersdorfer et al. (2012) proposed the depth-adaptive super-pixels method for the depth images which is used to reduce the complexity of the segmentation task. Besides, linear spectral clustering (Li and Chen, 2015) was introduced by Li et al. in which used kernel functions to transform the image pixels to the weighted super pixels in the kernel space. Next, they uniformly sample the seed pixels in the whole image.

The proposed method by den Bergh et al. (2012) starts from an initial super-pixel partitioning. By revising the boundaries relatively to the objective function defined on super-pixel boundaries and color histogram, the super-pixels are clarified. Papon et al. (2013) proposed the algorithm by using voxel kinships to generate super-pixels which are fully relevant with the spatial relationship in the three-dimensional-spaced scenery. Strategy saliency-based super-pixel (Xu et al., 2014), another type of super-pixels, attains super-pixels through an incorporating strategy on the basis of the salient values of sub-images. Generating super-pixel via the graph matching was developed by Xie et al. (2014) called the Co-super-pixel method. Co-super-pixel achieves super-pixels in a pair of images containing the same or similar objects. It firstly develops super-pixel for each image, after that, these super-pixels are brought together.

Data complexity has increased since the past few years, this poses a great challenge for clustering. The type of features, data size, temporal aspects, and the data diversity are various aspects entailed in the data complexity. Habitually, clustering algorithms and theories have been very well invented for the linear case. However finding clustering solutions for real world data often requires nonlinear methods and a positive definite kernel that can sometimes have the best of both cases (Hofmann et al., 2008; Shawe-Taylor and Cristianini, 2004). Kernel-based clustering (Girolami, 2002; Xiang et al., 2014; Yang and Tsai, 2008) is one of the many approaches to solve this problems that the datasets are reconstructed into a higher feature space, thereby enlarging the feasibility of having linear separable patterns and consequently shortening the associated data prototypes. Other research areas focused on the accomplishment of kernel methods (Graves and Pedrycz, 2010) or what we could gain from the kernel versions of FCM algorithm (Pal and Sarkar, 2014). Anand et al. (2014) introduced semi-supervised kernel mean shift clustering.

Many kernels (Chen et al., 2012; Huang et al., 2012; Sonnenburg et al., 2006) in replacement of a single fixed kernel boost adaptability of kernel gathering and exhibiting that real-life learning problems often consist of many difference data sources. Chen et al. (2012) approached to handle unrelated pixel information with a multiple-kernel FCM (MKFCM) using the combination kernel.

A linear combination of multiple kernels was presented in Chen et al. (2012) with updating the weights of the individual kernel functions. Mai and Ngo (2018) proposed a semi-supervised clustering method based fuzzy c-means algorithm using multiple kernel technique for land-cover classification. A generalized multiple-kernel FCM was proposed by Ganesh and Palanisamy (2012) for satellite image classification. A method of interval type-2 FCM with spatial information (IIT2-FCM) was introduced by Ngo et al. (2015) for the multi-spectral satellite based land cover classification.

Calculation of the membership function in fuzzy clustering may arise their hesitation degree, Chaira solved that problem by incorporating the intuitionistic fuzzy theory into fuzzy clustering (Chaira, 2011). He also expanded his method by combining kernel methods to solve non-linear and non spherical clustering problems (Chaira, 2014) and applications in medical image processing. Another challenge of clustering algorithms is parameter selection, in Kuo-Ping (2014) Kuo-Ping Lin used the genetic (GA) algorithm to set parameters and improve the performance of the KIFCM algorithm. Kuo et al. (2018) also uses GA and some other evolutionary algorithms such as particle swarm optimization and artificial bee colony algorithms to overcome the disadvantages of the KIFCM algorithm in initialing cluster centers.

There are lots of reviews and applications of collaborative clustering into real-world problems, and the kernel methods have widely applied to various fuzzy clustering algorithms. However, one challenge in enhancing the performance of collaborative clustering is the limited study of application of kernel method into collaborative clustering. This paper introduces a novel structure based on collaborative fuzzy clustering for multi-spectral satellite based land cover classification. The multiple-kernel technique is embedded into the collaborative fuzzy clustering with intention of handling the non-spherical complicated structure in the dataset, called Multiple kernel collaborative fuzzy clustering (MKCFCM). Besides, super-pixel technique is useful in grouping the individual pixels into the super-pixel granulation to lessen the complexity. Ideally, the larger the super-pixel granules are, the more stable the final clustering results are. Hence, the weights are designated to the super-pixel granules on the basis of their size. This algorithm is called weighted super-pixels based multiple-kernel collaborative fuzzy clustering (SMKCFCM). We have implemented the framework on the land-cover classification problem of multi-spectral satellite image datasets with three data sites capturing from the three individual areas. Comparison is made between the proposed algorithm and other algorithms like FCM, Collaborative FCM (CFCM), and Interval Type-2 FCM with spatial information (IIT2-FCM) and Multiple Kernels based Collaborative FCM (MKCFCM) on the support of validity indices and outcome of land-cover classification.

The remaining parts of this study are arranged in the following form: Section 2 describes brief exploratory of super-pixel technique, collaborative fuzzy clustering and multiple kernel method. Section 3 introduces the weighted multiple kernel collaborative fuzzy clustering with super-pixel granulation. Section 4 shows the outcome of our experiment on three data sites. Conclusion and future studies are covered in Section 5.

## 2. Preliminaries

### 2.1. Super-pixel

Super-pixels are used in image segmentation as a pre-processing step instead of segmenting pixels directly; the pixels are first grouped into "super-pixels" based on their similarity. These reduce the computational complexity of image processing problems. There are two major advantages when using super-pixels: estimate features on more meaningful regions and reduce the input entities for the subsequent algorithms.

We consider the super-pixel-based segmentation, called SLIC, which has only one input $k$ of the desired number of the approximately sized super-pixels. The segmentation is done on the basis of the clustering process which begins with $k$ initial centroids. The grid interval $S = \sqrt{N/k}$ is used to roughly generate the approximately sized super-pixels with the total number of pixels $N$ (Achanta et al., 2012). The centroids are moved to the seeding locations according to the lowest gradient direction within the spatial window $3 \times 3$. The process is to escape the concentration of an edging super-pixel and to decrease the ability of becoming the seeding location of super-pixels containing noisy pixels.

The distance $D$ based on color and spatial information is used to measure the similarity between pixels:

$$D = \sqrt{d_c^2 + (\frac{d_s}{S}) * m^2} \tag{1}$$

in which, $d_c = \|color_i - color_j\|$ is the Euclidean distance between color of pixel $i$ and color of pixel $j$ in color space and $d_s = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ is the Euclidean distance between pixel $i$ $(x_i, y_i)$ and pixel $j$ $(x_j, y_j)$ in the spatial space and $m$ is the adjusting parameter.

**Algorithm 1** SLIC super-pixel segmentation (Achanta et al., 2012)

Input: the input images, $k$, the threshold.
Output: the resulting super-pixels
1. Initialize cluster centroids by sampling pixels at regular grid spaced by $S$.
2. Move cluster centroids to the locations according to the lowest gradient direction in a $3 \times 3$ window.
3. For each pixel $i$:
4. Set $l(i) = -1$ (denote for label).
5. Set $d(i)=\infty$ (denote for distance).
6. ***do***
7. for each centroid $C_k$:
8. for individual pixel $j$ belongs to a spatial window $2S \times 2S$ around $C_k$:
9. Calculate $D$: the distance from $C_k$ to pixel $i$.
10. if $D < d(i)$ {
11. + Set $d(i) = D$ (set to new distance).
12. + Set $l(i) = k$ (set to new label).
13. }
14. Calculate new centroids.
15. Calculate residual error $E$.
16. ***while*** $E$ <=*threshold.*

In the formula $D$, parameter $m$ is also used to weight the relative importance between color similitude and spatial proximity. The larger the $m$ takes, more important the spatial proximity is and more compact the resulting super-pixels are.

### 2.2. Fuzzy clustering using kernel method

The kernel based clustering method is realized the clustering problems in the new feature space. Firstly, a transformation is done on the original data space by a nonlinear mapping into the higher dimensional feature space. Next, the clustering problem is done in the higher dimensional feature space with the kernel function can be used to calculate the product in the input space (Graves and Pedrycz, 2010).

Two main types are mentioned about the kernel-based FCM algorithm. The first type is to put the prototypes in the kernel space, called KFCM-F. In the second one, we form the prototypes in the original feature space, denoted KFCM-K, by calculating an inverse mapping from the kernel space to the original feature space (Graves and Pedrycz, 2010). The objective functions of KFCM-F and KFCM-K have the same constraints with FCM as follows

The KFCM-F objective function:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m(\varphi(x_k) - v_i)^2 \tag{2}$$

The KFCM-K objective function:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m(\varphi(x_k) - \varphi(v_i))^2 \tag{3}$$

in which $\varphi(x)$ is a mapping from the original feature space to the kernel space, $N$ is the number of patterns.

### 2.3. Collaborative Fuzzy clustering

Let $D[i_s]$, $i_s = 1, \ldots, K$ be $K$ data sites and data site $i_s$ contains $n_{is}$ data patterns determined in the same feature space $X$. All patterns are grouped into $c$ clusters at individual data sites. The clustering results at each data sites are effected by the remain data sites, this process is called collaborative process.

The objective function of individual data site based on the FCM algorithm is known as:

$$\sum_{k=1}^{n_{is}}\sum_{i=1}^{c} u_{ik}^2[i_s]d_{ik}^2, i_s = 1, \ldots, \tag{4}$$

The collaboration between each data site is done through other data sites, the intensity of the interaction is described by factor $\beta$. The collaborative fuzzy clustering extends the objective function becoming the form as follows.

$$J_{[i_s]} = \sum_{k=1}^{n_{is}}\sum_{i=1}^{c} u_{ik}^2[i_s]d_{ik}^2 + \beta\sum_{j_s=1}^{K}\sum_{k=1}^{n_{is}}\sum_{i=1}^{c}(u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 d_{ik}^2 \tag{5}$$

There are two parts in the formula in which the first one is the FCM-based objective function. The second part reproduces the influence of structural knowledge from other sites. Among parameters, $\tilde{u}[i_s|j_s]$ reflects the influence of the data site $i_s$ to the data site $j_s$, called the induced matrix and calculated by the following formula:

$$\tilde{u}_{ik}[i_s|j_s] = \frac{1}{\sum_{j=1}^{c}\left(\frac{|x_k[i_s] - v_i[j_s]|}{|x_k[i_s] - v_j[j_s]|}\right)^2} \tag{6}$$

The collaborative fuzzy clustering is considered as the optimization problem with constraints as follows:

1. Min J$[i_s]$
2. s.t. $\mathbf{u}[i_s] \in U$

where $\mathbf{u}$ is a set of membership matrices described as follows:

$$\mathbf{u} = \left\{ u_{ik}[i_s] \in [0,1] | \sum_{i=1}^{c} u_{ik}[i_s] = 1 \forall k, i \& 0 < \sum_{k=1}^{n_{is}} u_{ik}[i_s] < n_{is} \right\} \tag{7}$$

The objective function (5) is resolved by using the Lagrange method to find the matrix $\mathbf{u}$ and $\mathbf{v}$ as follows:

$$u_{rs}[i_s] = \frac{1}{\sum_{j=1}^{c} d_{rs}^2/d_{js}^2}\left[1 - \sum_{j=1}^{c}\frac{\beta\sum_{j_s=1, j_s\neq i_s}^{K}\tilde{u}_{js}[i_s|j_s]}{(1+\beta(K-1))}\right] + \frac{\beta\sum_{j_s=1, j_s\neq i_s}^{K}\tilde{u}_{rs}[i_s|j_s]}{(1+\beta(K-1))} \tag{8}$$

$$v_{rt}[i_s] = \frac{\sum_{k=1}^{n_{is}} u_{rk}^2[i_s]x_{kt} + \beta\sum_{j_s=1, j_s\neq i_s}^{K}\sum_{k=1}^{n_{is}}(u_{rk}[i_s] - \tilde{u}_{rk}[i_s|j_s])^2 x_{kt}}{\sum_{k=1}^{n_{is}} u_{rk}^2[i_s] + \beta\sum_{j_s=1, j_s\neq i_s}^{K}\sum_{k=1}^{n_{is}}(u_{rk}[i_s] - \tilde{u}_{rk}[i_s|j_s])^2} \tag{9}$$

for $r = 1, \ldots, c; s = 1, \ldots, n_{is}$ and $t = 1, \ldots, M$ (the number of attributes);

The coefficient $\beta[i_s|j_s]$ represents the collaborative level between the data site $i_s$ and $j_s$, the greater $\beta$ is, the more effective the collaborative level is and vice versa. When the data sites have the similar structure, the collaborative level will be greater i.e $\beta$ takes a higher value. The value of $\beta$ can be estimated by experts or computed on the basis of the structural similarity between the data sites. In collaborative model, the prototypes $v[j_s]$ were sent from data site $j_s$ to data site $i_s$ and the induced objective function can computed:

$$\tilde{J}[i_s|j_s] = \sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \tilde{u}_{ik}^2[i_s|j_s]|x_k - v_i[j_s]|^2 \tag{10}$$

The coefficient $\beta[i_s|j_s]$ denoted the interactive level between two sites $i_s$ and $j_s$ is computed as follows:

$$\beta[i_s|j_s] = \min\left\{1, \frac{J[i_s]}{\tilde{J}[i_s|j_s]}\right\} \tag{11}$$

So, the membership degree and cluster centroid matrix is calculated in the following equation:

$$u_{rs}[i_s] = \frac{1}{\sum_{j=1}^{c} d_{rs}^2/d_{js}^2}\left[1 - \sum_{j=1}^{c}\sum_{j_s=1, j_s\neq i_s}^{K} \frac{\beta[i_s|j_s]\tilde{u}_{js}[i_s|j_s]}{(1+\beta[i_s|j_s](K-1))}\right] + \sum_{j_s=1, j_s\neq i_s}^{K} \frac{\beta[i_s|j_s]\tilde{u}_{rs}[i_s|j_s]}{(1+\beta[i_s|j_s](K-1))} \tag{12}$$

$$v_{rt}[i_s] = \frac{\sum_{k=1}^{n_{is}} u_{rk}^2[i_s]x_{kt} + \sum_{j_s=1, j_s\neq i_s}^{K}\sum_{k=1}^{n_{is}} \beta[i_s|j_s](u_{rk}[i_s] - \tilde{u}_{rk}[i_s|j_s])^2 x_{kt}}{\sum_{k=1}^{n_{is}} u_{rk}^2[i_s] + \sum_{j_s=1, j_s\neq i_s}^{K}\sum_{k=1}^{n_{is}} \beta[i_s|j_s](u_{rk}[i_s] - \tilde{u}_{rk}[i_s|j_s])^2} \tag{13}$$

## 3. Collaborative Fuzzy C-means clustering with weighted super-pixels granulation and multiple kernels

### 3.1. Super-pixel granulation

Granular Computing (GrC) is considered as effectively computing method based on forming granules from intervals, classes or clusters to model a computational diagram for resolving complex problems (Pedrycz, 2001). In satellite imagery, granules can be treated as an area of pixels that have similar features. This coincides with the idea of creating super pixels where we can consider a super pixel as a granule. A granule is represented as a image super pixel, and the distance between granules is defined by the centers of super pixels. We use the SLIC algorithm to generate super-pixel granules from satellite images. Because the super-pixels are often of the different size, so the super-pixel based classification often produces the results limited than the pixel based classification. The larger super-pixels normally tend to attract the small super-pixels toward themselves, so the effect of individual super-pixels on the resulting clusters is different. To solve this, we propose the usage of the weight $\varphi$ for individual super-pixel, the weight $\varphi$ could be calculated based on the number of pixels in the super-pixels, could be as follows:

$$\varphi_k = \frac{n_k}{n_{is}} \tag{14}$$

where $0 < \varphi_k < 1$ be weight of the $k$th super-pixel, $n_k$ be the number of pixels of the $k$th super-pixel, $n_{is}$ be the total of pixels of the image at the data site $i_s$.

### 3.2. Collaborative Fuzzy C-means clustering with weighted super-pixels granulation and multiple kernels

We cluster the super pixels instead of pixels, each super pixel have the weight $\varphi$ so the objective function at the data site $i_s$ by using the standard FCM algorithm is described as follows

$$J_{[i_s]} = \sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s] d_{ik}^2 \quad i_s = 1, \ldots, K \tag{15}$$

Which is modified into the following function in collaborative clustering:

$$J_{[i_s]} = \sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s] d_{ik}^2 + \beta \sum_{j_s=1}^{K}\sum_{k=1}^{n_{is}}\sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 d_{ik}^2 \tag{16}$$

Kernel methods are used to originate nonlinear relationships among data by using the maps to transform the features into the kernel spaces. Let $\phi = \{\phi_1, \phi_2, \ldots, \phi_M\}$ be a set of $M$ maps. Each map $\phi_k$ decodes the pattern $x$ into a vector $\phi_k(x)$ in its $M$-dimensional feature space. Consider $\{P_1, P_2, \ldots, P_M\}$ is a set of Mercer kernels associating to these implicit maps, respectively.

$$P_k(x_i, x_j) = \phi_k(x_i)^T \phi_k(x_j) \tag{17}$$

We consider a resulting kernel satisfied Mercer's condition by using non-negative combination of the feature maps, $\phi'$, which is defined as follows:

$$\phi'(x) = \sum_{k=1}^{M} \omega_k \phi_k(x) \tag{18}$$

Let it is easy to form a set of the independent maps, $\psi = \{\psi_1, \psi_2, \ldots, \psi_M\}$, from the original maps $\phi$ as:

$$\psi_1(x) = \begin{bmatrix} \phi_1(x) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \psi_2(x) = \begin{bmatrix} 0 \\ \phi_2(x) \\ \vdots \\ 0 \end{bmatrix}, \ldots, \psi_M(x) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \phi_M(x) \end{bmatrix} \tag{19}$$

Constructing new maps by this way guarantees that there is the same dimensionality in the feature spaces of these maps and can form the linear combination well. In addition, a new set of orthogonal bases is defined since.

$$\begin{aligned} &\psi_k(x_i)^T \psi_k(x_j) = P_k(x_i, x_j) \\ &\psi_k(x_i)^T \psi_{k'}(x_j) = 0 | k \neq k' \end{aligned} \tag{20}$$

As the crossing terms from implicit maps may be obstructed by such orthogonal bases, the inner product in the same map may be well appraised by the kernel functions. Accordingly, we revise the objective function as follows:

$$J_{[i_s]} = \sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s](\psi(x_k) - v_i[i_s])^2 + \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s\neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 (\psi(x_k) - v_i[i_s])^2 \tag{21}$$

$$\psi(x) = \omega_1\psi_1(x) + \omega_2\psi_2(x) + \cdots + \omega_M\psi_M(x) \tag{22}$$

Subject to $\omega_1 + \omega_2 + \cdots + \omega_M = 1$, $\omega_k \geq 0, \forall k$; $\sum_{j=1}^{c} u_{js}[i_s] = 1, \forall s$; $u_{js}[i_s] \geq 0, \forall s, j$.

Where $v_i$ is the centroid $i$th defined in the implicit feature space, $(\omega_1, \omega_2, \ldots, \omega_M)$ is a weight vector for composed multiple kernel. The distance $d_{ik}$ concerns the $k$th pattern and the $i$th centroid in $D[i_s]$: $d_{ik}^2 = (\psi(x_k) - v_i)^2$.

The partition matrix satisfying the following conditions:

$$\mathbf{u} = \left\{ u_{ik} \in [0,1] | \sum_{i=1}^{c} u_{ik}[i_s] = 1, \forall k \& 0 < \sum_{k=1}^{n_{is}} u_{ik}[i_s] < n_{is}, \forall i \right\} \tag{23}$$

The distance $d_{ik}[i_s|j_s]$ defines the distance between the $k$th pattern in $D[i_s]$ and the $i$th prototype in $D[j_s]$.

We state the following theorem to determine the components of the objective function.

**Theorem 1.** *The objective function $J_{[i_s]}$ in (21) attain the local minima when $U = [u_{ik}]_{cn_{i_s}}$ satisfy the following relationship*

$$u_{ik}[i_s] = \frac{\sum_{j_s=1, j_s\neq i_s}^{K} \beta[i_s|j_s]\tilde{u}_{ik}[i_s|j_s]}{(\varphi_s + \sum_{j_s=1, j_s\neq i_s}^{K} \beta[i_s|j_s])}$$

$$+\frac{1}{\sum_{j=1}^{c}\frac{d_{ik}^2}{d_{jk}^2}}\left[1-\sum_{j=1}^{c}\frac{\sum_{j_s=1,j_s\neq i_s}^{K}\beta[i_s|j_s]\tilde{u}_{jk}[i_s|j_s]}{(\varphi_s+\sum_{j_s=1,j_s\neq i_s}^{K}\beta[i_s|j_s])}\right] \tag{24}$$

*where the distance between object k and cluster i is calculate as*

$$d_{ik}^2=\sum_{t=1}^{M}\alpha_{ikt}\omega_t^2 \tag{25}$$

*In which: Eq. (26) is given in Box I*

$$\omega_t=\frac{1}{\sum_{t=1}^{M}\frac{\sum_{k=1}^{n_{is}}\sum_{i=1}^{c}\varphi_k u_{ik}^2[i_s]\alpha_{ikt}+\sum_{k=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K}\beta[i_s|j_s]\sum_{i=1}^{c}(u_{ik}[i_s]-\tilde{u}_{ik}[i_s|j_s])^2\alpha_{ikt}}{\sum_{k=1}^{n_{is}}\sum_{i=1}^{c}\varphi_k u_{ik}^2[i_s]\alpha_{ikt}+\sum_{k=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K}\beta[i_s|j_s]\sum_{i=1}^{c}(u_{ik}[i_s]-\tilde{u}_{ik}[i_s|j_s])^2\alpha_{ikt}}} \tag{27}$$

*for* $s=1,\ldots,n_{is}$ *and* $r=1,\ldots,c$.

The proof of Theorem 3.1 is included in Appendix.

### 3.3. SMKCFCM algorithm

The algorithm of the multiple kernel collaborative fuzzy clustering with weighted super-pixels granulation (SMKCFCM) is described in Algorithm 2. The substance of collaborative clustering is to collaboratively discover the structures in individual data sites through exchanging prototypes between data sites. Two main phases are mentioned in collaborative fuzzy clustering: generating super-pixels and clustering data in each site by the clustering algorithm known as FCM, and implementation of re-clustering based on the collaboration of exchanging the data clustering results coming from phase 1.

Firstly, the SLIC algorithm is used to create super-pixels from the input data. Then, the FCM-typed algorithm is locally carried out for the individual sites. In the collaborative phase, the prototypes at the individual site are communicated to all the remaining sites. Then, the membership matrix and the prototype at the individual site are optimized after each iteration until the terminal criterion has been satisfied.

**Algorithm 2** SMKCFCM Algorithm

**Input:** $K$: the number of sites, $n_{is}$: the number of patterns in site $i_s$, $c[i_s]$: the number of clusters in data site $i_s$, $M$: the number of attributes, $X[i_s]$: patterns in the site $i_s$.

**Output:** **u**: membership matrix and $\omega$: feature weights.

1. **Phase 1: forming super-pixel using SLIC algorithm and locally FCM-type clustering**
2. Calculating Super-pixel granules by using Algorithm 1 for each data site.
3. Locally clustering by using FCM for each data site based on Super-pixels granules.
4. **Phase 2: Collaborative clustering with weighted super-pixels granulation and multiple kernels**
5. ***Repeat***
6. Communicate partition matrices and prototypes between data sites.
7. For each data site $D[i_s]$.
8. Calculate induced matrices using (6).
9. do
10. Calculate matrix $\alpha$ by the formula (26).
11. Calculate weight of attributes $\omega$ using (27).
12. Updating partition matrix **u** using (24) and (25).
13. while $\max\|\mathbf{u}[i_s]^{(k)}-\mathbf{u}[i_s]^{(k-1)}\|>\varepsilon$.
14. End for
15. ***Until*** $\max\|\mathbf{v}[i_s]^{(k)}-\mathbf{v}[i_s]^{(k-1)}\|<\varepsilon_v$.

## 4. Experimental studies

In this section, we carry out some experiments on real data to evaluate the performance of SMKCFCM. The collaborative fuzzy clustering with multiple kernel (MKCFCM) algorithm (Dang et al., 2016), IIT2-FCM (Ngo et al., 2015) algorithm, CFCM (Pedrycz, 2008) algorithm, KFCM (Bezdek et al., 1984) algorithm and Kernel Intuitionistic Fuzzy Clustering (KIFCM) (Chaira, 2014) algorithm are used to compare and analyze by following clustering validity indices: Fuzzy Silhouette Criterion, Sum of Squared Error (Dang et al., 0000), the Dunn's separation index (D-I), the Separation index (S-I) and Classification Entropy index (CE-I) (Wang and Zhang, 2007) and Davies–Bouldin's Index (DB-I), Bezdek's partition coefficient (PC-I) (Davies and Bouldin, 1979), Xie–Beni Index (XB-I) (Xie and Beni, 1991), Chen–Linkens Index (CL-I) (Chen and Linkens, 2004) and Generalized C Index (GC-I) (Bezdek et al., 2016). The better algorithms exhibit smaller values of S-I, CE-I, DB-I, SSE, XB-I and GC-I; the larger value of D-I, PC-I, FS and CL-I. To form multiple kernels for MKCFCM and SMKCFCM, we used three kernel functions consisting of Gaussian kernel $P_1$ (for both KFCM and KIFCM), Polynomial kernel $P_2$ and Hyperbolic Tangent (Sigmoid) Kernel $P_3$, set parameter $\delta^2=4$ in $P_1$; parameters $c=20$, $p=2$ in $P_2$; and parameters $\alpha=0.3$, $\beta=10$ in $P_3$ in the following form:

$$P_1(x,y)=\exp\left(\frac{\|x-y\|^2}{2\delta^2}\right),\delta\in\mathbb{R}^+$$

$$P_2(x,y)=(xy+c)^p,c\in\mathbb{R}^+,p\in\mathbb{N}^+$$

$$P_3(x,y)=tanh(\alpha xy+\beta),\alpha,\beta\in\mathbb{R}^+$$

The parameters of IIT2-FCM was chosen the same with Ngo et al. (2015) as: mask size is 5 × 5, $m_1$=1.8; $m_2$ = 3.5; all algorithm have maximize loop number = 100 and epsilon = 0.00001.

The experiment concerns data coming from satellite images with 3 data sites:

*Data site 1:* Thanh Hoa city (Vietnam) was captured in 2007 (20°29′27.81″N, 106°32′21.65″E to 21°27′15.41″N, 104°35′23.54″E). The image consists of 54,894 pixels with the resolution of 30 m × 30 m, i.e. the area is 49.405 km$^2$ (see Fig. 1).

Data site 2: Thai Nguyen city (Vietnam) was captured in 2010 ($11^0$18′29.13″N, $108^0$18′10.57″E to 11°58′29.63″N, 107°01′44.93″E) with the area of 6822.165 km$^2$. The image consists of 193,039 pixels and the area is 173.735 km$^2$ (see Fig. 2).

*Data site 3:* Quy Hop region(Nghe An province, Vietnam) captured in 2010 (19°06′36.68″N, 105°42′0.12″E to 19°34′33.50″N, 104°38′13.71″E) consists of 1,047,825 pixels and the area is 943.043 km$^2$ (see Fig. 3).

These images are captured by Landsat satellite with 7 bands and provided by Vietnam National Remote Sensing center (VNRS). We considered these image datasets as three data sites and only used two bands (band 3 and band 4) as two attributes of data patterns to realize collaborative clustering. The datasets are clustered into six classes as follows: Class 6: Jungles; Class 5: Perennial tree crops; Class 4: Planted forests, low woods; Class 3: Fields, grass; Class 2: Rocks, bare soil; Class 1: Rivers, ponds,

**Fig. 1.** Band 3 (left) and band 4 of Thanh Hoa city area.

$$\alpha_{ikt} = P_t(x_k, x_k) - 2\frac{\sum_{j=1}^{n_{is}} \varphi_j u_{ij}^2[i_s] P_t(x_k, x_j) + \sum_{j=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s] - \tilde{u}_{ij}[i_s|j_s])^2 P_t(x_k, x_j)}{\sum_{j=1}^{n_{is}} \varphi_j u_{ik}^2[i_s] + \sum_{j=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s] - \tilde{u}_{ij}[i_s|j_s])^2}$$
$$+\frac{\sum_{j1=1}^{n_{is}} \sum_{j2=1}^{N[i_s]} \varphi_{j1}\varphi_{j2} u_{ij1}^2[i_s] u_{ij2}^2[i_s] P_t(x_{j1}, x_{j2}) + 2\sum_{j1=1}^{n_{is}} \sum_{j2=1}^{N[i_s]} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] u_{ij1}^2 (u_{ij2}[i_s] - \tilde{u}_{ij2}[i_s|j_s])^2 P_t(x_{j1}, x_{j2})}{\left(\sum_{j1=1}^{n_{is}} \varphi_{j1} u_{ik}^2[i_s] + \sum_{j1=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s] - \tilde{u}_{ij}[i_s|j_s])^2\right)^2} \quad (26)$$
$$+\frac{\sum_{j_s=1, j_s \neq i_s}^{K} \sum_{j1=1}^{n_{is}} \sum_{j2=1}^{n_{is}} \varphi_{j1}\varphi_{j2}\beta^2[i_s|j_s](u_{ij1}[i_s] - \tilde{u}_{ij1}[i_s|j_s])^2 (u_{ij2}[i_s] - \tilde{u}_{ij2}[i_s|j_s])^2 P_t(x_{j1}, x_{j2})}{\left(\sum_{j1=1}^{n_{is}} \varphi_{j1} u_{ik}^2[i_s] + \sum_{j1=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s] - \tilde{u}_{ij}[i_s|j_s])^2\right)^2}$$

**Box I.**

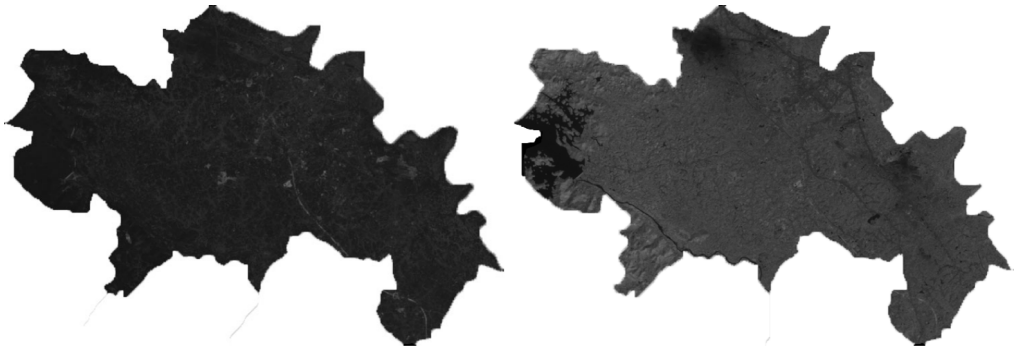

**Fig. 2.** Band 3 (left) and band 4 of Thai Nguyen city area.

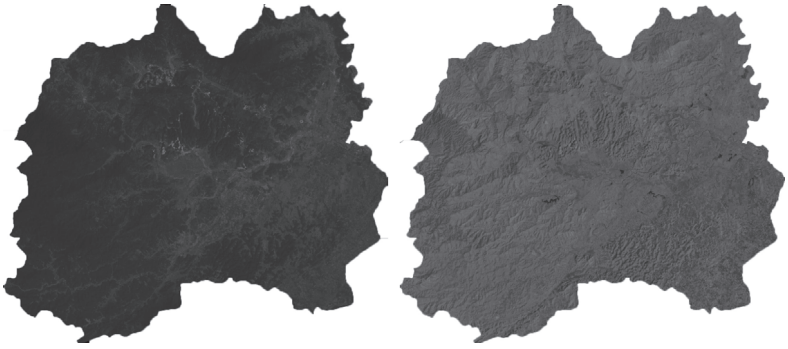

**Fig. 3.** Band 3 (left) and band 4 of Quy Hop area.

lakes [redacted]. The number of clusters is 6 and all algorithms initialize randomly cluster centers; the results coming from the KIFCM, KFCM, CFCM, IIT2-FCM, MKCFCM, and SMKCFCM are averaged over all three data sites.

Fig. 4 demonstrates the experimental diagram in which three datasets of Thanh Hoa, Thai Nguyen and Quy Hop are located at three data sites. The datasets are locally processed at Phase 1 by running the SLIC and FCM algorithms on the individual dataset. At the collaborative phase, three outputs produced from the algorithms on the data sites are clustered together by the SMKCFCM algorithm. The resulting clusters at three the data sites are displayed, fused or analyzed through the particular tasks.

Figs. 5, 7 and 9 display the clustered images obtained when running the individual algorithms on dataset of Thanh Hoa, Thai Nguyen and Quy Hop areas, respectively.

Table 1 shows the number of misclassified pixels in each class and Fig. 6 presents the percentage of area of each class which are produced by different algorithms and the surveying data provided by Vietnam National Remote Sensing center (VNRS) on Thanh Hoa dataset. The average percentage differences of KIFCM, KFCM, CFCM, IIT2-FCM, MKCFCM, SMKCFCM and VNRS are 1.11%, 5.78%, 6.74%, 1.97%, 0.60% and 0.56% respectively.

**Table 1**
The number of misclassified pixels producing by the algorithms on Thanh Hoa dataset.

| Class | KIFCM | KFCM | CFCM | IIT2-FCM | MKCFCM | SMKCFCM |
|---|---|---|---|---|---|---|
| Class 1 | 601 | 2 355 | 3 241 | 1277 | 393 | 43 |
| Class 2 | 921 | 1 886 | 4 224 | 295 | 98 | 265 |
| Class 3 | 691 | 4 120 | 4 911 | 1965 | 786 | 565 |
| Class 4 | 603 | 3 519 | 1 964 | 982 | 491 | 97 |
| Class 5 | 473 | 1 819 | 5 500 | 1572 | 98 | 54 |
| Class 6 | 704 | 5 352 | 2 357 | 392 | 98 | 829 |
| Total | 3993 | 19 051 | 22 197 | 6482 | 1964 | **1854** |

Table 3 shows the number of misclassified pixels in each classes and Fig. 8 presents the percentage of area of each class which are produced by different algorithms and the surveying data provided by VNRS on

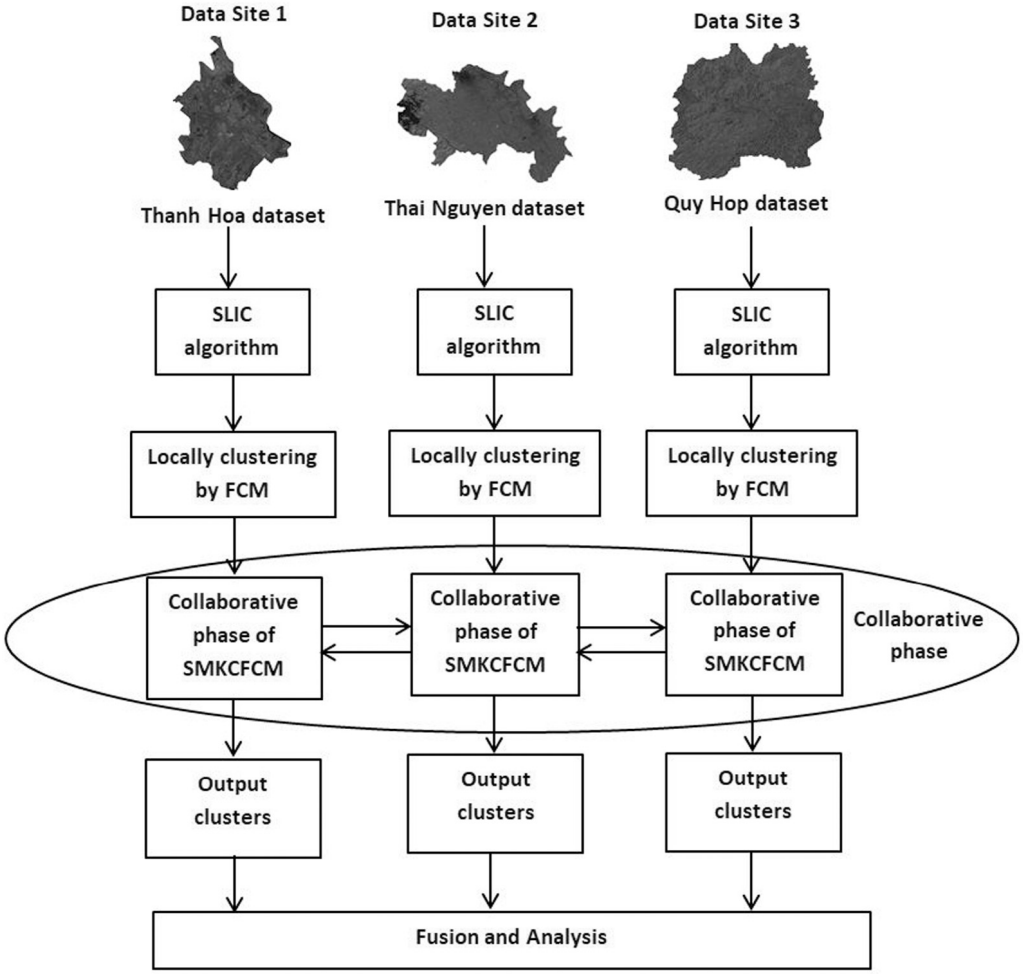


**Fig. 4.** Experimental Diagram.

**Table 2**
Validity Indices (VI) obtained for Thanh Hoa dataset.

| VI | KIFCM | KFCM | CFCM | IIT2-FCM | MKCFCM | SMKCFCM |
|---|---|---|---|---|---|---|
| FS | 4.3812 | 3.2478 | 3.7612 | 3.9871 | 4.6735 | **4.9241** |
| SSE | 92.0647 | 142.3514 | 122.8734 | 98.8745 | 88.7621 | **79.6247** |
| D-I | 0.7951 | 0.3150 | 0.3549 | 0.7875 | 0.7963 | **0.8012** |
| DB-I | 1.3283 | 4.0548 | 4.7643 | 1.3794 | **1.2623** | 1.2791 |
| PC-I | 0.8542 | 0.5127 | 0.6823 | 0.7862 | 0.8632 | **0.8924** |
| CE-I | 1.0025 | 3.2019 | 2.8961 | 0.9982 | 0.9985 | **0.9627** |
| S-I | 0.4163 | 0.7518 | 0.7862 | 0.4672 | 0.3672 | **0.2843** |
| XB-I | 0.1421 | 0.2694 | 0.2765 | 0.1382 | 0.1487 | **0.1305** |
| CL-I | 0.9521 | 0.8201 | 0.8970 | 0.9657 | 0.9553 | **0.9826** |
| GC-I | 2.7019 | 3.2410 | 2.9826 | 2.6769 | 2.8750 | **2.0034** |
| Mean | 1.11% | 5.78% | 6.74% | 1.97% | 0.60% | **0.56%** |

**Table 3**
The number of misclassified pixels producing by the algorithms on Thai Nguyen dataset.

| Class | KIFCM | KFCM | CFCM | IIT2-FCM | MKCFCM | SMKCFCM |
|---|---|---|---|---|---|---|
| Class 1 | 2 575 | 9 974 | 8 198 | 3 199 | 2 199 | 1 031 |
| Class 2 | 929 | 10 048 | 9 999 | 4 000 | 3 000 | 1 795 |
| Class 3 | 1 853 | 11 163 | 13 600 | 6 600 | 4 600 | 2 598 |
| Class 4 | 3 139 | 11 318 | 13 602 | 5 600 | 2 600 | 1 867 |
| Class 5 | 3 048 | 8 052 | 9 200 | 6 200 | 3 199 | 1 407 |
| Class 6 | 9 687 | 7 824 | 11 001 | 6 002 | 2 002 | 1 444 |
| Total | 21 230 | 58 379 | 65 600 | 31 601 | 17 600 | **10 142** |

Thai Nguyen dataset in which the average percentage differences of KIFCM, KFCM, CFCM, IIT2-FCM, MKCFCM, SMKCFCM and VNRS are 1.81%, 5.04%, 5.66%, 2.73%, 1.52% and 0.88% respectively.

**Table 4**
Validity Indices (VI) obtained for Thai Nguyen dataset.

| VI | KIFCM | KFCM | CFCM | IIT2-FCM | MKCFCM | SMKCFCM |
|---|---|---|---|---|---|---|
| FS | 4.2617 | 3.0427 | 3.2985 | 4.2761 | 4.3872 | **4.5102** |
| SSE | 101.4680 | 130.5861 | 119.2763 | 106.2745 | 98.8736 | **91.2381** |
| D-I | 0.8026 | 0.2683 | 0.3894 | 0.7658 | 0.8162 | **0.8703** |
| DB-I | 1.4293 | 6.0385 | 4.6253 | 1.5621 | 1.3876 | **1.231** |
| PC-I | 0.9078 | 0.4162 | 0.7982 | 0.8742 | 0.9031 | **0.9128** |
| CE-I | 0.6238 | 3.1038 | 2.0983 | 0.7761 | 0.5712 | **0.5143** |
| S-I | 0.4208 | 0.8315 | 0.7183 | **0.3981** | 0.4128 | 0.3997 |
| XB-I | 0.1926 | 0.4653 | 0.3828 | 0.2086 | 0.1873 | **0.1735** |
| CL-I | 0.9381 | 0.8041 | 0.8659 | **0.9593** | 0.9240 | 0.9472 |
| GC-I | 3.0528 | 4.1861 | 4.1948 | 3.6828 | 3.1950 | **2.7315** |
| Mean | 1.81% | 5.04% | 5.66% | 2.73% | 1.52% | **0.88%** |

**Table 5**
The number of misclassified pixels producing by the algorithms on Quy Hop dataset.

| Class | KIFCM | KFCM | CFCM | IIT2-FCM | MKCFCM | SMKCFCM |
|---|---|---|---|---|---|---|
| Class 1 | 1 352 | 102 624 | 80 944 | 30 953 | 9 818 | 6 507 |
| Class 2 | 15 487 | 99 019 | 95 698 | 35 699 | 19 626 | 12 899 |
| Class 3 | 19 605 | 54 089 | 57 903 | 32 902 | 19 626 | 8 121 |
| Class 4 | 13 276 | 98 936 | 99 439 | 34 442 | 13 737 | 9 221 |
| Class 5 | 3 657 | 121 810 | 122 218 | 52 213 | 3 929 | 10 436 |
| Class 6 | 14 177 | 80 546 | 55 922 | 5 920 | 4 | 2 934 |
| Total | 67 553 | 557 024 | 512 124 | 192 129 | 66 740 | **50 117** |

Similarly, Table 5 shows the number of misclassified pixels in each classes and Fig. 10 demonstrates the results produced by the algorithms on Quy Hop dataset with the average differences of KIFCM, KFCM, CFCM, IIT2-FCM, MKCFCM, SMKCFCM and VNRS are 0.81%, 5.04%,

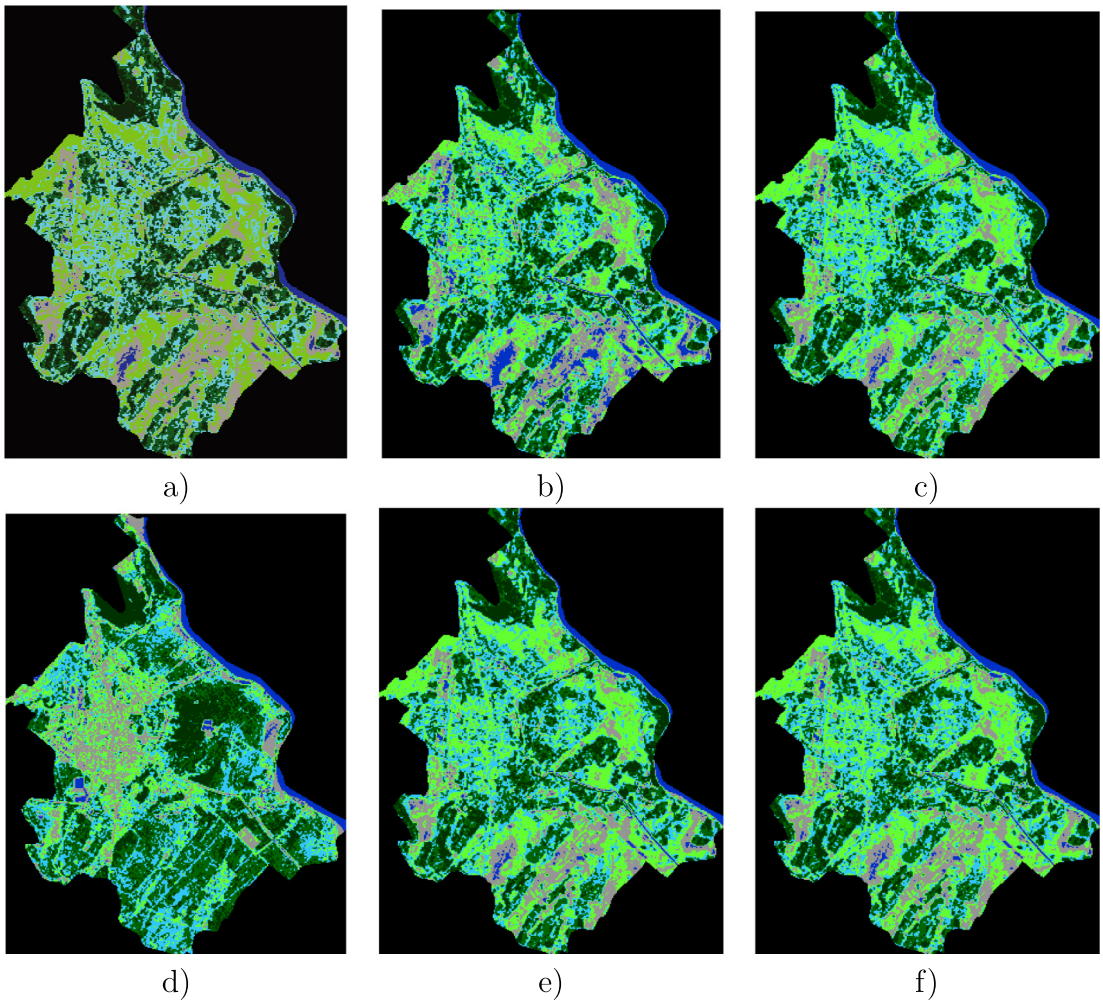


**Fig. 5.** Classification on Thanh Hoa dataset (a) KIFCM, (b) KFCM, (c) CFCM, (d) IIT2-FCM, (e) MKCFCM, (f) SMKCFCM.

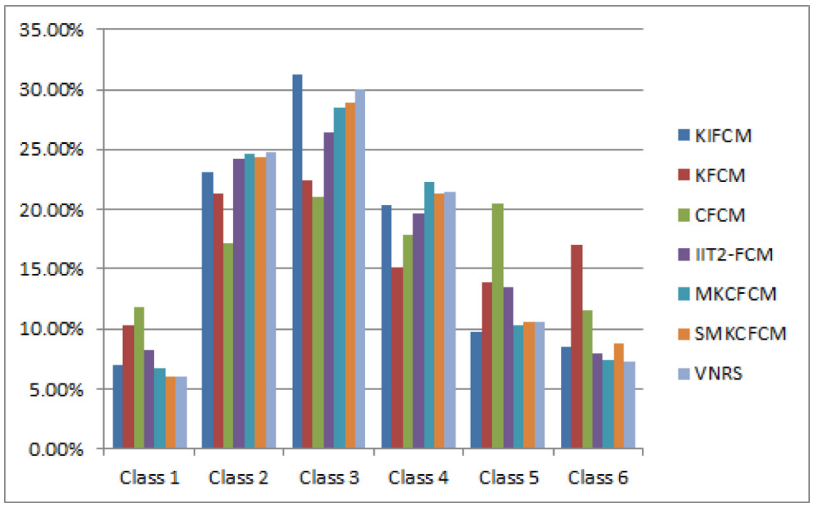


**Fig. 6.** The comparing chart between the classes of Thanh Hoa dataset.

5.66%, 2.73%, 1.52% and 0.80%, respectively. The vast majority, the SMKCFCM yields the best results.

In addition, the clustering results are analyzed on the basis of various assessment indices to evaluate performance of the considered algorithms on the experimental datasets. The results in Tables 2, 4, 6 and the average of validity indices for all datasets in Table 7 show that the SMKCFCM algorithm gains the better quality than the considered algorithms like KIFCM, KFCM, CFCM, IIT2-FCM and MKCFCM in almost indices on the experimental datasets .

The clustered images of each algorithm in Figs. 5, 7 and 9 show that the SMKCFCM algorithm gives clearer clustering results in overlapping regions between classes and the split line has a nonlinear shape. This

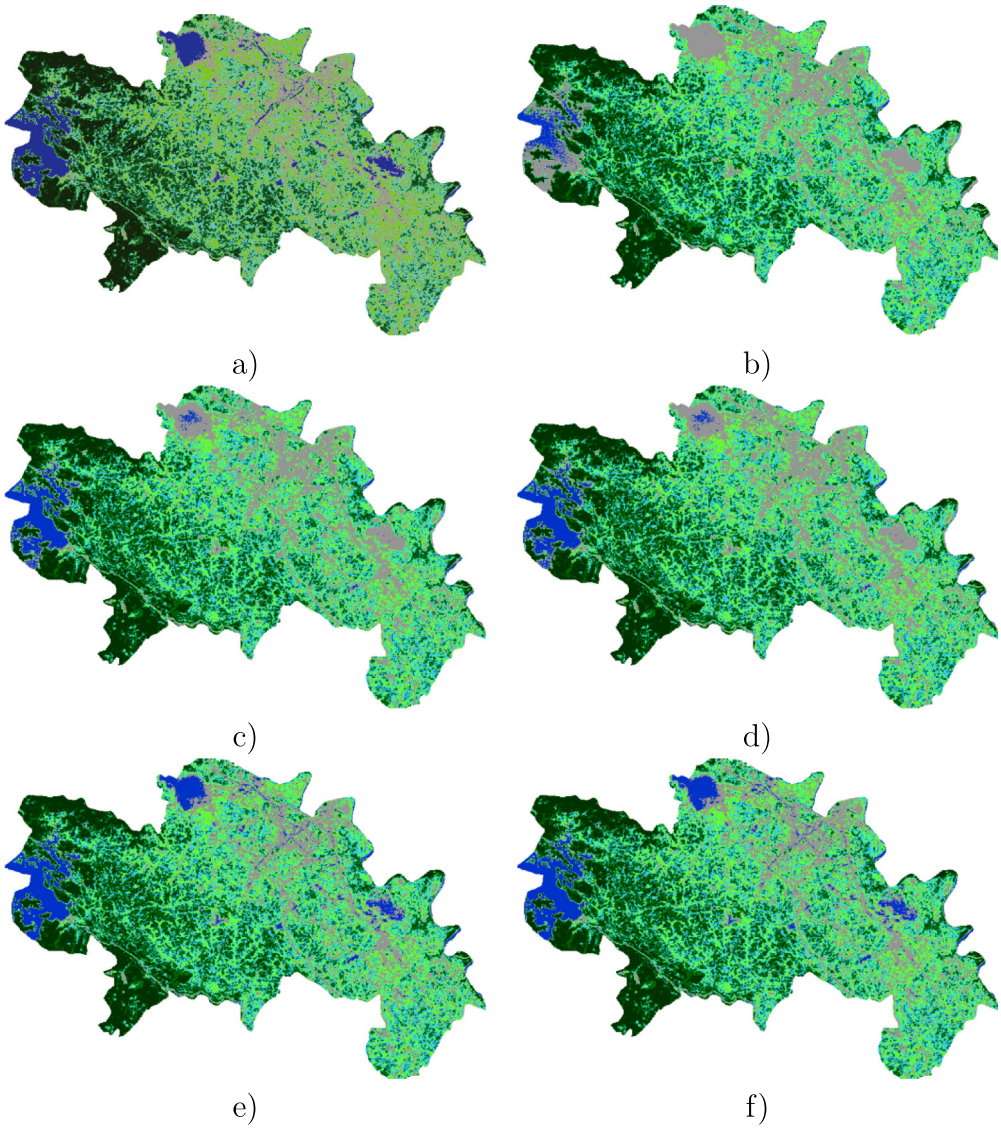


**Fig. 7.** Classification on Thai Nguyen dataset (a) KIFCM (b) KFCM, (c) CFCM, (d) IIT2-FCM, (e) MKCFCM, (f) SMKCFCM.

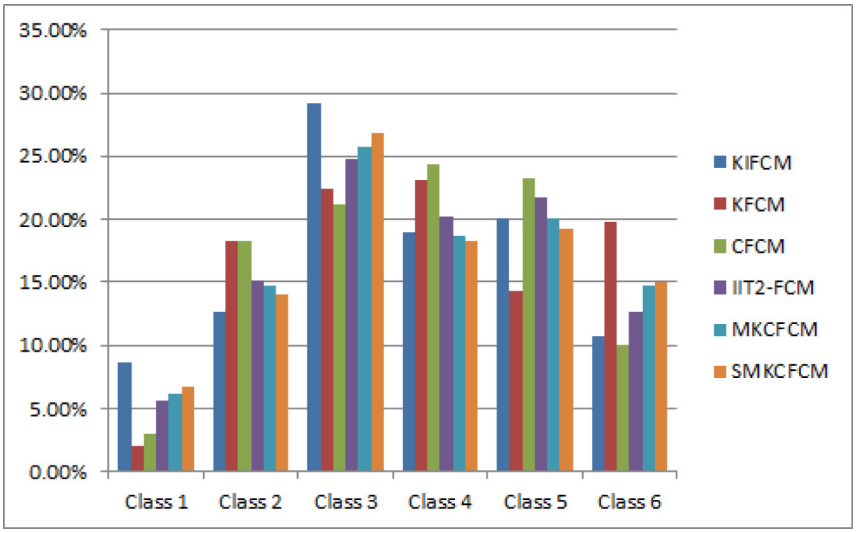


**Fig. 8.** The comparing chart between the classes of Thai Nguyen dataset.

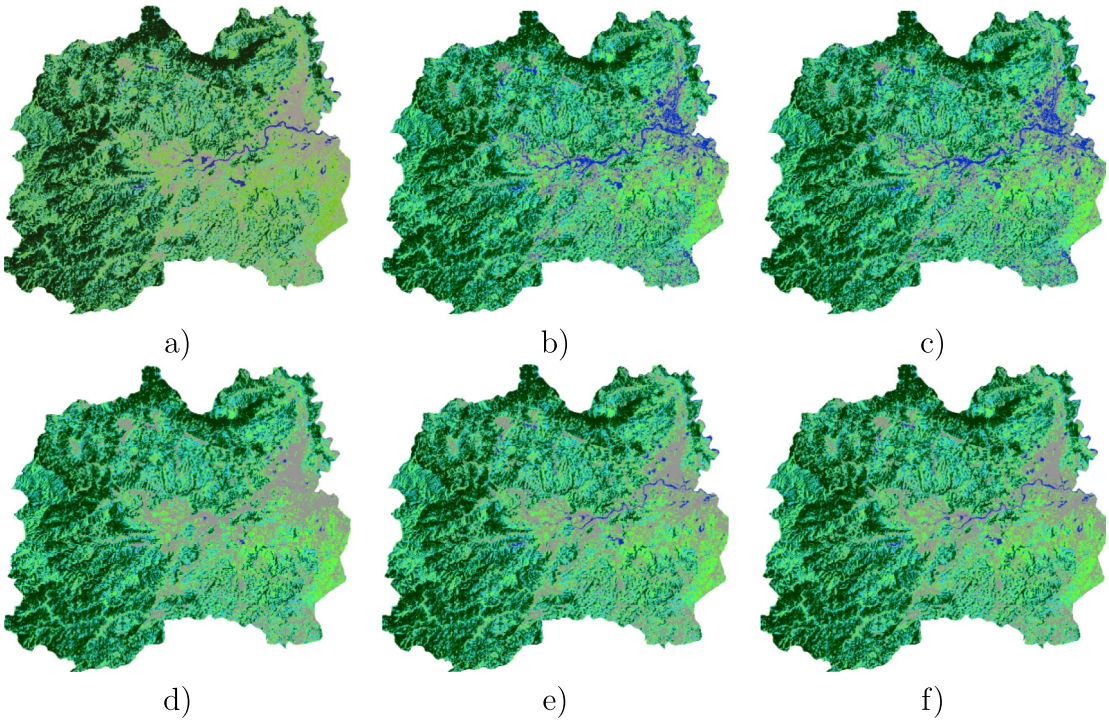


**Fig. 9.** Classification on Quy Hop dataset (a) KIFCM (b) KFCM, (c) CFCM, (d) IIT2-FCM, (e) MKCFCM, (f) SMKCFCM.

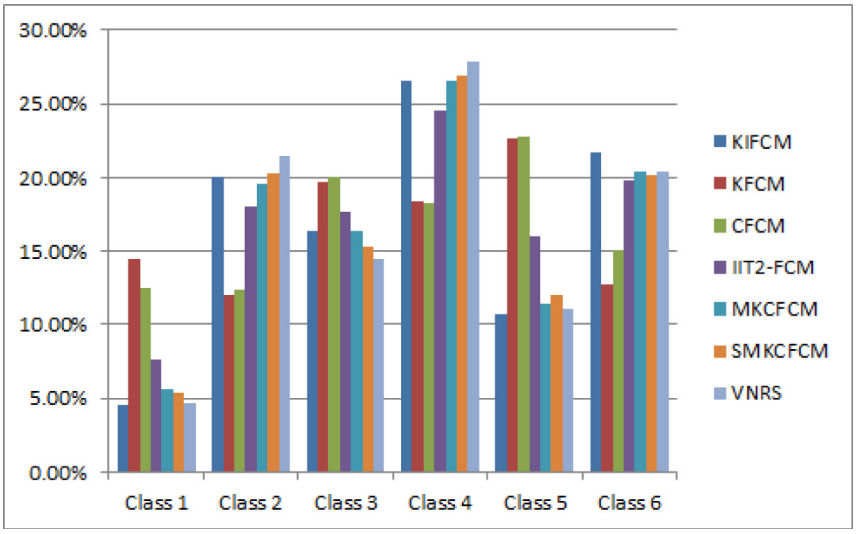


**Fig. 10.** The comparing chart between the classes of Quy Hop dataset.

**Table 6**
Validity Indices (VI) obtained for Quy Hop dataset.

| VI | KIFCM | KFCM | CFCM | IIT2-FCM | MKCFCM | SMKCFCM |
|---|---|---|---|---|---|---|
| FS | 4.3915 | 3.4021 | 3.8762 | 4.8723 | 4.8692 | **4.9071** |
| SSE | 93.2549 | 195.0342 | 178.3985 | 158.7814 | 140.8757 | **121.0687** |
| D-I | 0.8203 | 0.1836 | 0.3215 | 0.6534 | 0.6873 | **0.7216** |
| DB-I | 1.3087 | 4.6573 | 4.2761 | 2.6715 | 1.2313 | **1.1035** |
| PC-I | 0.8935 | 0.4208 | 0.7652 | 0.8802 | 0.8891 | **0.8906** |
| CE-I | 0.5314 | 2.1385 | 1.0963 | 0.4516 | 0.2511 | **0.2403** |
| S-I | 0.4038 | 0.8964 | 0.7965 | 0.4978 | 0.3123 | **0.3017** |
| XB-I | 0.4031 | 0.2872 | 0.2375 | 0.1690 | 0.1588 | **0.1136** |
| CL-I | 0.9367 | 0.7928 | 0.8288 | 0.9469 | **0.9608** | 0.9403 |
| GC-I | 2.9680 | 2.529 | 2.2761 | 1.8268 | 1.4392 | **1.318** |
| Mean | 0.81% | 8.86% | 8.15% | 3.06% | 1.06% | **0.80%** |

**Table 7**
Average of validity indices for Thanh Hoa, Thai Nguyen and Quy Hop datasets.

| VI | KIFCM | KFCM | CFCM | IIT2-FCM | MKCFCM | SMKCFCM |
|---|---|---|---|---|---|---|
| FS | 4.3448 | 3.1111 | 3.4527 | 4.1798 | 4.4826 | **4.6482** |
| SSE | 95.6558 | 134.5079 | 120.4753 | 103.8078 | 95.5031 | **87.3670** |
| D-I | 0.8060 | 0.2839 | 0.3779 | 0.7730 | 0.8096 | **0.8473** |
| DB-I | 1.3554 | 5.3773 | 4.6716 | 15012 | 1.3458 | **1.2470** |
| PC-I | 0.8846 | 0.4484 | 0.7596 | 0.8449 | 0.8898 | **0.9060** |
| CE-I | 0.7192 | 3.1365 | 2.3642 | 0.8501 | 0.7136 | **0.6638** |
| S-I | 0.4136 | 0.8049 | 0.7409 | 0.4211 | 0.3976 | **0.3612** |
| XB-I | 0.2459 | 0.4000 | 0.3474 | 0.1851 | 0.1744 | **0.1592** |
| CL-I | 0.9423 | 0.8094 | 0.8763 | **0.9614** | 0.9344 | 0.9590 |
| GC-I | 2.9076 | 3.8711 | 3.7907 | 3.3475 | 3.0883 | **2.4888** |
| Mean | 1.24% | 5.29% | 6.02% | 2.48% | 1.21% | **0.75%** |

**Table 8**
Computational time by running algorithms KIFCM, KFCM, CFCM, IIT2-FCM, MKCFCM and SMKCFCM on the individual datasets.

| Time (s) | KIFCM | KFCM | IIT2-FCM | CFCM | MKCFCM | SMKCFCM |
|---|---|---|---|---|---|---|
| Thanh Hoa (*3161.304* km$^2$) | 93.1825 | 81.6274 | 98.7263 | 326.7263 | 417.8723 | **216.0247** |
| Thai Nguyen (*6822.165* km$^2$) | 178.0279 | 143.5716 | 196.8772 | | | |
| Quy Hop (*5730.4* km$^2$) | 127.9623 | 98.0254 | 136.7824 | | | |
| The total clustering time | 399.1187 | 323.2244 | 432.3859 | 326.7263 | 417.8723 | **216.0247** |

result is due to the fact that the preprocessing of super pixel SLIC algorithm involves only nearly and similar points, so the accuracy of gathering points into super pixels is better than the general clustering quality for full images and SMKCFCM algorithms inherit this advantage when only super-pixel is clustered instead of clustering all the original pixels.

Besides analyzing the indicators for validating the quality, the computational time is also measured when running different algorithms which is presented in Table 8.

The summaries in Table 8 exhibit the total clustering time of the individual algorithms on all the three datasets or each individual dataset, it can be seen that collaborative fuzzy clustering with super-pixels segmentation has significantly reduced the computational time. When clustering on all the three satellite image datasets, the proposed algorithm takes 216.0247 s, the least value in comparison with the other algorithms (216.0247 s, the least value in comparison with the other algorithms taking 399.1187, 323.2244, 432.3859, 326.7263 and 417.8723 s by the KIFCM, KFCM, IIT2-FCM, CFCM and MKCFCM algorithms, respectively). While the proposed algorithm with weighted super-pixel obtains the equivalent accuracy or the better accuracy compared with the considering algorithms. Therefore the novel framework obtains not only the significant performance of computational time but also the accuracy.

## 5. Conclusions

We have proposed the idea of the collaborative fuzzy clustering by introducing multiple-kernel technique and the weighted super-pixels which helped improve the clustering results running on the satellite images. The proposed methods also enable to overcome drawbacks in the data complexity which data shape is non-spherical and cannot linearly separate. The experiments were implemented for land cover classification problem on multi-spectral satellite image datasets, which exhibited that the SMKCFCM algorithm obtains the better clustering quality than results produced from some considered clustering algorithms.

Several further studies may be concentrated on the usage of evolutionary methods to optimize parameters or implementing on the efficient platform like GPU platform could be investigated.

## Acknowledgement

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

## CRediT authorship contribution statement

**Trong Hop Dang:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & edit. **Dinh Sinh Mai:** Data curation, Formal analysis, Resources, Software, Visualization, Writing - original draft. **Long Thanh Ngo:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing - original draft.

## Appendix. The proof of Theorem 1

To use the technique of Lagrange multipliers, we reformulate the objective function (21) is reformulated in the form:

$$V_{[i_s]} = \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s] d_{ik}^2 + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 d_{ik}^2 - \lambda\Big(\sum_{i=1}^{c} u_{ik}[i_s] - 1\Big) \tag{A.1}$$

The necessary conditions for the minimum of $V[i_s]$ are expressed as $\frac{\partial V}{\partial u_{rs}} = 0$ and $\frac{\partial V}{\partial v_i} = 0$, $r = 1, \ldots, c; s = 1, \ldots, n_{is}$. After calculating the derivative corresponding to the individual elements of the membership matrix, we reach:

$$\frac{\partial V_{[i_s]}}{\partial u_{rs}} = 2\varphi_s u_{rs}[i_s] d_{rs}^2 + 2 \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] u_{rs}[i_s] d_{rs}^2 - 2 \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \tilde{u}_{rs}[i_s|j_s] d_{rs}^2 - \lambda = 0 \tag{A.2}$$

$$u_{rs}[i_s] = \left[ \frac{\sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \tilde{u}_{rs}[i_s|j_s] d_{rs}^2 + \lambda/2}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s]) d_{rs}^2} \right] \tag{A.3}$$

Applying the constraint in the form $\sum_{j=1}^{c} u_{js}[i_s] = 1$, we obtain:

$$\sum_{j=1}^{c} \left[ \frac{\sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \tilde{u}_{js}[i_s|j_s] d_{js}^2 + \lambda/2}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s]) d_{js}^2} \right] = 1 \tag{A.4}$$

$$\sum_{j=1}^{c} \frac{\sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \tilde{u}_{js}[i_s|j_s]}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s])} + \frac{\lambda}{2} \sum_{j=1}^{c} \frac{1}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s]) d_{js}^2} = 1$$

$$\frac{\lambda}{2} = \frac{1 - \sum_{j=1}^{c} \frac{\sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \tilde{u}_{js}[i_s|j_s]}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s])}}{\sum_{j=1}^{c} \frac{1}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s]) d_{js}^2}} \tag{A.5}$$

By plugging (A.5) into (A.3) we have:

$$u_{rs}[i_s] = \frac{\sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \tilde{u}_{rs}[i_s|j_s] d_{rs}^2 + \frac{1 - \sum_{j=1}^{c} \frac{\sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \tilde{u}_{js}[i_s|j_s]}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s])}}{\sum_{j=1}^{c} \frac{1}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s]) d_{js}^2}}}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s]) d_{rs}^2} \tag{A.6}$$

$$u_{rs}[i_s] = \frac{\sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \tilde{u}_{rs}[i_s|j_s]}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s])} + \frac{1}{\sum_{j=1}^{c} \frac{d_{rs}^2}{d_{js}^2}} \left[ 1 - \sum_{j=1}^{c} \frac{\sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \tilde{u}_{js}[i_s|j_s]}{(\varphi_s + \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s])} \right] \tag{A.7}$$

for $s = 1, \ldots, n_{is}$ and $r = 1, \ldots, c$.

To calculate the centroids of clusters we rewrite the objective function as following:

$$J_{[i_s]} = \sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s] (\psi(x_k) - v_i[i_s])^2 + \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 (\psi(x_k) - v_i[i_s])^2 \tag{A.8}$$

$$\frac{\partial J_{[i_s]}}{\partial v_i[i_s]} = 0$$

$$d_{ik}^2 = \psi(x_k)^T\psi(x_k) - 2\frac{\sum_{j=1}^{n_{is}} \varphi_k u_{ij}^2[i_s]\psi(x_k)^T\psi(x_j) + \sum_{j=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2\psi(x_k)^T\psi(x_j)}{\sum_{j=1}^{n_{is}} \varphi_k u_{ik}^2[i_s] + \sum_{j=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2}$$
$$+\left(\frac{\sum_{j1=1}^{n_{is}} \varphi_k u_{ij1}^2[i_s]\psi(x_{j1}) + \sum_{j1=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij1}[i_s]-\tilde{u}_{ij1}[i_s|j_s])^2\psi(x_{j1})}{\sum_{j1=1}^{n_{is}} \varphi_k u_{ik}^2[i_s] + \sum_{j1=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij1}[i_s]-\tilde{u}_{ij1}[i_s|j_s])^2}\right)^T$$
$$\left(\frac{\sum_{j2=1}^{n_{is}} \varphi_k u_{ij2}^2[i_s]\psi(x_{j2}) + \sum_{j2=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij2}[i_s]-\tilde{u}_{ij2}[i_s|j_s])^2\psi(x_{j2})}{\sum_{j2=1}^{n_{is}} \varphi_k u_{ik}^2[i_s] + \sum_{j2=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij2}[i_s]-\tilde{u}_{ij2}[i_s|j_s])^2}\right)$$

$$d_{ik}^2 = \psi(x_k)^T\psi(x_k) - 2\frac{\sum_{j=1}^{n_{is}} \varphi_j u_{ij}^2[i_s]\psi(x_k)^T\psi(x_j) + \sum_{j=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2\psi(x_k)^T\psi(x_j)}{\sum_{j=1}^{n_{is}} \varphi_j u_{ik}^2[i_s] + \sum_{j=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2}$$
$$+\frac{\sum_{j1=1}^{n_{is}}\sum_{j2=1}^{N[i_s]} \varphi_{j1}\varphi_{j2}u_{ij1}^2[i_s]u_{ij2}^2[i_s]\psi(x_{j1})^T\psi(x_{j2}) + 2\sum_{j1=1}^{n_{is}}\sum_{j2=1}^{N[i_s]}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s]u_{ij1}^2(u_{ij2}[i_s]-\tilde{u}_{ij2}[i_s|j_s])^2\psi(x_{j1})^T\psi(x_{j2})}{\left(\sum_{j1=1}^{n_{is}} \varphi_{j1}u_{ik}^2[i_s] + \sum_{j1=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2\right)^2}$$
$$+\frac{\sum_{j_s=1,j_s\neq i_s}^{K}\sum_{j1=1}^{n_{is}}\sum_{j2=1}^{n_{is}} \varphi_{j1}\varphi_{j2}\beta^2[i_s|j_s](u_{ij1}[i_s]-\tilde{u}_{ij1}[i_s|j_s])^2(u_{ij2}[i_s]-\tilde{u}_{ij2}[i_s|j_s])^2\psi(x_{j1})^T\psi(x_{j2})}{\left(\sum_{j1=1}^{n_{is}} \varphi_{j1}u_{ik}^2[i_s] + \sum_{j1=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2\right)^2}$$

**Box II.**

$$d_{ik}^2 = \sum_{t=1}^{M}\omega_t^2P_t(x_k,x_k) - 2\frac{\sum_{j=1}^{n_{is}} \varphi_j u_{ij}^2[i_s]\sum_{t=1}^{M}\omega_t^2P_t(x_k,x_j) + \sum_{j=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2\sum_{t=1}^{M}\omega_t^2P_t(x_k,x_j)}{\sum_{j=1}^{n_{is}} \varphi_j u_{ik}^2[i_s] + \sum_{j=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2}$$
$$+\frac{\sum_{j1=1}^{n_{is}}\sum_{j2=1}^{N[i_s]} \varphi_{j1}\varphi_{j2}u_{ij1}^2[i_s]u_{ij2}^2[i_s]\sum_{t=1}^{M}\omega_t^2P_t(x_{j1},x_{j2}) + 2\sum_{j1=1}^{n_{is}}\sum_{j2=1}^{N[i_s]}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s]u_{ij1}^2(u_{ij2}[i_s]-\tilde{u}_{ij2}[i_s|j_s])^2\sum_{t=1}^{M}\omega_t^2P_t(x_{j1},x_{j2})}{\left(\sum_{j1=1}^{n_{is}} \varphi_{j1}u_{ik}^2[i_s] + \sum_{j1=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2\right)^2} \quad (A.11)$$
$$+\frac{\sum_{j_s=1,j_s\neq i_s}^{K}\sum_{j1=1}^{n_{is}}\sum_{j2=1}^{n_{is}} \varphi_{j1}\varphi_{j2}\beta^2[i_s|j_s](u_{ij1}[i_s]-\tilde{u}_{ij1}[i_s|j_s])^2(u_{ij2}[i_s]-\tilde{u}_{ij2}[i_s|j_s])^2\sum_{t=1}^{M}\omega_t^2P_t(x_{j1},x_{j2})}{\left(\sum_{j1=1}^{n_{is}} \varphi_{j1}u_{ik}^2[i_s] + \sum_{j1=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2\right)^2}$$

**Box III.**

$$\alpha_{ikt} = P_t(x_k,x_k) - 2\frac{\sum_{j=1}^{n_{is}} \varphi_j u_{ij}^2[i_s]P_t(x_k,x_j) + \sum_{j=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2P_t(x_k,x_j)}{\sum_{j=1}^{n_{is}} \varphi_j u_{ik}^2[i_s] + \sum_{j=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2}$$
$$+\frac{\sum_{j1=1}^{n_{is}}\sum_{j2=1}^{N[i_s]} \varphi_{j1}\varphi_{j2}u_{ij1}^2[i_s]u_{ij2}^2[i_s]P_t(x_{j1},x_{j2}) + 2\sum_{j1=1}^{n_{is}}\sum_{j2=1}^{N[i_s]}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s]u_{ij1}^2(u_{ij2}[i_s]-\tilde{u}_{ij2}[i_s|j_s])^2P_t(x_{j1},x_{j2})}{\left(\sum_{j1=1}^{n_{is}} \varphi_{j1}u_{ik}^2[i_s] + \sum_{j1=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2\right)^2} \quad (A.13)$$
$$+\frac{\sum_{j_s=1,j_s\neq i_s}^{K}\sum_{j1=1}^{n_{is}}\sum_{j2=1}^{n_{is}} \varphi_{j1}\varphi_{j2}\beta^2[i_s|j_s](u_{ij1}[i_s]-\tilde{u}_{ij1}[i_s|j_s])^2(u_{ij2}[i_s]-\tilde{u}_{ij2}[i_s|j_s])^2P_t(x_{j1},x_{j2})}{\left(\sum_{j1=1}^{n_{is}} \varphi_{j1}u_{ik}^2[i_s] + \sum_{j1=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ij}[i_s]-\tilde{u}_{ij}[i_s|j_s])^2\right)^2}$$

**Box IV.**

$$\sum_{k=1}^{n_{is}} \varphi_k u_{ik}^2[i_s]\psi(x_k) - \sum_{k=1}^{n_{is}} \varphi_k u_{ik}^2[i_s]v_i[i_s]$$
$$+\sum_{k=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ik}[i_s]-\tilde{u}_{ik}[i_s|j_s])^2\psi(x_k)$$
$$-\sum_{k=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ik}[i_s]-\tilde{u}_{ik}[i_s|[i_s|j_s])^2v_i[i_s] = 0$$

$$v_i[i_s] = \frac{\sum_{k=1}^{n_{is}} \varphi_k u_{ik}^2[i_s]\psi(x_k) + \sum_{k=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ik}[i_s]-\tilde{u}_{ik}[i_s|j_s])^2\psi(x_k)}{\sum_{k=1}^{n_{is}} \varphi_k u_{ik}^2[i_s] + \sum_{k=1}^{n_{is}}\sum_{j_s=1,j_s\neq i_s}^{K} \beta[i_s|j_s](u_{ik}[i_s]-\tilde{u}_{ik}[i_s|j_s])^2} \quad (A.9)$$

We have the distance between the $k$th pattern and centroid of the $i$th cluster: $d_{ik}^2$ is given in Box II and Eq. (A.11) is given in Box III

$$d_{ik}^2 = (\psi(x_k)-v_i)^T(\psi(x_k)-v_i) = \psi(x_k)^T\psi(x_k) - 2\psi(x_k)^Tv_i + v_i^Tv_i \quad (A.10)$$

Eq. (A.11) can be rewritten as follows:

$$d_{ik}^2 = \sum_{t=1}^{M} \alpha_{ikt}\omega_t^2 \tag{A.12}$$

in which $a_{ikt}$ may be described in the following form: Eq. (A.13) is given in Box IV

By using Eq. (A.12) to calculate distance between the $k_{th}$ pattern and centroid of the $i_{th}$ cluster, the objective function (A.1) becomes:

$$J_{[i_s]} = \sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s] \sum_{t=1}^{M} \alpha_{ikt}\omega_t^2 + \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 \sum_{t=1}^{M} \alpha_{ikt}\omega_t^2 \tag{A.14}$$

Subject to $\sum_{t=1}^{M} \omega_t = 1,\ \omega_k \geq 0, \forall k;\ \sum_{j=1}^{c} u_{js}[i_s] = 1, \forall s;\ u_{js}[i_s] \geq 0, \forall s, j.$

The objective function (A.14) can be resolved by using a Lagrange multiplier as follows:

$$J_{[i_s]} = \sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s] \sum_{t=1}^{M} \alpha_{ikt}\omega_t^2 + \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 \times \sum_{t=1}^{M} \alpha_{ikt}\omega_t^2 - 2\lambda(\sum_{t=1}^{M} \omega_t - 1) \tag{A.15}$$

$$\frac{\partial J_{[i_s]}}{\partial \omega_t} = 2\omega_t \sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s]\alpha_{ikt} + 2\omega_t \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 \alpha_{ikt} - 2\lambda = 0$$

$$\omega_t = \frac{\lambda}{\sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s]\alpha_{ikt} + \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 \alpha_{ikt}} \tag{A.16}$$

we know that $\sum_{t=1}^{M} \omega_t = 1$.

$$\sum_{t=1}^{M} \frac{\lambda}{\sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s]\alpha_{ikt} + \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 \alpha_{ikt}} = 1$$

$$\lambda = \frac{1}{\sum_{t=1}^{M} \frac{1}{\sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s]\alpha_{ikt} + \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 \alpha_{ikt}}} \tag{A.17}$$

By plugging (A.17) into (A.16) we have

$$\omega_t = \frac{\frac{1}{\sum_{t=1}^{M} \frac{1}{\sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s]\alpha_{ikt} + \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|j_s])^2 \alpha_{ikt}}}}{\sum_{k=1}^{n_{is}} \sum_{i=1}^{c} \varphi_k u_{ik}^2[i_s]\alpha_{ikt} + \sum_{k=1}^{n_{is}} \sum_{j_s=1, j_s \neq i_s}^{K} \beta[i_s|j_s] \sum_{i=1}^{c} (u_{ik}[i_s] - \tilde{u}_{ik}[i_s|jj])^2 \alpha_{ikt}} \tag{A.18}$$

We have found the formulas (24), (25), (26) and (27) in the formula (A.7) , (A.12), (A.13) and (A.18), i.e. Theorem 1 is proven.

## References


Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S., 2012. Slic superpixels compared to state-of-the-art super-pixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34 (11), 2274–2282.

Anand, S., Mittal, S., Tuzel, O., Meer, P., 2014. Semi-supervised kernel mean shift clustering. IEEE Trans. Pattern Anal. Mach. Intell. 36, 1201–1215.

den Bergh, M.V., Boix, X., Roig, G., de Capitani, B., Van Gool, L., 2012. Seeds: Superpixels extracted via energy-driven sampling. In: Computer Vision, ECCV 2012. Springer.

Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The Fuzzy c-means clustering algorithm. Comput. Geosci. 10 (2–3), 191–203.

Bezdek, J.C., Moshtaghi, M., Runkler, T., 2016. The generalized c index for internal fuzzy cluster validity. IEEE Trans. Fuzzy Syst. 1500–1512.

Chaira, Tamalika., 2011. A novel intuitionistic fuzzy c means clustering algorithm and its application to medical images. Appl. Soft Comput. 17 (2), 1711–1717.

Chaira, Tamalika., 2014. An atanassov's intuitionistic Fuzzy kernel clustering for medical image segmentation. Int. J. Comput. Intell. Syst. 7 (2), 360–370.

Chen, L., Chen, C.L.P., Lu, M., 2012. A multiple-kernel Fuzzy c-means algorithm for image segmentation. IEEE Trans. Syst. Man Cybern. 41 (5), 1263–1274.

Chen, M.Y., Linkens, D.A., 2004. Rule-base generation and simplification for data-driven fuzzy models. Fuzzy Sets and Systems 142, 243–265.

Coletta, L.F.S., Vendramin, L., Hruschka, E.R., Campello, R.J.G.B., Pedrycz, W., 2012. Collaborative Fuzzy clustering algorithms: Some refinements and design guidelines. IEEE Trans. Fuzzy Syst. 20 (3), 444–462.

Dang, T.H., Ngo, L.T., Witold, P., Interval type-2 fuzzy CMeans approach to collaborative clustering. In: The 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015).

Dang, T.H., Ngo, L.T., Witold, P., 2016. Multiple Kernel Based Collaborative Fuzzy Clustering Algorithm. ACIIDS 2016-Springer LNAI.

Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1 (2), 224–227.

Felzenszwalb, P., Huttenlocher, D., 2004. Efficient graph-based image segmentation. Int. J. Comput. Vis. 59 (2), 167–181.

Ganesh, M., Palanisamy, V., 2012. Multiple-kernel Fuzzy c-means algorithm for satellite image segmentation. Eur. J. Sci. Res. 83 (2), 255–263.

Girolami, M., 2002. Mercer kernel-based clustering in feature space. IEEE Trans. Neural Netw. 13, 780–784.

Graves, D., Pedrycz, W., 2010. Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study. Fuzzy Sets and Systems 161, 522–543.

Hofmann, T., Scholkopf, B., Smola, A.J., 2008. Kernel methods in machine learning. Ann. Statist. 36 (3), 1171–1220.

Huang, H.-C., Chuang, Y.-Y., Chen, C.S., 2012. Multiple kernel Fuzzy clustering. IEEE Trans. Fuzzy Syst. 20 (1), 120–134.

Jiang, Y., Chung, F.L., Wang, S., Deng, Z., Wang, J., Qian, P., 2014. Collaborative fuzzy clustering from multiple weighted views. IEEE Trans. Cybernet. 1–13.

Kuo, R.J., Lin, T.C., Zulvia, F.E., Tsai, C.Y., 2018. A hybrid metaheuristic and kernel intuitionistic Fuzzy c-means algorithm for cluster analysis. Appl. Soft Comput. 67, 299–308.

Kuo-Ping, Lin, 2014. A novel evolutionary kernel intuitionistic Fuzzy c-means clustering algorithm. IEEE Trans. Fuzzy Syst. 22 (5), 1074–1087.

Li, Z., Chen, J., 2015. Super-pixel segmentation using linear spectral clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1356–1363.

Liu, M.Y., Tuzel, O., Ramalingam, S., Chellappa, R., 2011. Entropy rate superpixel segmentation. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2097–2104.

Mai, S.D., Ngo, L.T., 2018. Multiple kernel approach to semi-supervised Fuzzy clustering algorithm for land-cover classification. In: Engineering Applications of Artificial Intelligence, Vol. 68. Elsevier, pp. 205–213.

Mitra, S., Banka, H., Pedrycz, W., 2006. Rough–Fuzzy collaborative clustering. IEEE Trans. Syst. Man Cybern. B 36, 795–805.

Ngo, L.T., Mai, D.S., Pedrycz, W., 2015. Semi-supervising interval type-2 Fuzzy C-Means clustering with spatial information for multi-spectral satellite image classification and change detection. Comput. Geosci. 83, 1–16.

Pal, N.R., Sarkar, K., 2014. What and when Can we gain from the kernel versions of c-means algorithm?. IEEE Trans. Fuzzy Syst. 22, 363–379.

Papon, J., Abramov, A., Schoeler, M., Worgotter, F., 2013. Voxel cloud connectivity segmentation-supervoxels for point clouds. In: The IEEE Conference on Computer Vision and Pattern Recognition. pp. 2027–2034.

Pedrycz, W., 2001. Granular computing: an introduction. In: Joint 9th IFSA World Congress and 20th NAFIPS International Conference.

Pedrycz, Witold, 2002. Witold pedrycz collaborative fuzzy clustering. Pattern Recognit. Lett. 23, 1675–1686.

Pedrycz, Witold, 2007. Witold pedrycz collaborative and knowledge based Fuzzy clustering. Int. J. Innovative Comput. Inf. Control 3, 1–12.

Pedrycz, Witold., 2008. Witold pedrycz collaborative clustering with the use of Fuzzy c-means and its quantification. Fuzzy Sets and Systems 2399–2427.

Perbet, F., Maki, A., 2011. Homogeneous super-pixels from random walks. In: MVA. pp. 2–30.

Prasad, M., Li, D.L., Liu, Y.T., Siana, L., Lin, C.T., Saxena, A., 2014. A preprocessed induced partition matrix based collaborative Fuzzy clustering for data analysis. In: IEEE International Conference on Fuzzy Systems. pp. 1553–1558.

Ren, Xiaofeng, Malik, Jitendra, 2003. Learning a Classification Model for Segmentation. In: ICCV '03, vol. 1, pp. 10–17, Nice.

Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22 (8), 888–905.

Sonnenburg, S., Ratsch, G., Schafer, C., Scholkopf, B., 2006. Large scale multiple kernel learning. J. Mach. Learn. Res. 7, 1531–1565.

Veksler, O., Boykov, Y., Mehrani, P., 2010. Super-pixels and supervoxels in an energy optimization framework. In: Computer Vision, ECCV 2010. Springer, pp. 211–224.
Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. Mach. Intell. 6, 583–598.
Wang, W., Zhang, Y., 2007. On fuzzy cluster validity indices. Fuzzy Sets and Systems 158, 2095–2117.
Weikersdorfer, D., Gossow, D., Beetz, M., 2012. Depth-adaptive super-pixels. In: 2012 21st International Conference on Pattern Recognition (ICPR). pp. 2087–2090.
Xiang, D., Tang, T., Hu, C., Li, Y., Su, Y., 2014. A kernel clustering algorithm with Fuzzy factor: Application to SAR image segmentation. IEEE Geosci. Remote Sens. Lett. 11, 1290–1294.
Xie, X.L., Beni, G.A., 1991. A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 13 (8), 841–846.
Xie, Y., Xu, L., Wang, Z., 2014. Automated co-super-pixel generation via graph matching signal. Image Video Process. 8 (4), 753–763.
Xu, L., Zeng, L., Wang, Z., 2014. Saliency-based super-pixels. Signal Image Video Process. 8 (1), 181–190.
Yang, M.S., Tsai, H.S., 2008. A Gaussian kernel-based fuzzy c-means algorithm with a spatial bias correction. Pattern Recognit. Lett. 29, 1713–1725.
Yu, F., Tang, J., Cai, R., 2007. Partially horizontal collaborative Fuzzy c-means. Int. J. Fuzzy Syst. 9 (4), 198–204.
Yu, Fusheng., Yu, Shengli., 2009. Prototypes-based horizontal collaborative Fuzzy clustering. Web Min. Web-based Appl. 66–69.
Zhou, J., Chen, C.L.P., Chen, L., Li, H.X., 2014. Collaborative Fuzzy clustering algorithm in distributed network environments. IEEE Trans. Fuzzy Syst. 22 (6), 1–14.