J Indian Soc Remote Sens (March 2014) 42(1):35–49
DOI 10.1007/s12524-013-0296-x

RESEARCH ARTICLE

ISRS

Robust Semi-supervised Kernel-FCM Algorithm Incorporating Local Spatial Information for Remote Sensing Image Classification

Chengjie Zhu · Shizhi Yang · Qiang Zhao · Shengcheng Cui · Nu Wen

Received: 20 February 2013 / Accepted: 14 May 2013 / Published online: 24 July 2013

Abstract Fuzzy c-means (FCM) algorithm is a popular method in image segmentation and image classification. However, the traditional FCM algorithm cannot achieve satisfactory classification results because remote sensing image data are not subjected to Gaussian distribution, contain some types of noise, are nonlinear, and lack labeled data. This paper presents a robust semi-supervised kernel-FCM algorithm incorporating local spatial information (RSSKFCM_S) to solve the aforementioned problems. In the proposed algorithm, insensitivity to noise is enhanced by introducing contextual spatial information. The non-Euclidean structure and the problem in nonlinearity are resolved through kernel methods. Semi-supervised learning technique is utilized to supervise the iterative process to reduce step number and improve classification accuracy. Finally, the performance of the proposed RSSKFCM_S algorithm is tested and compared with several similar approaches. Experimental results for the multispectral remote sensing image show that the RSSKFCM_S algorithm is more effective and efficient.

Keywords Kernel-FCM · Remote sensing image · Image classification · Semi-supervised · Local spatial information

C. Zhu · S. Yang · Q. Zhao · S. Cui · N. Wen
Center of Optical Remote Sensing, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China

C. Zhu
College of Electronic Engineering, Anhui University of Science and Technology, Huainan 232001, China

C. Zhu (✉)
Huaibei, Anhui, China
e-mail: ahhbzcj@126.com

Introduction

Remote sensing image is utilized in various applications of natural resources inventory and management and weather services. Image classification plays a key role in these applications. Many algorithms and approaches were proposed in the past to classify satellite images. These algorithms and approaches have achieved huge progress. Fuzzy c-means (FCM) algorithm is a popular method in image classification because of its structural and computational simplicity. The success of FCM is mainly attributed to the introduction of fuzziness for the belongingness of each image pixel.

Some of the complications of real-image data, such as image context spatial information and nonlinear space, are not considered by standard FCM (Rafael 1997; Liew et al. 2000; Pham 2002; Chen and Zhang 2004; Tamma et al. 2011). For example, the traditional FCM algorithm does not consider the context spatial information of the image, indicating that FCM is very sensitive to noise and outliers. In addition, standard FCM utilizes Euclidean distance to compute the objective function, leading to discontented classification results. Many

authors proposed several approaches and algorithms to improve FCM's capability to resolve the aforementioned problems. For example, Yang et al. (2011) introduced kernel-FCM (KFCM) algorithms with spatial constraints (KFCM_S) for image segmentation. Chen and Zhang (2004) proposed several FCM algorithms with constraints (FCM_S1, FCM_S2, and KFCM_S1) and utilized mean or median filter to incorporate local spatial information with FCM or KFCM in MR image segmentation. Liu's study made use of partial labeled data-supervised iterative process in KFCM (Liu et al. 2008). Cai et al. (2007) integrated local spatial and gray information with FCM in MR image segmentation. The above mentioned approaches have been proven effective to some extent in specific domains.

Remote sensing images actually have the following characteristics: 1) they are generally multispectral or hyperspectral; the nonlinear degree and complication of data are higher than those of gray images; 2) they are not subjected to Gussian distribution; 3) they do not have a single type of noise; they are usually mixed with several types of noise; and 4) they lack labeled data. These characteristics of remote sensing images make the aforementioned algorithms and approaches disadvantageous, thereby reducing their classification accuracy and efficiency.

Considering the characteristics of remote sensing images, this study proposes a robust semi-supervised KFCM algorithm incorporating local spatial information (RSSKFCM_S). The proposed algorithm aims to

1) utilize FCM to eliminate data fuzziness
2) transform low-dimensional nonlinear input space into a high-dimensional linear feature space to resolve complex nonlinear problems linearly with the well-known kernel theorem.
3) supervise the iterative process with few labeled data; this procedure can improve classification accuracy and reduce step number and computation time
4) enhance insensitivity to noise by incorporating local spatial information.

Traditional FCM and KFCM Algorithm

FCM Algorithm

FCM was first introduced by J. C. Bezdek in 1981. In the algorithm, $X=\{x_1,x_2\cdots x_n\}\subset R^S$ denotes an observed monochromatic intensity field where n is the number of data, R^S is the vector space, and x_k is the value of a pixel at position k.

Mathematically, the standard FCM objective function for partitioning a dataset X into clusters is

$$\begin{cases} J_m=\min\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m d_{ik}^2=\min\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m\|x_k-v_i\|^2 \\ \text{s.t.}\sum_{i=1}^{c}u_{ik}=1,\ \ 0\le u_{ik}\le 1,\ \ i=1,\cdots,\text{c},k=1,\cdots,n \end{cases} \quad (1)$$

In the above equation, U is membership function and u_{ik} describes the degree to which x_k belongs to class i. Following bi-element logic, membership has two values, 0 or 1, indicating that x_k is either a member or not a member of class i. Fuzzy logic allows membership to have multiple values; thus, x_k can belong to any class to some degree. The last constraint guarantees that U is nondegenerate; thus, the resulting partition is nontrivial. c is the number of class. m is the weighted exponent that controls the fuzziness of the membership function, $1\le m\le\infty$. m is usually between 2 and 4.5 in most of practical applications (Pal and Bezdek 1995). d_{ik} is the distance measure from x_k to cluster center v_i. The value of d_{ik} is calculated by Euclidean distance.

The FCM algorithm searches for optimal partition U and optimal prototype V by minimizing an objective function subject to the constraints on U, which can be solved through the method of Lagrange multipliers as follows:

$$J_m=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^m\|x_k-v_i\|^2+\sum_{k=1}^{n}\lambda_k\left(1-\sum_{i=1}^{c}u_{ik}\right) \quad (2)$$

where λ_k s are Lagrange multipliers. We set $\frac{\partial}{\partial u_{ik}}J_{\min}=0$ and $\frac{\partial}{\partial v_i}J_{\min}=0$.

We derive and gain

$$u_{ik}=1/\sum_{j=1}^{c}\left(d_{ik}/d_{jk}\right)^{2/m-1} \quad i=1,\cdots,c\ ,\ \ k=1,\cdots,n \quad (3)$$

$$v_i=\sum_{k=1}^{N}u_{ik}^m x_j/\sum_{k=1}^{N}u_{ik}^m \quad i=1,\cdots,c\,. \quad (4)$$

With Eqs. [3] and [4], U and V can be updated step by step from an arbitrary partition until the iteration process converges.

KFCM Algorithm

Kernel method is one of the most researched subjects in machine-learning domain and has been widely applied in pattern recognition and function approximation. Kernel method involves the transformation of low-dimensional nonlinear input space into a high-dimensional linear feature space to linearly treat and solve complex nonlinear problems in the transformed space. Unlike FCM, KFCM utilizes kernel-induced distance instead of Euclidean distance; hence, KFCM is more suitable for nonlinear data than FCM (Liu et al. 2008).

We suppose that $x \in X \subseteq R^l \mapsto \Phi(x) \in F \subseteq R^h (l << h)$ is a nonlinear low-dimensional feature space transformed into linear high-dimensional feature space (Chen and Zhang 2004) where $x=[x_1,x_2]^T$, $\Phi(x) = \left[x_1^2, \sqrt{2}x_1x_2, x_2^2\right]^T$.

In F space,

$$\Phi(x)^T\Phi(y) = \left[x_1^2, \sqrt{2}x_1x_2, x_2^2\right]^T \left[y_1^2, \sqrt{2}y_1y_2, y_2^2\right] = \left(x^Ty\right)^2 = K(x,y).$$

In the above formula, $K(x,y)$ is the kernel function. Several typical kernel functions are available, such as, Gaussian kernel (RBF): $K(x,y)=\exp(-\|x-y\|^2/\sigma^2)$, Sigmoid kernel: $K(x,y)=\tanh[\beta<x\cdot y>+\gamma]$; and polynomial kernel: $K(x,y)=(x^Ty+1)^d$. For the RBF kernel, $K(x,x)=1$.

$$\begin{aligned}\text{Thus},\quad d(x,y) &= \|\Phi(x)-\Phi(y)\| \\ &= \sqrt{(\Phi(x)-\Phi(y))^T(\Phi(x)-\Phi(y))} \\ &= \sqrt{\Phi(x)^T\Phi(x)-\Phi(x)^T\Phi(y)-\Phi(y)^T\Phi(x)+\Phi(y)^T\Phi(y)} \\ &= \sqrt{K(x,x)+K(y,y)-2K(x,y)} \\ &= \sqrt{2-2K(x,y)}.\end{aligned}$$

The objective function is then changed to

$$\begin{aligned}J_{Km} &= \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \|\Phi(x_k)-\Phi(v_i)\|^2 \\ &= 2\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m (1-K(x_k,v_i)).\end{aligned} \tag{5}$$

The conditions of restriction are similar to those in the FCM algorithm. J_{Km} can be minimized under such conditions, and U and V can be obtained through iteration until the maximal iteration step or terminative condition is reached as follows:

$$u_{ik} = (1-K(x_k,v_i))^{-1/(m-1)} \Big/ \sum_{j=1}^{c}\left(1-K\left(x_k,v_j\right)\right)^{-1/(m-1)} \quad i=1,\cdots,c \quad k=1,\cdots,n. \tag{6}$$

$$v_i = \sum_{k=1}^{N} u_{ik}^m K(x_k,v_i)x_k \Big/ \sum_{k=1}^{N} u_{ik}^m K(x_k,v_i) \quad i=1,\cdots,c \tag{7}$$

Equation [7] shows that V lies in the original space and is endowed with additional weight $K(x_k,v_i)$; thus, computational simplicity of KFCM is similar to that of FCM. $K(x_k,v_i)$ in the RBF formula measures the similarity between x_k and v_i. However, when the distance between the two is large, $K(x_k,v_i)$ is small and the weighted sum of the data points is suppressed and insensitive to noise and outliers.

FCM and KFCM can be summarized in the following steps:

Step 1: Fix c, m, T_{max}, ε, σ, and initialize prototypes $V_{(0)}$ and $U_{(0)}$ under restriction conditions; $t=0$.
Step 2: Calculate $K(x_k,v_i)$ for KFCM with RBF.
Step 3: Update prototype matrix $V_{(t)}$ through Eq. [5] for FCM and Eq. [7] for KFCM.
Step 4: Calculate objective function J_m for FCM and J_{Km} for KFCM.
Step 5: Update fuzzy partition matrix $U_{(t)}$ through Eq. [4] for FCM and Eq. [6] for FCM.
Step 6: If $\|u_{ik(t+1)}-u_{ik(t)}\| \le \varepsilon$ or $t=T_{\max}$, then stop; otherwise, go back to step 2, $t=t+1$.

The final step in FCM and KFCM is to interpret U, which is usually a hard cluster, by classifying pixel x_k to class i with the largest u_{ik}.

FCM and KFCM Algorithms with Local Spatial Information

FCM with Local Spatial Information

Ahmed et al. (2002) proposed an approach to eliminate noise by modifying the objective function by introducing

a term that allows the labeling of a pixel to be influenced by the labels in its neighborhood as follows:

$$J_{sm} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\|x_k - v_i\|^2 + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\sum_{r\in N_k}\|x_r - v_i\|^2 \tag{8}$$

where N_k stands for the set of neighbors existing in a window around x_k and N_R is the cardinality of N_k. The effect of the neighbors is controlled by parameter α. The last term in the formula denotes the effect of local spatial information and aims to maintain continuity in the neighboring pixel around x_k. J_{sm} can also be minimized while being subjected to restriction conditions. U and V are obtained with the following formula:

$$u_{ik} = \frac{\left(\|x_k - v_i\|^2 + \frac{\alpha}{N_R}\sum_{r\in N_k}\|x_r - v_i\|^2\right)^{-1/(m-1)}}{\sum_{j=1}^{c}\left(\|x_k - v_j\|^2 + \frac{\alpha}{N_R}\sum_{r\in N_k}\|x_r - v_j\|^2\right)^{-1/(m-1)}} \quad i = 1,\cdots,c \quad k = 1,\cdots,n \tag{9}$$

$$v_i = \sum_{k=1}^{n} u_{ik}^{m}\left(x_k + \frac{\alpha}{N_R}\sum_{r\in N_k} x_r\right) \Big/ \left((1+\alpha)\sum_{k=1}^{n} u_{ik}^{m}\right) \quad i = 1,\cdots,c \tag{10}$$

Chen and Zhang (2004) utilized $\|\bar{x} - v_i\|^2$ to replace $\frac{\alpha}{N_R}\sum_{r\in N_k}\|x_r - v_i\|^2$ because computing the neighbor- hood terms consumes more time in KFCM than in FCM. $\bar{x}$ can be computed in advance and only once. Chen utilized mean and median filter to compute $\bar{x}$ (referred to as FCM_S1 and FCM_S2, respectively) and modified the objective function as follows:

$$J_{s12m} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\|x_k - v_i\|^2 + \alpha\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\left\|\overline{x_k} - v_i\right\|^2 \tag{11}$$

Similarly, U and V are obtained with the following solution by minimizing J_{s12m}.

$$u_{ik} = \frac{\left(\|x_k - v_i\|^2 + \alpha\left\|\overline{x_k} - v_i\right\|^2\right)^{-1/(m-1)}}{\sum_{j=1}^{c}\left(\|x_k - v_j\|^2 + \alpha\left\|\overline{x_k} - v_j\right\|^2\right)^{-1/(m-1)}} \quad i = 1,\cdots,c \quad k = 1,\cdots,n \tag{12}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m}\left(x_k + \alpha\overline{x_k}\right)}{(1+\alpha)\sum_{k=1}^{n} u_{ik}^{m}} \quad i = 1,\cdots,c \tag{13}$$

This formula is advantageous because it reduces execution time and improves robustness to Gaussian noise or salt and pepper noise. The center pixel and its neighbors have the same prototypes or segmentation results, thereby guaranteeing pixel homogeneity.

KFCM with Local Spatial Information

Similar to the derivant of KFCM, Chen et al. kernelize the objective function of FCM_S and obtain the following objective function with kernel-induced distance instead of Euclidean distance (KFCM_S).

$$J_{ksm} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}(1 - K(x_k, v_i)) + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\sum_{r\in N_k}(1 - K(x_r, v_i)) \tag{14}$$

where $K(x_k, v_i)$ is RBF. The parameters are consistent with those in FCM_S.

The following iterative formula for U and V is obtained by minimizing J_{ksm}.

$$u_{ik} = \frac{\left((1 - K(x_k, v_i)) + \frac{\alpha}{N_R}\sum_{r\in N_k}(1 - K(x_k, v_i))^m\right)^{-1/(m-1)}}{\sum_{j=1}^{c}\left((1 - K(x_k, v_j)) + \frac{\alpha}{N_R}\sum_{r\in N_k}(1 - K(x_k, v_j))^m\right)^{-1/(m-1)}} \quad i = 1,\cdots,c \quad k = 1,\cdots,n \tag{15}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m}\left(K(x_k, v_i) + \frac{\alpha}{N_R}\sum_{r\in N_k} K(x_k, v_i)\right)x_k}{\sum_{k=1}^{n} u_{ik}^{m}\left(K(x_k, v_i) + \frac{\alpha}{N_R}\sum_{r\in N_k} K(x_k, v_i)\right)} \quad i = 1,\cdots,c \tag{16}$$

Similarly, the kernelized objective function for FCM_S1 and FCM_S2 (refereed to as KFCM_S1 and KFCM_S2, respectively) is

$$J_{ks12m} = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m}(1-K(x_k, v_i)) + \alpha \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m}\left(1-K\left(\overline{x_k}, v_i\right)\right). \quad (17)$$

The objective function above is minimized with the following iteration functions:

$$u_{ik} = \frac{\left((1-K(x_k, v_i)) + \alpha\left(1-K\left(\overline{x_k}, v_i\right)\right)\right)^{-1/(m-1)}}{\sum_{j=1}^{c}\left((1-K(x_k, v_j)) + \alpha\left(1-K\left(\overline{x_k}, v_j\right)\right)\right)^{-1/(m-1)}} \quad i = 1, \cdots, c \quad k = 1, \cdots, n \quad (18)$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m}\left(K(x_k, v_i)x_k + \alpha K\left(\overline{x_k}, v_i\right)\overline{x_k}\right)}{\sum_{k=1}^{n} u_{ik}^{m}\left(K(x_k, v_i) + \alpha K\left(\overline{x_k}, v_i\right)\right)} \quad i = 1, \cdots, c \quad (19)$$

The procedures for FCM and KFCM calculation with local spatial information can be summarized as follows:

Step 1: Fix c, m, T_{max}, ε, σ, β, and N_R; initialize prototypes $V_{(0)}$ and $U_{(0)}$ under the restriction conditions, set t=0, and compute the mean or median filtered image.

Step 2: Compute $K(x_k, v_i)$ for KFCM_S, and compute $K(\overline{x}_k, v_i)$ of the mean or median filtered image with RBF for KFCM_S1 and KFCM_S2.

Step 3: Update V through Eq. [10] for FCM_S, Eq. [13] for FCM_S1 and FCM_S2, Eq. [16] for KFCM_S, and Eq. [19] for KFCM_S1 and KFCM_S2.

Step 4: Calculate objective function J_{sm} for FCM_S, J_{ksm} for KFCM_S, J_{s12m} for FCM_S1 and FCM_S1, and J_{s12m} for KFCM_S1 and KFCM_S2.

Step 5: Update U through Eq. [9] for FCM_S, Eq. [12] for FCM_S1 and FCM_S2, Eq. [15] for KFCM_S, and Eq. [18] for KFCM_S1 and KFCM_S2.

Step 6: If $\|u_{ik(t+1)} - u_{ik(t)}\| \le \varepsilon$ or $t = T_{\max}$, then stop; otherwise go back to step 2, $t = t+1$.

Robust Semi-supervised KFCM Algorithm with Local Spatial Information

Robust FCM and KFCM Algorithm with Local Spatial Information (RFCM_S and RKFCM_S)

The approaches presented in "FCM and KFCM Algorithms with Local Spatial Information" have a common parameter α, which controls the influence of the local spatial neighbors. The value of α has a crucial impact on the performance of the approaches. Setting the appropriate value is important to obtain relatively optimal classification image. However, the value of α is difficult to set because information on noise, such as its type and intensity, is unavailable. The value of α should be large enough to eliminate noise. At the same time, the value should be small enough to preserve sharpness and details. The value of α is generally set based on experience or by trial-and-error experiments.

Cai et al. (2007) modified the expression of local spatial information in the objective function and a novel spatial function is defined to eliminate difficulty in the selection of α and to improve image classification performance.

$$s_k = \sum_{j\in N_k} s_{kj}x_j \Big/ \sum_{j\in N_k} s_{kj} \quad (20)$$

$$s_{kj} = \exp\left(\frac{-\|x_j - x_k\|^2}{\beta\cdot\lambda_k}\right) \quad (21)$$

$$\lambda_k = \frac{1}{N_R}\sum_{j\in N_i}\|x_j - x_k\|^2 \quad (22)$$

where x_k is the center of the local spatial window (window size is 3×3), x_j is the jth local spatial neighbors falling into the window around x_k, N_k is the set of all the neighbors, and N_R is the cardinality of N_k. λ_k is the mean of the distance between the neighbors of x_k and x_k. The value of λ_k reflects the homogeneity degree of the local spatial window. The smaller theλ_k value is, the more homogenous the local spatial window is. However, setting the different weighted values for

different pixels within the window can mitigate the influence of noise to some extent; the value of α in all the approaches presented in "FCM and KFCM Algorithms with Local Spatial InformationFCM and KFCM Algorithms with Local Spatial Information" is similar in every pixel. In this study, the values ofλ_k are different and can be computed in advance. s_{kj} reflects the damping extent in the pixel value. Based on the definition of s_{kj}, s_{kj} s can change automatically with different pixel values. Furthermore, the closer the value of the jth neighbors of x_k with the value of x_k, the larger s_{kj} is. β, the scale factor in s_{kj} function formula, determines the change characteristic. β is robustly insensitive to change in the value of s_{kj} to some extent because of the geometrical implication of s_{kj}; hence, selectingβ is much easier than selecting parameter α.

s_k is the weighted means of the homogeneity degree of the neighbors of x_k. Such direct transformation provides the following benefits: (1) the transformation makes the algorithm relatively independent of prior knowledge on noise; (2) the transformation can automatically be determined based on local spatial information relationship rather than by artificial or empirical selection; and (3) the transformation provides different weighted values to the local spatial neighbors of every pixel according to homogeneity degree. The influence of noise can be mitigated to some extent.

In our algorithm, the objective function is

$$J_{Rsm} = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\|s_k - v_i\|^2. \tag{23}$$

U and V are obtained by minimizing J_{Rsm}.

$$u_{ik} = \left(\|s_k - v_i\|^2\right)^{-1/(m-1)} \Big/ \sum_{j=1}^{c}\left(\|s_k - v_j\|^2\right)^{-1/(m-1)} \quad i = 1,\cdots,c \quad k = 1,\cdots,n \tag{24}$$

$$v_i = \sum_{k=1}^{n} u_{ik}^{m} s_k \Big/ \sum_{k=1}^{n} u_{ik}^{m} \quad i = 1,\cdots,c \tag{25}$$

The kernelized objective function of RFCM_S (RKFCM_S) is

$$J_{Rksm} = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m}(1 - K(s_k, v_i)). \tag{26}$$

This objective function is minimized through the following iterations:

$$u_{ik} = (1-K(s_k, v_i))^{-1/(m-1)} \Big/ \sum_{j=1}^{c}\left(1-K\left(s_k, v_j\right)\right)^{-1/(m-1)} \quad i = 1,\cdots,c \quad k = 1,\cdots,n \tag{27}$$

$$v_i = \sum_{k=1}^{n} u_{ik}^{m} K(s_k, v_i) s_k \Big/ \sum_{k=1}^{n} u_{ik}^{m} K(s_k, v_i) \quad i = 1,\cdots,c \tag{28}$$

Robust FCM and KFCM algorithms with local spatial information can be summarized as follows:

Step 1: c, m, T_{max}, ε, σ, β, and N_R; initialize prototypes $U_{(0)}$ and $V_{(0)}$ under the restriction conditions, set t=0, and computeλ_k, s_{kj}, and s_k.

Step 2: For RKFCM_S, compute $K(s_k, v_i)$ with RBF.

Step 3: Update V with Eq. [25] for RFCM_S and with Eq. [28] for RKFCM_S.

Step 4: Calculate objective function J_{Rsm} for RFCM_S and J_{Rksm} for RKFCM_S.

Step 5: Update U with Eq. [24] for RFCM_S and with Eq. [27] for RKFCM_S.

Step 6: If $\|u_{ik(t+1)} - u_{ik(t)}\| \le \varepsilon$ or $t = T_{\max}$, then stop; otherwise, go back to step2, t=t+1.

Robust Semi-supervised KFCM Algorithm with Local Spatial Information (RSSKFCM_S)

Remote sensing image classification is a challenging task in real-world applications because image data are nonlinear and fuzzy, are not subject to Gaussian distribution, mingled with a few types of noise, and lack labeled

Fig. 1 Original image

data. A robust semi-supervised KFCM algorithm incorporating local spatial information (RSSKFCM_S) is proposed in this study to improve the performance of remote sensing image classification methods.

In practical applications that involve large-sized image data, only few data can be labeled because the labeling process is usually costly, time consuming, and requires an experienced human annotator. Given that labeled training data are expensive, unlabeled data are readily available in large quantities (Bensaid et al. 1992). Semi-supervised method, which is learned from both labeled and unlabeled data, is a trade-off (Bouchachia and Pedrycz 2006). FCM and its derivants are unsupervised during clustering. If semi-supervised learning technique is introduced to the algorithms in "FCM and KFCM Algorithms with Local Spatial Information", some of the known classes can be utilized as labeled data to supervise the iterative process. Thus, semi-supervised FCM or KFCM is useful in remote sensing image classification where minimal prior knowledge and few labeled data are available. In this study, we introduce the semi-supervised learning technique to RFCM_S and

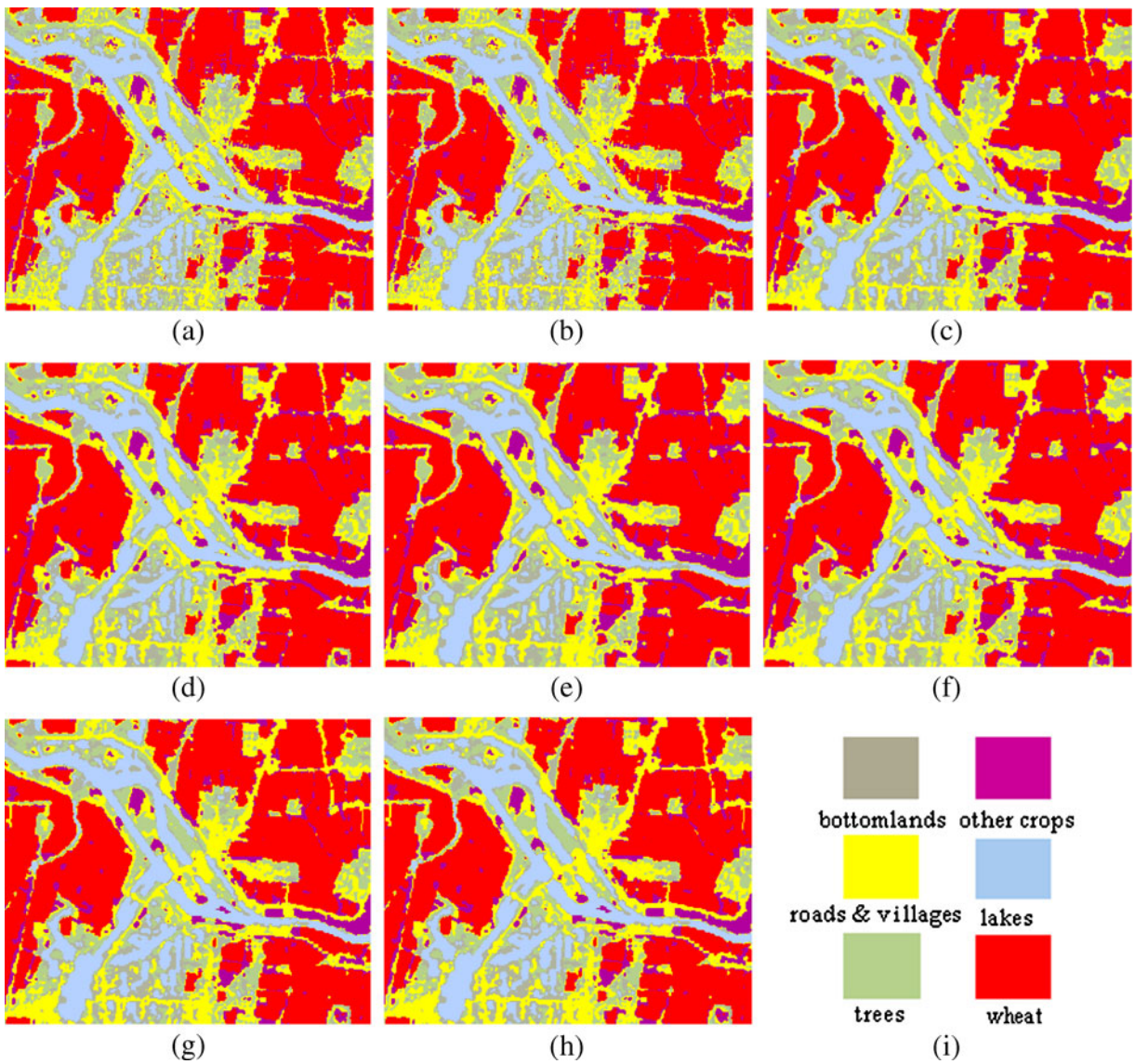

Fig. 2 Comparison of the eight algorithms results for an actual image. **a** FCM, **b** KFCM, **c** RFCM_S, **d** RKFCM_S, **e** KFCM_S1, **f** KFCM_S2, **g** RSSFCM_S, and **h** RSSKFCM_S. **i** Color map board of each class

RKFCM_S and referred to them as RSSFCM_S and RSSKFCM_S, respectively.

In the algorithm, given the data set $X=\{X_l^{nl},X_u^{nu}\}$ where X_l^{nl} is the set of labeled data, X_u^{nu} is the set of unlabeled data, nl is the number of the labeled data, and nu is the number of the unlabeled data, generally, $nl<<nu$. Fuzzy partition matrix is set to $U=\{U_l^{nl},U_u^{nu}\}$ where U_l^{nl} and U_u^{nu} are partition matrices corresponding to X_l^{nl} and X_u^{nu}, respectively. X_l^{nl} is the hard-partition matrix, and its element value is set to 1 or 0. X_u^{nu} is the fuzzy-partition matrix, and its values are set the one of {0, 1} (Liu et al. 2008).

In minimizing the objective function, the set of U_l^{nl} is utilized to aid and train the set of U_u^{nu}. U_l^{nl} holds the line. $U_{(0)}=\{U_l^{nl},U_u^{nu}{}_{(0)}\}$, $V_{(0)}$ is the clustering center of the set of X_l^{nl}. The algorithm can be summarized as follows:

Step 1: Fix c, m, T_{max}, ε, σ, β, and N_R; set $t=0$, initialize prototypes $V_{(0)}$ and $U_{(0)}$ under the restriction conditions, and compute λ_k , s_{kj}, and s_k.

Step 2: For RSSKFCM_S, compute $K(s_k, v_i)$ with RBF.

Step 3: Update V with Eq. [25] for RSSFCM_S an with Eq. [28] for RSSKFCM_S.

Step 4: Calculate objective function J_{Rsm} for RSSFCM_S and J_{Rksm} for RSSKFCM_S.

Step 5: Update U with Eq. [24] for RSSFCM_S and with Eq. [27] for RSSKFCM_S, and then modify $U_{(t)}=\{U_l^{nl},U_u^{nu}{}_{(t)}\}$.

Step 6: If $\|u_{ik(t+1)}-u_{ik(t)}\|\leq\varepsilon$ or $t=T_{\max}$, then stop; otherwise, go back to step 2, $t=t+1$.

The process of defuzzification is similar to that in the FCM algorithm; hence, the maximum membership rule is utilized, namely, pixel x_k. is classified to class i with the largest u_{ik}.

Experimental Results and Analysis

The experiment is performed with a multispectral remote sensing image acquired by the Japanese ALOS satellite AVNIR-2 sensor in March 2009. The data is employed to analyze the statistics of the wheat area in Huaibei plain. Winter wheat is mainly grown in the region. Mid-March is the returning green stage of winter wheat, which is a typical wheat phenology in the region. Figure 1 shows the original image, which is a false color composite of three bands: band 2 (520 nm to 600 nm, green), band 3 (610 nm to 690 nm, red), and band 4 (760 nm to 890 nm, infrared). Given that the image pixel is set to $X=\{x_1,x_2,\cdots x_n\}\subset R^3$, the number of pixels is 251×301, the area of experiment image data is E115°2′39.57″–N33°0′24.5″ to E115°4′ 26.20″–N33°0′53.9″. Six types of classes exist, including wheat, trees, roads and villages, lakes, other crops, and bottomlands. The computer hardware workbench consists of Pentium dual core (2.01 and 2.0 GHz),

Table 1 Comparison of the evaluating indexes of the eight algorithms for Fig. 1(a)

Class	FCM	KFCM	RFCM_S	RKFCM_S	KFCM_S1	KFCM_S2	RSSFCM_S	RSSKFCM_S
Roads & villages	75.5 %	80.7 %	77.8 %	77.8 %	78.2 %	79.2 %	88.2 %	87.2 %
Wheat	98.5 %	98.5 %	98.6 %	98.6 %	97.8 %	98.3 %	98.7 %	98.8 %
Trees	73.1 %	79.4 %	77.0 %	79.6 %	83.7 %	83.5 %	86.1 %	87.7 %
Other crops	81.6 %	81.8 %	87.8 %	87.5 %	89.1 %	86.7 %	88.4 %	88.1 %
Lakes	97.8 %	97.8 %	99.3 %	99.6 %	99.7 %	99.5 %	99.1 %	99.3 %
Bottomlands	59.7 %	60.2 %	58.4 %	60.2 %	56.1 %	64.2 %	65.1 %	69.7 %
Roads & villages	76.9 %	82.1 %	86.1 %	88.7 %	90.9 %	90.9 %	89.1 %	90.1 %
Wheat	95.0 %	94.8 %	97.5 %	97.4 %	98.1 %	97.3 %	97.3 %	97.2 %
Trees	65.2 %	72.0 %	69.1 %	70.6 %	69.0 %	72.5 %	79.9 %	84.9 %
Other crops	88.9 %	88.9 %	91.5 %	91.8 %	89.5 %	90.6 %	89.1 %	89.6 %
Lakes	91.5 %	91.5 %	85.2 %	86.3 %	84.4 %	87.8 %	92.7 %	91.9 %
Bottomlands	88.3 %	88.3 %	92.2 %	89.1 %	90.4 %	90.7 %	95.0 %	96.3 %
	87.6 %	89.0 %	88.8 %	89.4 %	89.1 %	90.2 %	92.2 %	92.7 %
	71.2 %	73.8 %	74.0 %	75.0 %	74.5 %	76.6 %	80.2 %	81.8 %
	84.1 %	85.9 %	85.6 %	86.4 %	86.0 %	87.5 %	90.0 %	90.7 %

2 GB memory, Windows XP operating system, and Matlab R2009a software.

The effectiveness and efficiency of the eight algorithms, namely, FCM, KFCM, KFCM_S1, KFCM_S2, RFCM_S, RKFCM_S, RSSFCM_S, and RSSKFCM_S, are compared based on the above image. In the succeeding experiments, three types of noise are utilized to test the performance of classification, anti-jamming, and de-noising, respectively, to be incorporated in the images. These noises include the salt and pepper noise, Gaussian noise, and mixed noise. Mixed noise usually occurs in many practical applications and is a mixture of Gaussian and impulsive noises. Mixed noise distribution follows

$P=(1-\eta)G+\eta S$ (Cai et al. 2007), where G is Gaussian noise with zero mean ($\mu=0$) and variance of σ_G^2, and S is $S\alpha S$ distribution, namely, symmetric α-stable (including abuse parameter α ($0<\alpha\leq 2$)), location parameter θ ($\theta\in(0,\infty)$), dispersion parameter σ ($\sigma>0$), and index of symmetryβ ($-1\leq\beta\leq 1$)) in which $\theta=0$ and

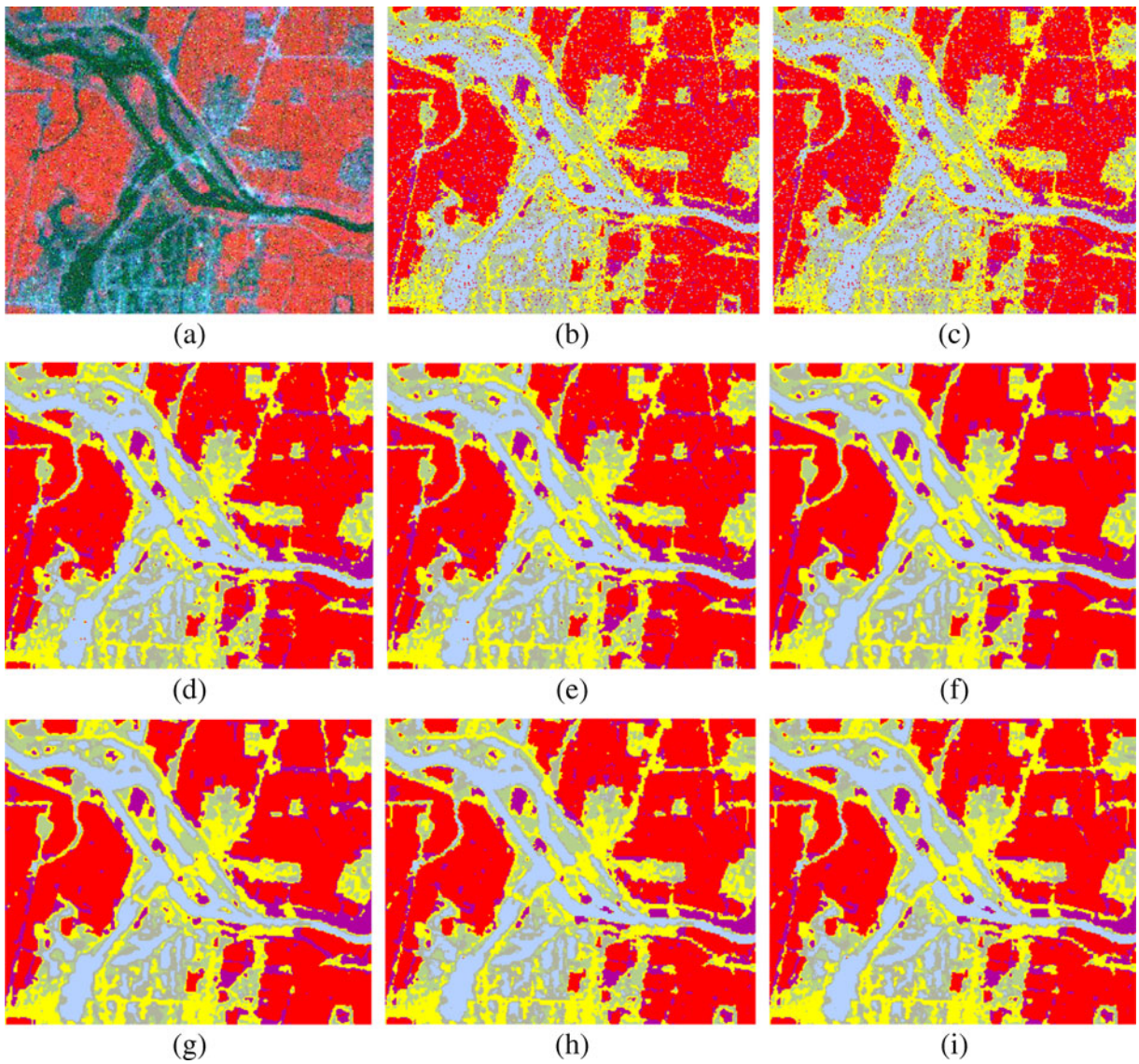

Fig. 3 Comparison of the eight algorithm results for an actual image with 5 % Gaussian noise. **a** Image with 5 % Gaussian noise; **b** FCM; **c** KFCM; **d** RFCM_S; **e** RKFCM_S; **f** KFCM_S1; **g** KFCM_S2; **h** RSSFCM_S; and **i** RSSKFCM_S

$\sigma=1$ (Xiaorui 2008). Although no uniform function of P exists, the characteristic function $\varphi(t)$ of P can be formulated as follows:

$$\phi(t) = \exp\left(j\eta\theta t - (1-\eta)^2 \frac{\sigma_G}{2} t^2 - \eta^\alpha \sigma |t|^\alpha\right) \quad (29)$$

In the formula, η is set to 2/ $(2+\pi)$ for the succeeding experiments (Ben Hamza and Krim 2001).

Evaluating Index System

Confusion matrix is selected to compare the accuracy of the eight approaches. Confusion matrix shows the accuracy of a classification result by comparing a classification result with ground truth information of the region of interests (ROIs). The ROIs do not include the labeled data set in the semi-supervised algorithms. The ROIs of the

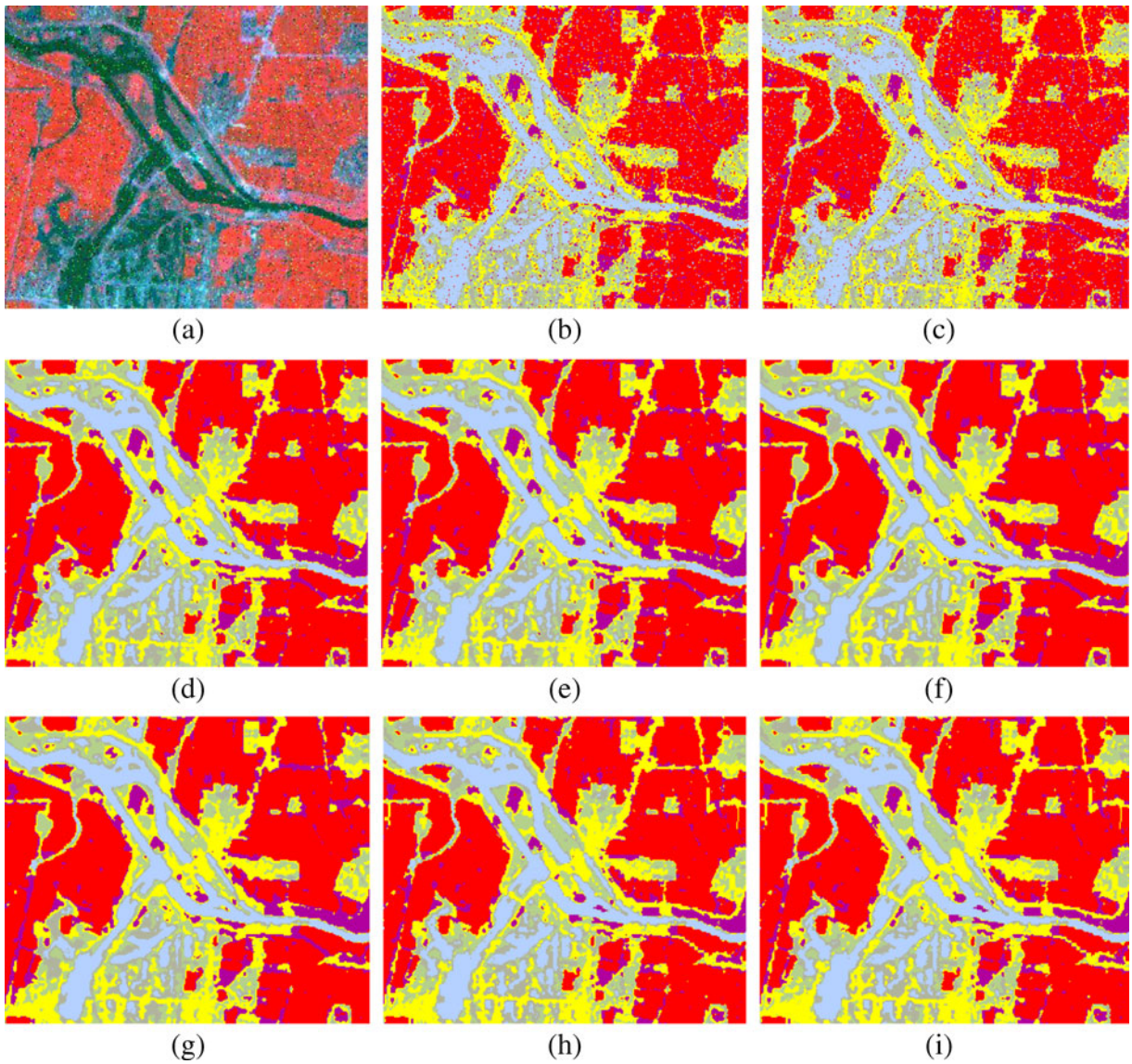

Fig. 4 Comparison of the eight algorithm results for an actual image with 9 % salt and pepper noise. **a** Image with 9 % salt and pepper noise; **b** FCM; **c** KFCM; **d** RFCM_S; **e** RKFCM_S; **f** KFCM_S1; **g** KFCM_S2; **h** RSSFCM_S; and **i** RSSKFCM_S

experiments is selected as follows: the number of roads and villages, other crops, wheat, lakes, trees, and bottomland is 1,187, 855, 2,850, 2,500, 870, and 540, respectively.

The evaluating index system is based on the confusion matrix. The evaluating indexes include user accuracy (UA), mapping accuracy (MA), overall accuracy (OA), kappa coefficient (PC), and comparison scores (CS), where CS is defined through the following formula (Chen and Zhang 2004; Cai et al. 2007):

$$S_{ij} = M_{ij} \cap M_{\text{int}} / M_{ij} \cup M_{\text{int}}. \tag{30}$$

In the formula, M_{ij} denotes the set of pixels belonging to the jth class found by the ith algorithm and M_{int} represents the set of pixels belonging to the ith class in the ROIs.

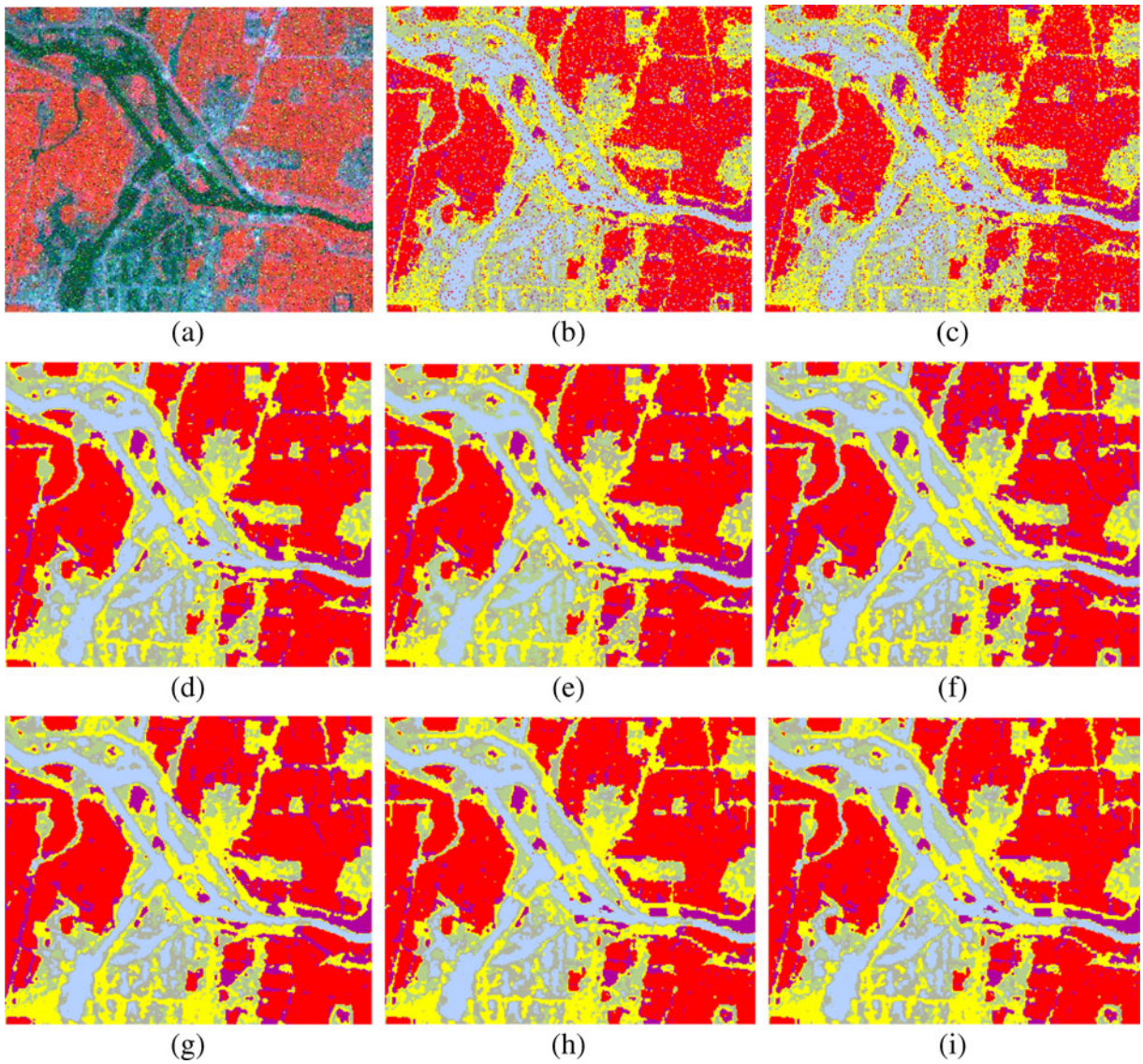

Fig. 5 Comparison of the eight algorithm results for an actual image with mixed noise (α=0.5). **a** Image with mixed noise (α=0.5); **b** FCM; **c** KFCM; **d** RFCM_S; **e** RKFCM_S; **f** KFCM_S1; **g** KFCM_S2; **h** RSSFCM_S; and **i** RSSKFCM_S

Table 2 Comparison of the evaluating indexes of the eight algorithms for Figs. 3(a), 4(a) and 5(a)

Evaluating index	FCM	KFCM	RFCM_S	RKFCM_S	KFCM_S1	KFCM_S2	RSSFCM_S	RSSKFCM_S
OA	80.0 %	78.6 %	87.6 %	87.5 %	87.4 %	89.8 %	89.7 %	90.7 %
CS	59.7 %	58.0 %	71.7 %	71.8 %	71.5 %	75.5 %	77.0 %	77.8 %
PC	74.5 %	72.9 %	84.0 %	84.0 %	83.9 %	86.9 %	87.1 %	88.1 %
OA	75.1 %	74.1 %	87.4 %	87.3 %	83.9 %	89.6 %	88.7 %	89.4 %
CS	53.9 %	52.8 %	71.6 %	71.6 %	65.7 %	74.9 %	74.9 %	75.3 %
PC	68.4 %	67.2 %	84.0 %	83.8 %	79.5 %	86.6 %	85.4 %	86.4 %
OA	75.2 %	74.6 %	86.3 %	86.6 %	85.2 %	86.8 %	89.0 %	89.1 %
CS	53.9 %	53.6 %	69.3 %	70.2 %	67.2 %	70.5 %	74.0 %	74.6 %
PC	69.7 %	68.1 %	82.5 %	82.9 %	81.2 %	83.1 %	85.4 %	86.0 %

Experimental Results and Comparison

The first experiment is performed to test the performance of classification as shown in Fig. 1. In Fig. 2(i) is the color map board of each class, and (a) to (h) is the classification results of the eight algorithms.

Trial and error techniques are adopted to select the parameters of the eight algorithms. The parameters are c=6, m=2, T_{max}=100, $\varepsilon=10^{-4}$, and N_R=8.

In the kernel function of all experiments throughout this paper, parameter σ=200. In the methods incorporating local information, α=3.2 and β=1.2. In the semi-supervised methods, the labeled data number of roads and villages, other crops, wheat, lakes, trees, and bottomland is 478, 427, 1,346, 757, 372, and 223, respectively.

Combined with ground information, the pictures show that the semi-supervised methods are better than the unsupervised methods. The kernel versions are better than Euclidean distance methods. RFCM_S, RKFCM_S, KFCM_S1, and KFCM _S2 are better than FCM and KFCM because neither of the latter utilized spatial information in both algorithms. RSSFCM_S and RSSKFCM _S are better than KFCM_S1, KFCM_S2, RFCM _S, and RKFCM _S because a few labeled data supervise the classification process. Furthermore, RSSKFCM_S achieves the best classification results. The results can be verified by the evaluating indexes. Table 1 shows the comparison of the evaluating indexes of the eight methods in Fig. 1.

The improvement of anti-jamming performance of these algorithms is also discussed. Three types of noises are incorporated into the above image, namely, the image is artificially corrupted by mixed noise, Gaussian noise, and salt and pepper noise. Figures 3, 4 and 5 show the comparison of the classification results for the image.

Figures 3(b to i), 4(b to i) and 5(b to i) show the classification results from applying the eight algorithms in Figs. 3(a), 4(a) and 5(a), respectively. Without loss of

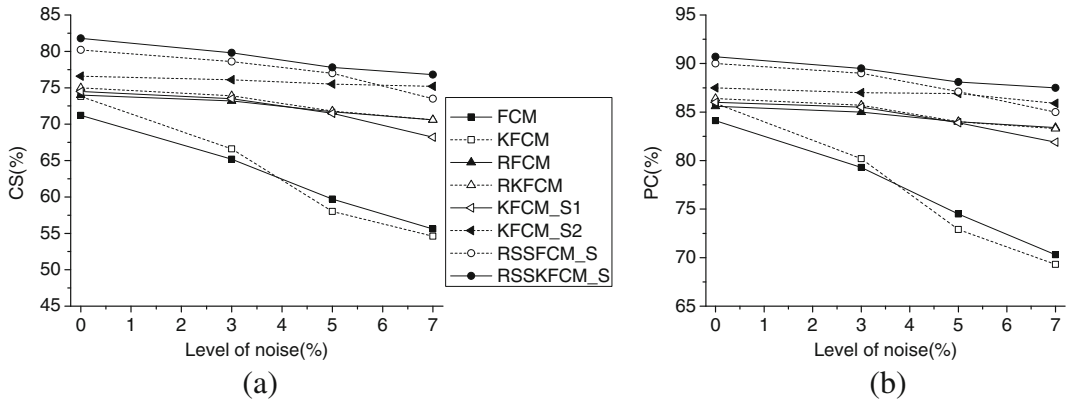

Fig. 6 Comparison of the ten methods in Fig. 1(a) under different levels of Gaussian noise

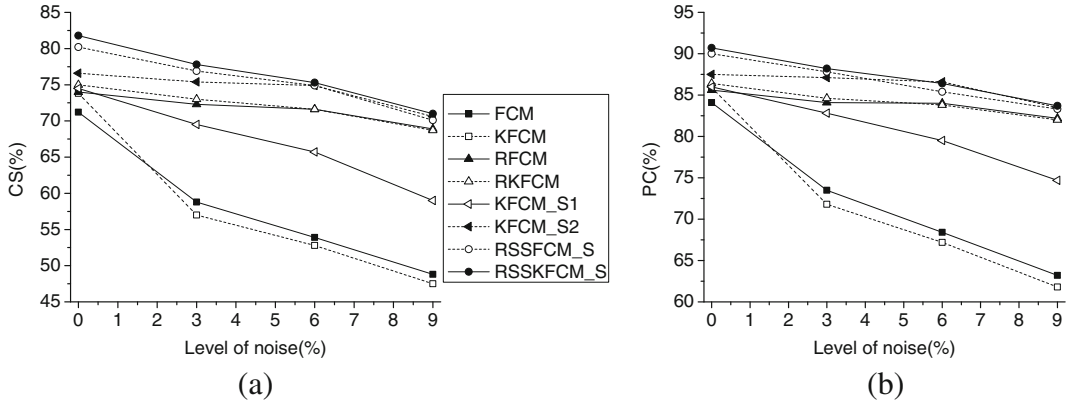

Fig. 7 Comparison of the ten methods in Fig. 1(a) under different levels of salt and pepper noise

generality, a moderate level of noise is added to the image, where the level of Gaussian noise is 5 %, the level of salt and pepper noise is 9 %, and mixed noise parameter α is 0.5.

Table 2 shows the comparison of the main evaluating indexes of the eight algorithms for Figs. 3(b), 4(b) and 5(b). The indexes show overall accuracy (OA), average comparison scores (CS), and Kappa coefficient (PC).

Figures 3, 4 and 5 and Table 2 show that FCM and KFCM exhibit poor performance in the presence of any type of noise. When mixed noise is added, KFCM_S2 performs better than RFCM_S and RKFCM_S. RFCM_S and RKFCM_S perform better than KFCM_S1, and comparatively speaking, RSSFCM_S and RSSKFCM_S provide favorable results. RSSKFCM_S is the best algorithm. When Gaussian noise is added, KFCM_S2 performs better than RFCM_S and RKFCM_S. RSSFCM_S and RSSKFCM_S provide satisfactory classification results. RSSKFCM_S is the best algorithm. When salt and pepper noise is added, the semi- supervised version algorithms of these eight algorithms exhibit efficient performance, where RSSKFCM_S and RSSFCM_S are tolerant of impulse noise and KFCM_S2 has robust anti-jamming performance. Overall, RSSKFCM_S provides favorable classification results and robust anti-jamming performance for the three types of noise.

The last experiment is performed to test the effect of the different levels of the three types of noise added to Fig. 1(a). Figures 6, 7 and 8 show the relationship curves for Kappa coefficient as well as the average comparison scores and the level of noise. Table 3 presents the values of the main evaluating index with different noises and concentrations. The pictures and Table 3 show that the performance of all the methods is gradually reduced with the increase in the level of Gaussian and salt and pepper noise but gradually increases with the increase in the value of parameter $\alpha(0<\alpha<1)$ under mixed noise. Overall, the proposed RSSKFCM_S algorithm is more flat than the others, and its anti-noising performance is the best.

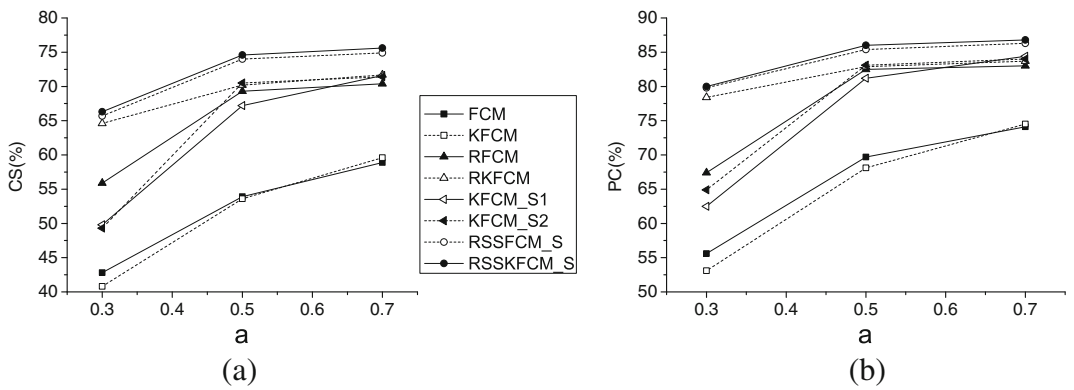

Fig. 8 Comparison of the ten methods in Fig. 1(a) under different levels of mixed noise

Table 3 Comparison of the eight algorithms under different levels of noise corresponding to Figs. 6, 7 and 8

Evaluating index	FCM	KFCM	RFCM_S	RKFCM_S	KFCM_S1	KFCM_S2	RSSFCM_S	RSSKFCM_S
CS	65.2 %	66.6 %	74.2 %	73.9 %	73.5 %	77.1 %	78.6 %	79.8 %
PC	79.3 %	80.2 %	86.0 %	85.7 %	85.5 %	88.0 %	89.0 %	89.5 %
CS	59.7 %	58.0 %	71.7 %	71.8 %	71.5 %	75.5 %	77.0 %	77.8 %
PC	74.5 %	72.9 %	84.0 %	84.0 %	83.9 %	86.9 %	87.1 %	88.1 %
CS	55.6 %	54.6 %	70.6 %	70.6 %	68.2 %	75.2 %	73.5 %	76.8 %
PC	70.3 %	69.3 %	83.4 %	83.3 %	81.9 %	86.9 %	85.0 %	87.5 %
CS	58.8 %	57.0 %	72.3 %	73.0 %	69.5 %	75.4 %	76.9 %	77.8 %
PC	73.5 %	71.8 %	84.1 %	84.6 %	82.8 %	87.1 %	87.8 %	88.2 %
CS	53.9 %	52.8 %	71.6 %	71.6 %	65.7 %	74.9 %	74.9 %	75.3 %
PC	68.4 %	67.2 %	84.0 %	83.8 %	79.5 %	86.6 %	85.4 %	86.4 %
CS	48.8 %	47.5 %	68.9 %	68.7 %	59.0 %	70.6 %	70.1 %	71.0 %
PC	63.2 %	61.8 %	82.2 %	82.0 %	74.7 %	83.5 %	83.3 %	83.7 %
CS	42.8 %	40.8 %	55.9 %	64.6 %	49.8 %	49.3 %	65.7 %	66.3 %
PC	55.6 %	53.1 %	67.4 %	78.4 %	62.5 %	64.9 %	79.8 %	80.0 %
CS	53.9 %	53.6 %	69.3 %	70.2 %	67.2 %	70.5 %	74.0 %	74.6 %
PC	69.7 %	68.1 %	82.5 %	82.9 %	81.2 %	83.1 %	85.4 %	86.0 %
CS	58.9 %	59.6 %	70.4 %	71.7 %	71.6 %	71.4 %	74.9 %	75.6 %
PC	74.1 %	74.5 %	83.0 %	83.7 %	84.4 %	84.0 %	86.3 %	86.8 %

Conclusion

A novel robust FCM framework for multispectral remote sensing image classification was presented in this paper. Spatial constraint and kernel method in the objective function of conventional FCM were employed with both labeled and unlabeled data to effectively classify images and improve anti-jamming performance under different types of noise. With kernel method, a kernel-based FCM algorithm can address nonlinear separation problems better than a Euclidean-based algorithm. Incorporating local spatial information with a novel function into the algorithm, namely, utilizing weighted means of local neighbors to replace the value of the pixel, allows the method to become more robust to noise compared with algorithms without local information. Partially labeled data utilized to train and supervise the iterative process can significantly improve classification accuracy. The convergent speed of the objective function is also further accelerated.

An actual multispectral remote sensing image was utilized to verify the performance of RSSKFCM_S, including its classification and anti-jamming accuracy. The experimental results show that the proposed algorithm is more efficient in multispectral remote sensing image classification and more robust to noise than algorithms without kernel substitution and lacking labeled data. Therefore, the proposed method can be utilized to improve the performance of hyperspectral remote sensing images.

Acknowledgments The authors would like to thank the anonymous referees for their helpful comments and suggestions to improve the presentation of the paper. Special thanks should be given to Dr. Banglong Pan and his colleagues, for providing the remote sensing image data for validation of our algorithm. This research was supported by Anhui Provincial Natural Science Foundation under Grant 1308085QD70.

References

Ahmed, M. N., Yamany, S. M., Mohamed, N., Farag, A. A., Moriarty, T. (2002). A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging, 21*, 193–199.

Ben Hamza, A., & Krim, H. (2001). Image denoising: a nonlinear robust statistical approach. *IEEE Transactions on Signal Processing, 49*, 3045–3053.

Bensaid, A. M., Bezdek, J. C., & Hall, L. O. (1992). Partially supervised fuzzy c-means algorithm for segmentation of MR

images. *Proceedings of SPIE, Science of Artificial Neural Networks, 1710*, 522–528.

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. NewYork: Plenum Press.

Bouchachia, A., & Pedrycz, W. (2006). Enhancement of fuzzy clustering by mechanisms of partial supervision. *Fuzzy Sets and Systems, 157*, 1733–1759.

Cai, W., Chen, S., & Zhang, D. (2007). Fast and robust fuzzy C-means clustering algorithms incorporating local information for image segmentation. *Pattern Recognition, 40*, 825–838.

Chen, S., & Zhang, D. (2004). Robust image segmentation using FCM with spatial constraints based on new Kernel-induced distance measure. *IEEE Transactions on Systems Man and Cybernetics Part B, 34*, 1907–1916.

Liew, A. W. C., Leung, S. H., & Lau, W. H. (2000). Fuzzy image clustering incorporating spatial continuity. *Institute of Electrical Engineers Vision, Image and Signal Processing, 147*, 185–192.

Liu, X., He, B., & Li, X. (2008). Semi-supervised classification for hyperspectral remote sensing image based on PCA and kernel FCM algorithm. *Proceedings of SPIE GIS and Built Environment, 7147*, 1–10.

Pal, N. R., & Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems, 3*, 370–379.

Pham, D. L. (2002). Fuzzy clustering with spatial constraints. In IEEE Proc. Int. Conf. Image Processing, NewYork, pp. II-65–II-68.

Rafael, W. (1997). Unsupervised fuzzy classification of multi-spectral imagery using spatial-spectral features. 21st Annual Meeting of the Gesellschaft fur Klassikation, GfKl'97, Potsdam.

Tamma, R., Rao, T. C. M., & Jaisankar, G. (2011). An efficient method for joint spatial and spectral classification of remote sensed images. *International Journal of Computer Science and Telecommunications, 2*, 262–265.

Xiaorui, Lv. (2008). Model simulation and parameter estimation of alpha stable distribution. Master Dissertation, Huazhong University of Science & Technology, China.

Yang, Y., Guo, S-X., Ren, R-Z., Yu, Y-L. (2011). Modified kernel-based fuzzy C-means algorithm with spatial information for image segmentation. *Journal of Jilin University (Engineering and Technology Edition), 41*, 283–287.