Remote Sensing Classification Using Fuzzy C-means Clustering with Spatial Constraints Based on Markov Random Field

Yang HongLei1,2, Peng JunHuan1*, Xia BaiRu2 and Zhang DingXuan1

1School of Land Science and Technology, China University of Geosciences, Beijing, 100083, China
2School of Engineering and Technology, China University of Geosciences, Beijing 100083, China
*Corresponding author, e-mail address: pengjunhuan@163.com

Abstract
This paper proposes a new clustering algorithm which integrates Fuzzy C-means clustering with Markov random field (FCM). The density function of the first principal component which sufficiently reflects the class differences and is applied in determining of initial labels for FCM algorithm. Thus, the sensitivity to the random initial values can be avoided. Meanwhile, this algorithm takes into account the spatial correlation information of pixels. The experiments on the synthetic and QuickBird images show that the proposed method can achieve better classification accuracy and visual qualities than the general FCM algorithm.

Keyword: Fuzzy c-means clustering, Markov random field, remote sensing classification, Kernel density function.

Introduction
Fuzzy C-Means clustering algorithm (FCM) was introduced by Ruspini [1969], developed by Dunn [1973], generalized by Bezdek [1981]. The FCM algorithm and its derivatives have been used successfully in many applications, such as pattern recognition, classification, data mining, and image segmentation.

Compared with crisp or hard clustering methods [Pham et al., 1999] which force pixels to belong exclusively to one class, FCM is the fuzzy variant of K-means [Hartigan and Wong, 1979], which allows pixels to belong to multiple clusters with varying degrees of membership. Indeed, FCM is an improvement over K-means when the data set can not clearly subdivide into underlying partitions. Due to the iterative nature, it is very important for an FCM algorithm to choose a good set of initial cluster centers randomly. If a good set of initial cluster centers were chosen, the algorithm may take less iteration to find the actual cluster centers. The pixels on an image are highly correlated, i.e. the pixels in the
Immediate neighborhood possess nearly the same feature data. Therefore, the spatial relation of neighboring pixels is an important characteristic which can be of great aid in image classification. However, the standard FCM only considers the pixel’s spectral information, ignoring the spatial information in image context, which makes it very sensitive to noise, outliers, and other imaging artifacts.

Recently, many researchers have proposed several modified techniques for the FCM algorithm in determining initial centers. Cheng et al. presented a multistage random sampling FCM algorithm to get initial cluster centers by using a series of the full data, which significantly reduced the computation time of partitioning a data set into C class [Cheng et al., 1996]. Hung et al. [2001] proposed a modified FCM called the partition simplification FCM in order to simplify the data set and find an initial candidate set of cluster as close as possible to the actual cluster centers. Kannan [2005] implemented the silhouette method based on cluster center initialization instead of random initialization to improve the segmentation efficiency of the FCM algorithm. Hiren et al. [2003] used the Fuzzy C-Medoids algorithm to select C representative cluster centers for FCM. Hou et al. [2005] inducted Genetic algorithm into FCM by using its searches and parallelism to solve the locality and the sensitivity of the initial condition of FCM. Kuang et al. [2006] proposed the Polynomial Fuzzy C-Mean (PFCM) based on the solving multinomial root to avoid selecting initial centers randomly. Yu et al. [2008] integrated Gustafson-Kessel and Gath-Geva algorithm into FCM to avoid falling into a local minimum. Yang et al. [2010] described a new strategy to determine the cluster number and initial cluster center according to the actual situation of intrusion detection data. Shamsi et al. [2011] developed a specific method by incorporating the spatial neighborhood information to calculate the initial cluster centers to improve the strength of the clusters. Megha et al. [2011] described a novel algorithm by incorporating distribution of the gray level information in the image and a new objective function which ensures better stability and compactness of clusters. To deal with the inhomogeneity problem, many algorithms have been proposed by adding correction steps before classifying the image or by modeling the image as the product of original image and a smooth varying multiplier field. Many researchers have incorporated spatial information into the original FCM algorithm to get better classification results. Tolias and Panas [1998a] proposed a fuzzy rule-based system to impose spatial continuity on FCM, and the other paper [Tolias and Panas, 1998b], they used a small positive constant to modify the membership of the center pixel in a 3×3 window. Pham et al. [1999] modified the objective function in the FCM algorithm to include a multiplier field containing the first-order and second-order information of the image. Similarly, Ahmed et al. [2000] proposed an algorithm to compensate for the intensity inhomogeneity and to label a pixel by considering its immediate neighborhood. Pham [2002] presented an approach to penalize the FCM objective function to constrain the behavior of the membership functions, similar to methods used in the regularization and Markov random field (MRF) theory [Li, 1995]. Chen et al. [2004] proposed a robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. Lung et al. [2009] proposed a Generalized Spatial Fuzzy C-Means (GSFCM) algorithm that utilizes both given pixel attributes and the spatial local information which is weighted correspondingly to neighbor elements based on their distance attributes.
In spite of above methods used different strategies that elevate performance of the FCM algorithm, none of them can provide reasonable initial centers and incorporate spatial information simultaneously. To overcome such drawbacks, we develop a modified FCM algorithm which has the advantages of both of the FCM and MRF, and can eliminate the effects of noise. The algorithm introduces the principal component transformation [Fung and LeDrew, 1987; Xu and Gong, 2007] and MRF [Li, 1995]. The initial class number and initial centers are determined by analyzing the density function of the first principal component which sufficiently reflects the class differences to improve speed of the algorithm. A powerful model for the membership functions that incorporates neighborhood information is given by MRF which is defined through a Gibbs function. The spatial context constraint is introduced into the objective function of FCM to get a modified FCM algorithm. Experimental results on the synthetic and real-world remote sensing images are given to demonstrate the robustness and validity of the proposed algorithm.

The modified FCM algorithm

FCM algorithm

Mathematically, conventional FCM is formulated to minimize the following objective function, the sum of the errors between intensity at every pixel and the centroids of each class, with respect to the membership \( \mu_k(i, j) \) and the centroids \( v_k \).

\[
J = \sum\limits_{i,j} \sum\limits_{k=1}^{C} \mu_k(i,j)^q \|I(i,j) - v_k\|^2 \quad [1]
\]

Where \( I(i,j) \) is the observed intensity at pixel \( (i, j) \), \( (i, j) \) is the set of pixel locations in the image, \( C \) is the number of clusters or classes. The membership functions are constrained to satisfy:

\[
\sum\limits_{k=1}^{C} \mu_k(i, j) = 1 \quad [2]
\]

Here, the objective function [1] is minimized only when high values are assigned to pixels whose intensities are close to the centroids and low values are assigned to pixels whose intensities are distant from the centroids. The parameter \( q \) is the constant parameter that controls the degree of fuzziness of clustering result and satisfies \( q > 1 \). The membership functions become increasingly fuzzy as \( q \) increases.

A pixel \( I(i, j) \) belongs to a specific cluster \( V_k \) that is given by the membership value \( \mu_k(i, j) \) of the pixel to that classification. Local minimization of the object function [1] is accomplished by repeatedly adjusting the values of \( \mu_k(i, j) \) and \( V_k \) according to the following equations:
As the object function [1] is iteratively minimized, \( V_k \) becomes more stable. Iteration is terminated when the termination measurement \( \max \{ v_k \} \leq \epsilon \) is satisfied. Where \( v_k \) is new centers, \( v_k \) is previous centers, and \( \epsilon \) is the predefined termination.

It can easily be seen from the object function [1] that the objective function of FCM does not take into account any spatial information; i.e., classification is solely based on the histogram of image. This limitation will make the FCM algorithm exhibit sensitivity to noise in the observed image.

**MRF Theory**

MRF theory provides a convenient and consistent way to model context dependent constraint through neighborhood system: \( N = \{ N_i, i \in S \} \), where \( S \) is the class label, \( N_i \) is the set of sites neighboring \( i \), and has the properties: (1) \( i \notin N_i \); (2) \( i \in N_j \iff j \in N_i \).

\( X \) is said to be a Markov random field (MRF) on \( S \) with respect to a neighborhood system \( N \) if and only if the following two conditions are satisfied:

\[
P(x) > 0, \forall x \in X \quad [5]
\]

\[
P(x_i | x_{S-\{i\}}) = P(x_i | x_{N_i}) \quad [6]
\]

MRF theory depicts the local characteristics of \( X \): a label interacts with only its neighboring labels. In other words, only neighboring labels have direct interactions with each other. According to the Hammersley-Clifford theorem, an MRF can be equivalently characterized by a Gibbs distribution:

\[
P(x) = \exp(-U(x)) / Z \quad [7]
\]

where \( Z = \sum_{x \in X} \exp(-U(x)) \) is a normalizing constant called the partition function and \( U(x) \) is the energy function. The energy is the sum of clique potentials \( V_c(x) \) over all possible
cliques $U(x) = \sum_{c \in \partial} V_c(x)$ . A clique is defined as a subset of sites in S in which every pair of distinct sites are neighbors. The value of $V_c(x)$ depends on the local configuration on the clique $\partial$.

**The modified FCM algorithm**

Given one pixel $(i, j)$ and the set of its neighboring sites $N_{(i,j)}$, the prior probability $P_k(i, j)$ of labeling the pixel to $k \in C$ can be calculated in term of the Gibbs model, $1 - P_k(i, j)$ can be considered as the resistance of neighbors $N_{(i,j)}$ to assigning pixel $(i, j)$ the label $k$. The resistance is defined as reusable level in this paper. The resistance can be introduced into the objective function [1] of FCM as following:

$$J = \sum_{ij} \sum_{k=1}^{C} (1 - p_k(i, j))(\mu_k(i, j))^q \left\| I(i, j) - v_k \right\|^2 \quad [8]$$

Using Lagrange multipliers to impose the constraint in [2] and evaluating the centers and membership functions that satisfy a zero gradient condition yield the two necessary conditions for [8] to be at a minimum:

$$\mu_k(i, j) = \frac{\left\| I(i, j) - v_k \right\|^2 (1 - p_k(i, j))^2}{\sum_{i=1}^{C} \left\| I(i, j) - v_k \right\|^2 (1 - p_k(i, j))^2} \quad [9]$$

$$v_k = \frac{\sum_{ij} \mu_k(i, j) (1 - p_k(i, j))I(i, j)}{\sum_{ij} \mu_k(i, j)^2 (1 - p_k(i, j))} \quad [10]$$

When minimizing [8], membership values are assigned to the pixel not only according to the distance from the centers, but also taking into account the resistance of the neighboring pixels to the label. Therefore, the proposed algorithm is robust to noise.

In our case, $\partial$ is the set of spatial second-order cliques. The prior $P(x)$ will represent the simple fact that classification should be locally homogeneous. Therefore, in our model, these potentials favor similar classes in neighboring pixels:

$$V_c(x) = \begin{cases} +1, & \text{ if } x_i \neq x, \\ -1, & \text{ otherwise} \end{cases}$$

From equations [7], the full prior can be obtained.
The discrete steps of the modified FCM algorithm are as follows:
1. Initial estimates of the centers and initial segmentation;
2. Calculate the prior probability \( P_k(i, j) \) in term of [7] from the traditional FCM classification result;
3. Update the membership functions and centers according to [9] and [10];

A block scheme describing the modified FCM algorithm is shown in Figure 1.

**Experimental results and analysis**
Several experimental trials were carried out to aim at assessing classification accuracy analysis of the proposed approach. In particular, two data sets corresponding to the noisy synthetic image and the Quick Bird image of Maricopa County, Arizona, USA are applied.

**Synthetic image**
In this section, the synthetic image was used to evaluate the anti-noise performance of the proposed method. The synthetic image with 512×512 pixels included three classes with three
intensity values taken as (55 110 225) was sampled from MRF model using a Gibbs sampler (Li S. Z. et al., 1995). The image was divided into four parts corrupted separately by Gaussian noise (upper left part, mean: 0, variance: 0.01), Poisson noise (upper right part, mean: 140), speckle noise (lower left part, mean: 0, variance: 0.04) and salt pepper noise (lower right part, the noise density: 5%). The synthetic image and the noise image were shown in Figure 2 and Figure 4a.

![Figure 2 - The simulated image.](image)

![Figure 3 - Density function of the contaminated image on Gibbs Sampler (The horizontal axis is gray value and vertical axis is the probability density).](image)

![Figure 4 - (a) The noise image, (b) the FCM classification image, (c) the modified FCM classification image.](image)

Table 1 - Classification accuracy.

<table>
<thead>
<tr>
<th>Classification method</th>
<th>Overall classification accuracy</th>
<th>Kappa coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>The FCM algorithm</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>The modified FCM algorithm</td>
<td>0.98</td>
<td>0.97</td>
</tr>
</tbody>
</table>

In this experiment, the initial class number 3 and initial centers (55 110 225) were selected from Figure 3, which can decrease iteration. The modified FCM algorithm using the given initial centers took 19 iterations to terminate at the given $\varepsilon$ while it took 27 iterations using
initial centers randomly. The classification result of the modified algorithm was superior to that obtained using the FCM algorithm from the Table 1. Figure 4b and Figure 4c showed the classification images by FCM algorithm and the modified FCM algorithm respectively. It can be seen that the result of the modified FCM classification is less speckled and smoother. The most Poisson noise and salt & pepper noise was reduced by the modified algorithm. The classification error is mostly of Gaussian noise, speckle noise and some faint distortion at edges caused by the influence of the penalty function in neighborhoods containing pixels from different classes. The performance study on visual as well as quantitative measure showed that the modified FCM can improve the classification accuracy due to the combined benefits of FCM and MRF model.

**Maricopa Study Area**

In order to assess classification accuracy of the proposed method, the real-world data set was used in the experiment that acquired by the QuickBird satellite imagery. The image (300×300 pixels) was used in UTM coordinate and georeferenced in the WGS84 system, which covers over Maricopa County, Arizona, USA. on Mar 17, 2004, including a multi-spectral image in 4 bands (blue, green, red and near infrared) with the spatial resolution of 2.4 m and a panchromatic image with the resolution of 0.6m. The area mainly consists of a green belt, water, roads, residential buildings and bare land shown in Figure 5a. In this experiment, the images firstly were atmospherically corrected to at-surface reflectance using the ITT ENVI 4.5 FLAASH module. Linear equalization was used to enhance the contrast between objects and background. The panchromatic image was resampled to a spatial resolution of 2.4m in order to combine with multi-spectral images. The class number was fixed at 6 classes (water, building, bare land 1, bare land 2, green belt 1 and green belt 2) and the corresponding initial centers were set to (-1 -0.5 0.3 0.8 1.8 2.2) based on the density function of the first principal component (shown in Fig. 5a). The modified FCM algorithm took 56 seconds to terminate at the given $\varepsilon$ while it took 1 minute 12 seconds using initial centers randomly. Therefore, the proposed method for initial centers can give a good approximation of the final centers, which allowed for less iteration of the modified FCM algorithm and, a faster classification.

![Figure 5 - Density function of Log-Principal Component Transformation, a is the 1st PC, b to d, the 2nd to 4th PC respectively.](image-url)
Figure 6b and 6c shows the classification result of the modified FCM algorithm and the standard FCM, respectively. By visual inspection, the image quality obtained by the standard FCM clustering is poor, with many spurious blobs. This again demonstrates the drawback of failing to incorporate spatial dependencies information into classification scheme. 500 ground truth points were randomly sampled on the reference image (shown in Fig. 6a) without any consideration of informational class distribution to avoid statistical bias, whose true class were assessed by visual interpretation technique. The overall classification accuracy for the proposed algorithm is 82%. There is an enhancement of about 7% of overall classification accuracy relative to the traditional FCM algorithm from Tables 2-3. The image generated by the proposed method also reveals “clean” (especially within class “bare land 1” area) and patch-like visual appearance. However, there are a little blurred at edges within the image.

Figure 6 - (a) Original image, (b) The modified FCM Clustering image, (c) FCM Clustering image.
Table 2 - Classification accuracy (A: Actual class; P: Predicted class).

<table>
<thead>
<tr>
<th></th>
<th>Water</th>
<th>Building</th>
<th>Bare land 1</th>
<th>Greenbelt 1</th>
<th>Bare land 2</th>
<th>Greenbelt 2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>87</td>
</tr>
<tr>
<td>Building</td>
<td>0</td>
<td>50</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>69</td>
</tr>
<tr>
<td>Bare land 1</td>
<td>1</td>
<td>7</td>
<td>79</td>
<td>8</td>
<td>14</td>
<td>4</td>
<td>113</td>
</tr>
<tr>
<td>Greenbelt 1</td>
<td>0</td>
<td>5</td>
<td>9</td>
<td>63</td>
<td>1</td>
<td>14</td>
<td>92</td>
</tr>
<tr>
<td>Bare land 2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>32</td>
<td>7</td>
<td>44</td>
</tr>
<tr>
<td>Greenbelt 2</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>20</td>
<td>4</td>
<td>64</td>
<td>95</td>
</tr>
<tr>
<td>Sum</td>
<td>88</td>
<td>63</td>
<td>107</td>
<td>97</td>
<td>53</td>
<td>92</td>
<td>500</td>
</tr>
</tbody>
</table>

Overall classification accuracy: 75%
Kappa coefficient: 70%

Table 3 - Classification accuracy (A: Actual class; P: Predicted class).

<table>
<thead>
<tr>
<th></th>
<th>Water</th>
<th>Building</th>
<th>Bare land 1</th>
<th>Greenbelt 1</th>
<th>Bare land 2</th>
<th>Greenbelt 2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>87</td>
</tr>
<tr>
<td>Building</td>
<td>0</td>
<td>54</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>69</td>
</tr>
<tr>
<td>Bare land 1</td>
<td>1</td>
<td>5</td>
<td>88</td>
<td>5</td>
<td>11</td>
<td>3</td>
<td>113</td>
</tr>
<tr>
<td>Greenbelt 1</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>71</td>
<td>0</td>
<td>9</td>
<td>92</td>
</tr>
<tr>
<td>Bare land 2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>35</td>
<td>6</td>
<td>44</td>
</tr>
<tr>
<td>Greenbelt 2</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>15</td>
<td>2</td>
<td>73</td>
<td>95</td>
</tr>
<tr>
<td>Sum</td>
<td>88</td>
<td>63</td>
<td>110</td>
<td>96</td>
<td>50</td>
<td>93</td>
<td>500</td>
</tr>
</tbody>
</table>

Overall classification accuracy: 82%
Kappa coefficient: 78%

Conclusion
A new scheme of incorporating spatial information based on introducing MRF into the traditional FCM algorithm has been proposed, which presents some important advantages over the traditional FCM algorithm used in remote sensing applications:
(1) The initial centers of the FCM algorithm are determined by analyzing the density function of the first principal component. In this way, the proposed approach avoids selecting random initial values.
(2) The proposed algorithm adopts the MRF theory accompanied with the traditional FCM algorithm, both spectral and 2D spatial information can be combined.

The results as shown in this study have revealed the success in achieving higher accuracy and more patch-like patterns compared with the traditional FCM algorithm. The simulation experiment verified that the proposed algorithm is effective and more robust to noisy images than the conventional FCM algorithm. The real-world image experiment verified the proposed method yields considerably lower classification error, which each class comprising more homogeneous regions. 56.3% classification errors occurred near edges within the image. Because edges do not conform to the assumption that neighboring pixels
should be similar, the membership value of the actual class is lowered slightly, thereby slightly increasing the membership values for other classes. Moreover, the initial number and initial centers are determined by analyzing the density function of the first principal component to avoid randomly select, which improves the speed of the algorithm, and it is especially significant for high-dimensional spaces and large data sets.

Acknowledgements
This work was supported by the following research projects: the National High-tech R&D Program of China (2007AA12Z226, SS2012AA120804); the National Natural Science Foundation of China (40674015, 41074009); the Doctoral Fund of Ministry of Education of China (20100022110008); the Fundamental Research Funds for the Central Universities (2-9-2011-227); the Open Research Fund of Key Laboratory of Digital Earth Science, Center for Earth Observation and Digital Earth, Chinese Academy of Sciences (2010LDE002). The authors would like to thank anonymous reviewers who gave valuable suggestion that has helped to improve the quality of the manuscript.

References


Kannan A S.R. (2005) - *Segmentation of MRI using new unsupervised fuzzy C-means*

Received 01/06/2012, accepted 20/02/2013