Automatic Change Detection in High-Resolution Remote Sensing Images by Using a Multiple Classifier System and Spectral–Spatial Features

Kun Tan, *Member, IEEE*, Xiao Jin, Antonio Plaza, *Fellow, IEEE*, Xuesong Wang, Liang Xiao, *Member, IEEE*, and Peijun Du

***Abstract*—Change detection (CD) is an active research topic in remote sensing applications including urban studies, disaster assessment, and deforestation monitoring. In this paper, we propose an automatic method for CD in high-resolution remote sensing images that uses a novel strategy for the selection of training samples and an ensemble of multiple classifiers. As for the selection of training samples, our proposed method uses two groups of thresholds instead of just one threshold to enhance the quality of the selected training samples by allowing for their selection in an intelligent manner. In order to achieve high CD accuracy, spatial information such as texture and morphological profiles are utilized in conjunction with spectral information. Our multiple classifier system (MCS) exploits the extreme learning machine (ELM), multinomial logistic regression (MLR), and K-nearest neighbor (KNN) classifiers. To validate our newly proposed approach, we conduct experiments using multispectral images collected by ZY-3. The proposed method provides state-of-the-art CD accuracies as compared with other approaches widely used in the literature for CD purposes.**

***Index Terms*—Change detection (CD), extreme learning machines (ELMs), multiple classifier, spatial information.**

I. Introduction

THE DETECTION of changes occurred over a region of interest is a challenging goal for Earth observation applications. In remote sensing, change detection (CD) generally focuses on extracting change information by analyzing multitemporal images of the same geographical area [1]. CD techniques are often applied in natural disaster evaluation, urban sprawl assessment, and environmental monitoring [2]–[5].

According to the availability of the training samples, CD methods can be broadly grouped into two categories: 1) supervised [6]–[10] and 2) unsupervised [11]–[14]. Supervised CD methods exploit the learning capacity of classifiers to extract change information, and are generally robust to different atmospheric conditions and acquisition dates. In turn, supervised techniques generally require ground-truth information, which limits their application and reduces the degree of automation. Supervised CD methods include postclassification comparison [15], [16] and direct classification [9], [17], [18]. Postclassification comparison identifies class transitions by comparing multiple maps, which are widely applied to Landsat imagery and VHR imagery. However, the accuracy of change map is entirely dependent on the quality of the input maps. Direct classification just requires one classification stage and provides an effective framework to mine a complicated time series [19]. On the contrary, unsupervised methods can obtain change information without any other additional information [17]. This makes these methods widely used such as CVA, thresholding (Otsu algorithm [20] and expectation maximization (EM) algorithm [21]), and Markov random field-based methods [22], [23]. This paper focuses on the direct classification approach.

To improve the accuracy of supervised CD techniques, a new set of methods for sample selection was proposed by Huo *et al.* [18]. These methods identify a threshold to select training samples based on change vector analysis (CVA). Cao *et al.* also used this method to select training samples for automatic CD in high-resolution remote sensing images [9]. However, the samples are often concentrated in one single area and the number of samples for different classes is the same, which does not meet the actual distribution of pixels in different classes of the scene. Moreover, with the rich spatial resolution now available in remote sensing images, CD has been notably developed for high- and very high-resolution (VHR) images in recent years. Normally, in VHR images spectral information is scarce and spatial information is very rich [24]. As a result, making full use of the spatial information has become an exploitation goal for these data. The most commonly used methods for extracting information from VHR images is object-based image analysis [2], [25], [26]. However, this approach

Manuscript received October 25, 2015; revised February 29, 2016; accepted March 07, 2016. Date of publication May 02, 2016; date of current version August 24, 2016. This work was supported in part by the Natural Science Foundation of China under Grant 41471356), in part by the Fundamental Research Funds for the Central Universities (2014ZDPY14), and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions and Jiangsu Key Laboratory of Spectral Imaging and Intelligence Sense.

K. Tan and X. Jin are with Jiangsu Key Laboratory of Resources and Environment Information Engineering, and School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China.

A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politecnica, University of Exremadura, Cáceres 10003, Spain.

X. Wang is with the School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China.

L. Xiao is with the Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping, and Geoinformation of China, Nanjing University, Nanjing 210094, China.

P. Du is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210023, China (e-mail: tankuncu@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2016.2541678

heavily depends on the accuracy of the initial segmentation [9], [19], [27]. Another important method is the fusion of multiple features. Murray *et al.* extracted texture features with the gray level co-occurrence matrix (GLCM) and combined them with spectral features to classify vegetation groups by using high-resolution IKONOS imagery [28]. Tuia *et al.* employed morphological features and support vector machines (SVMs) to investigate the relevance of morphological operators for classification performed on QuickBird panchromatic images [29]. Volpi *et al.* stacked spectral features, texture features, and morphological features to perform supervised CD in VHR images [10]. Above all, these studies prove that the inclusion of texture features and morphological features can compensate for the lack of detailed spectral information.

On the other hand, traditional supervised CD methods usually utilize one classifier to extract change information [9], [10], [18]. But a single classifier cannot detect all kinds of changes that may happen in the image effectively. To address this issue, ensemble learning has been applied to the research and application of CD and classification. Ensemble learning is to solve the same problem by using multiple classifiers (also called base classifiers), which can effectively improve its generalization. In general, the construction of a classifier ensemble consists of two stages: 1) the generation of base classifiers and 2) the combination of their outputs. For the first stage, one possible strategy relies on using the same training data and different learning algorithms to build the base classifiers. Another approach is to build a set of base classifiers by using a single learning method and different training sets [30]. The main issue of this approach is the conversion of the original training set to obtain different training sets. Commonly used methods for this purpose are bagging [31]–[33], boosting [34], [35], and random subspace [36]. The second step is to combine the outputs of base classifiers to provide a final decision. There are a number of different methods, such as weighting methods [37], [38], probabilistic methods [39], evidential reasoning-based methods [40], and meta-learning methods [41], which can be used for this purpose. Many studies have proved that ensemble learning can improve the accuracy of CD and classification. For example, Roy *et al.* used ensemble learning for CD purposes [17], showing that CD using semisupervised multiple classifier systems (MCSs) generally performs better than other state-of-the-art techniques; Yang *et al.* constructed three MCSs for remote sensing image classification and the results show that MCS is superior to base classifiers [42].

In order to avoid the problems related with traditional methods for sample selection and take advantage of ensemble learning, we propose in this work a new automatic CD method for high-resolution remote sensing images. Here, according to the characteristics of data distribution, we use two groups of threshold ranges in order to replace the single threshold value that is commonly used to select training samples. Then, the extreme learning machine (ELM) [43]–[45], multinomial logistic regression (MLR) [46]–[48], and K-nearest neighbor (KNN) [49], [50] are used to build a serial-distributed MCS, in which each classifier uses the random subspace method (RSM) [30], [51]–[53] to produce different base classifiers. To make full use of the rich spatial information present in VHR images, the method not only extracts the texture features and morphological features to fill the lack of spectral information, but also utilizes the objects and spatial-contextual information to impose constraints on the obtained change map.

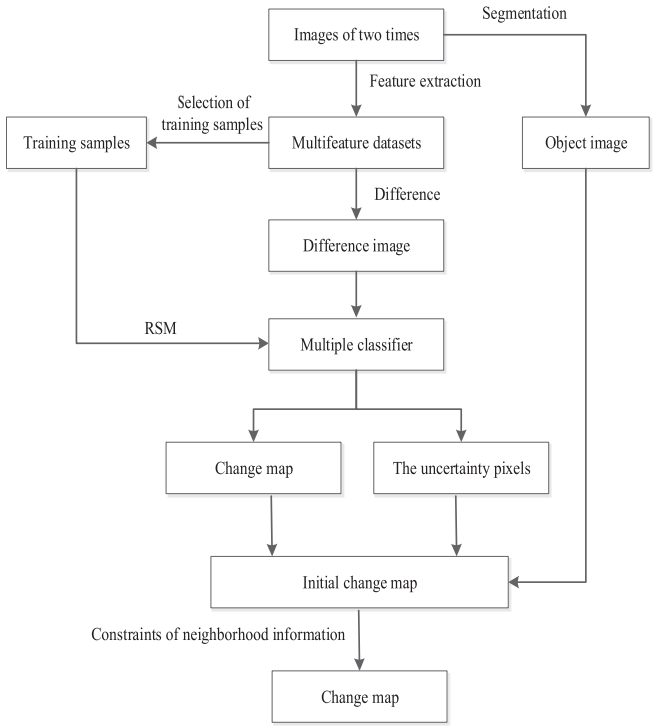


Fig. 1. Flowchart of the proposed method.

In order to test the effectiveness of the proposed method, experiments are carried out on two multitemporal and VHR images. The proposed method is compared with a variety of methods such as ELM, MLR, KNN, ELM ensemble with RSM (ELM-RSM), MCS, pixel-based CVA, and object-based CVA (CVA-OB). This paper is organized as follows. Section II presents the proposed methodology. A brief introduction to the considered data sets is given in Section III. The experimental results are discussed in Section IV. Finally, Section V concludes with some remarks and hints at plausible future research lines.

II. Proposed Method

The main contributions of the present work are twofold: 1) a new method for sample selection is designed based on CVA and 2) a serial-distributed MCS is used for CD purposes. The new sample selection method uses two groups of threshold ranges in order to obtain more evenly distributed training samples. In addition, the RSM is used to increase the diversity of the base classifiers. As already mentioned, our MCS consists of three learning algorithms: 1) ELM; 2) MLR; and 3) KNN. Last but not least, we employ objects and spatial-contextual information to impose constraints in the initial change information. A detailed description of the proposed CD method is presented in subsequent sections. A flowchart of the proposed algorithm is given in Fig. 1.

A. Generation of Input Dataset

The input dataset consists of the difference images of three features: 1) spectral features; 2) texture features; and 3) morphological features. The GLCM is a conventional method of extracting texture features effectively, and it provides comprehensive information about the gray level of the image from the viewpoints of direction, distance, and amplitude of variation [54], [55]. In this paper, five second-moment descriptors are adopted: 1) mean; 2) variance; 3) homogeneity; 4) contrast; and 5) dissimilarity. The mathematical morphology framework defines a series of operators to emphasize homogeneous spatial structures in a gray-level image [10], [56]–[58]. The opening by reconstruction has a good inhibitory effect on the brighter elements that are smaller than structuring element. The closing by reconstruction can filter out the darker elements smaller than structuring element. The opening and closing by reconstruction inherits the advantages of the aforementioned methods regarding their capacity to preserve original shapes of spatial structures. Therefore, these three morphological reconstruction filters are used to construct the input dataset.

The difference image X is produced from two superimposed images. If we suppose that the original images have *r* spectral bands and can be represented by x^1 and x^2, then X is calculated as follows:

$$\begin{cases} X = \left|x^1 - x^2\right| \\ X = X_S \cup X_T \cup X_M \end{cases} \tag{1}$$

where X_S represents the *r*-dimensional spectral features, X_T is the 5*r*-dimensional texture feature, and X_M is the 3*r*-dimensional morphological feature. Therefore, the dimensionality of the input dataset X is 9*r*. Each band of X must be normalized in the range [0,1], and the *i*th dimensional input data X_i is normalized as follows:

$$X_i = \frac{X_i - V_{\min}}{V_{\max} - V_{\min}}, \quad i = (1, \ldots, 9 * r) \tag{2}$$

where $V_{\min}$ and $V_{\max}$ are the minimum and maximum of the *i*-dimensional input data.

B. Selection of Training Samples

A common method of sample selection is based on the assumption that the greater the value of a pixel in the difference image, the greater the possibility of the pixel belonging to the change category. So, the samples corresponding to the first $\alpha \times N$ largest portion of the pixels in the difference image are regarded as positive samples and the samples corresponding to the first $\alpha \times N$ smallest portion of the pixels in the difference image are regarded as negative samples [18]. N is the number of pixels and α is the threshold. However, the training samples selected by this method are distributed unevenly, which may reduce the accuracy of CD. The error of samples increases with the increase in α. When the value of α is too large, the training samples are too many, which reduces the speed of the training process. When the value of α is too small, the samples are concentrated in one or several areas, which cannot provide accurate classification results for the whole image. In addition, the number of samples belonging to different categories is the same, which may not meet the actual distribution of pixels from different classes.

To solve this problem, an improved sampling method is proposed in this paper. The new method uses two groups of threshold ranges to replace one threshold value to get more evenly distributed training samples. The general procedure of the new method is given as follows:

Step 1) Extract the change feature with CVA [18]. The change feature map can be represented by $D = \{d_{ij}, 1 \le i \le m, 1 \le j \le n\}$, which is usually termed as "difference image"

$$d_{ij} = \sqrt{\sum_{l}^{9r} \left(x_{ij,l}^1 - x_{ij,l}^2\right)^2} \tag{3}$$

where $x_{ij,l}^1$ and $x_{ij,l}^2$ are the gray values of the pixels at the position (i, j) in the lth band of the two images, and *m* and *n* are the number of row and column of the images.

Step 2) Calculate the minimum (*minv*) and maximum (*maxv*) and an adaptive threshold (T) of the change feature map. For details on how to calculate the adaptive threshold *T*, we refer to [59]. The process of calculating the adaptive threshold T can be summarized as follows.

1) Randomly determine a value T_0, $T_0 \in \{0, 1, \ldots, \max \nu\}$, and divide the difference image into two sections

$$d_{ij} = \begin{cases} d_{ij}^a, & 0, \le d_{ij}^a \le T_0 \\ d_{ij}^b, & 0, \le d_{ij}^b \le \max \nu \end{cases} \tag{4}$$

2) Calculate the average of the two sections

$$M^a = \frac{\sum_{a=0}^{m} d_{ij}^a}{m} \tag{5}$$

$$M^b = \frac{\sum_{b=0}^{m} d_{ij}^b}{n} \tag{6}$$

where *m* is the number of d_{ij}^a and n is the number of d_{ij}^b. If the value of *m* or *n* is 0, then M^a or M^b is made equal to 0.

3) Determine a terminal value *e* and compare the average of M^a and M^b to T^0. If $\left|\frac{\left(M^a + M^b\right)}{2} - T_0\right| < e$, output the final threshold. If this condition is false, return to (1) and restart the convergence iteration computation, and get the threshold $T_{last} = \frac{\left(M^a + M^b\right)}{2}$. According to Wei *et al.* [59], the only T_{last} can be obtained when the terminal value e is excessively large. But, T_{last} is not the optimal threshold of the whole image but the optimal threshold of the largest change area. When value e is small enough, there will be two T_{last} values through an iteration computation, which is named as T_{last1} and T_{last2}

$$T_{last} = \begin{cases} T_{last1}, & 0, \le T_0 \le T_{last1} \\ T_{last2}, & T_{last1}, \le T \le \max \nu \end{cases}. \tag{7}$$

4) The preliminary threshold scope of the change feature map is $\{T_{last1}, T_{last2}\}$ and divided the change feature map outside the preliminary threshold scope into two parts

$$d_{ij}^{l} = \begin{cases} d_{ij}^{l1}, & 0, \le d_{ij}^{l1} \le T_{last1} \\ d_{ij}^{l2}, & T_{last2}, \le d_{ij}^{l2} \le \max \nu \end{cases}. \quad (8)$$

5) Calculate the standard deviation S_1 and S_2 of the two parts

$$S_1 = \sqrt{\frac{\sum_{l1}^{m1} \left(d_{ij}^{l1} - T_{last1}\right)^2}{m1}} \quad (9)$$

$$S_2 = \sqrt{\frac{\sum_{l2}^{m2} \left(d_{ij}^{l2} - T_{last2}\right)^2}{m2}} \quad (10)$$

where $m1$ is the number of the pixels that belong to d_{ij}^{l1}, $m2$ is the number of the pixels that belong to d_{ij}^{l2}. Then, the final threshold scope can be presented as $\{T_{last1} - S_1, T_{last2} + S_2\}$.

6) Calculate the final threshold *T* as follows:

$$T = \frac{\left(T_{last1} - \sqrt{\frac{\sum_{l1}^{m1} \left(d_{ij}^{l1} - T_{last1}\right)^2}{m1}}\right) + \left(T_{last2} - \sqrt{\frac{\sum_{l2}^{m2} \left(d_{ij}^{l2} - T_{last2}\right)^2}{m2}}\right)}{2}. \quad (11)$$

Step 3) Divide the change feature map into two parts by using *T*, and calculate their standard deviation (S_c and S_nc) and the ratio of the number of two parts (*P*).

Step 4) Identify two groups of threshold ranges. The threshold range of change category $C = (C_1, C_2, \ldots, C_l)$ can be represented as follows:

$$\begin{cases} C_1 = [T + S\_c, T + (1 + \alpha 1/L - 1) * S\_c] \\ C_2 = [T + 2 * S\_c, T + (2 + \alpha 2/L - 1) * S\_c] \\ \cdots \cdots \\ C_L = [T + L * S\_c, \max \nu] \end{cases} \quad (12)$$

On the other hand, the threshold range of unchanged category $U = (U_1, U_2, \ldots, U_k)$ can be represented as follows:

$$\begin{cases} \alpha 2 = \alpha 1/P \\ U_1 = [T - S\_nc, T - (1 - \alpha 2/K - 1)] \\ U_2 = [T - 2 * S\_nc, T - (2 - \alpha 2/K - 1)] \\ \cdots \cdots \\ U_K = [\min \nu, T - K * S\_nc] \end{cases} \quad (13)$$

where $L = fix\left(\left(\max v - T\right)/S\_c\right)$, $K = fix\left(\left(T - \min v\right)/S\_nc\right)$. The value of $\alpha 1$ can be set between 0 and 1 based on the number of training samples selected. As the original sample selection, the parameter values of the new method are set in this work empirically, through trial and error.

Step 5) Select the training samples from the change feature map by two groups of threshold ranges.

C. Construction of MCS

The ELM, MLR, and KNN are chosen to construct our MCS. ELM is proposed for single-hidden layer feedforward neural network (SLFN) in [60]. Unlike other traditional approaches (such as back-propagation algorithms), in the ELM the input weights and hidden biases are randomly chosen, and the output weights are analytically determined by using the Moore–Penrose generalized inverse. This makes ELM learn much faster than the traditional gradient-based learning algorithms and helps avoiding many difficulties faced by gradient-based learning methods, such as stopping criteria, learning rate, learning epochs, and local minima [43]. Multinomial logistic regression estimates the probability of each class and the algorithm of the odds ratio is assumed to be a linear function of the influencing variables. For a more detailed introduction of MLR, we refer to Bourennane *et al.* [61].The MLR model is relatively simple and, contrary to conventional linear discriminant analysis, it requires few restrictive assumptions [62]. Specifically, the features for classification do not need to be normally distributed and linearly related to the classes of interest. KNN determines the category of a pixel based on the categories of the nearest *k* samples, and it is a nonparametric classification technique based on learning by analogy [49]. This algorithm is robust and simple to implement, and it is frequently used in pattern recognition. However, all three learning algorithms have their limits, so ensemble learning can be used to improve their performance. For example, Samat *et al.* proposed two novel algorithms for ensemble ELMs (bagging-based and AdaBoost-based ELMs) for the classification task, which can improve the stability of ELM [63]. Xia *et al.* presented a classification scheme based on ELMs and KNN for cloud classification [64]; Tan proposed a new algorithm for hyperspectral image semisupervised classification, which combined with MLR and KNN [65]. These examples demonstrate the reasonableness of the combination of ELM, MLR, and KNN.

In this work, ELM is used to classify the pixels of the input dataset first. Thus, a large part of the pixels in the input image can be quickly labeled and the speed of MCS will be increased because of the high computational efficiency of the ELM. Then, the remaining unlabeled pixels are further classified by MLR and KNN in sequence with a certain rule. The specific process for constructing our MCS can be summarized as follows.

1) After selecting the training samples, the RSM is employed for ELM to produce *N* base classifiers, which are used to classify all the pixels of the input dataset, and their outputs for each pixel will be counted. A large part of pixels can be labeled using the following rule. Each pixel has *N* labels (changed class and unchanged class) after classification, and $n\_c + n\_nc = N$, where n_c is the number of labels for the changed class and n_nc is the number of labels for the unchanged class. If $|n\_c - n\_nc| \ge \frac{N+1}{2}$, the pixel can be labeled by

Fig. 2. True color images of the first dataset. High-resolution images acquired in (a) November 2012 and (b) November 2013.

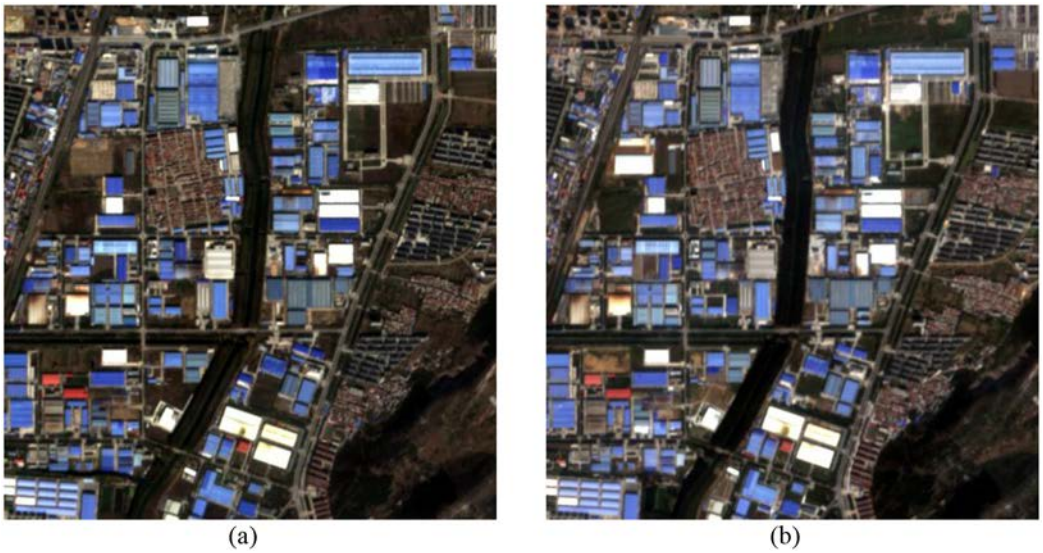


Fig. 3. True color images of the second dataset. High-resolution images acquired in (a) November 2013 and October 2014.

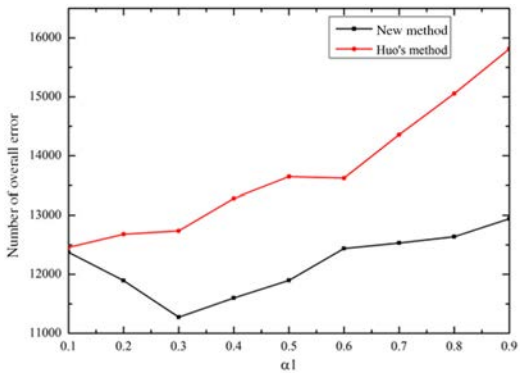


Fig. 4. Number of overall error with different $\alpha 1$ values for the new method and $\alpha * 10$ values for Huo's method to the first dataset.

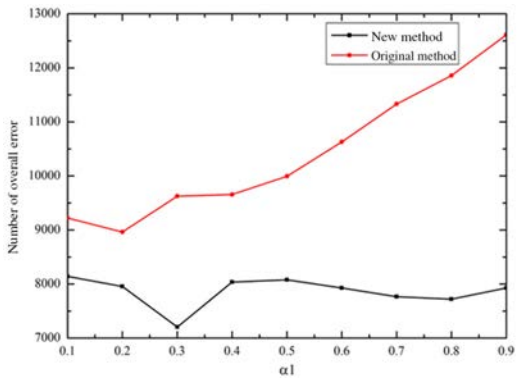


Fig. 5. Number of overall error with different $\alpha 1$ values for the new method and $\alpha * 10$ values for Huo's method to the second dataset.

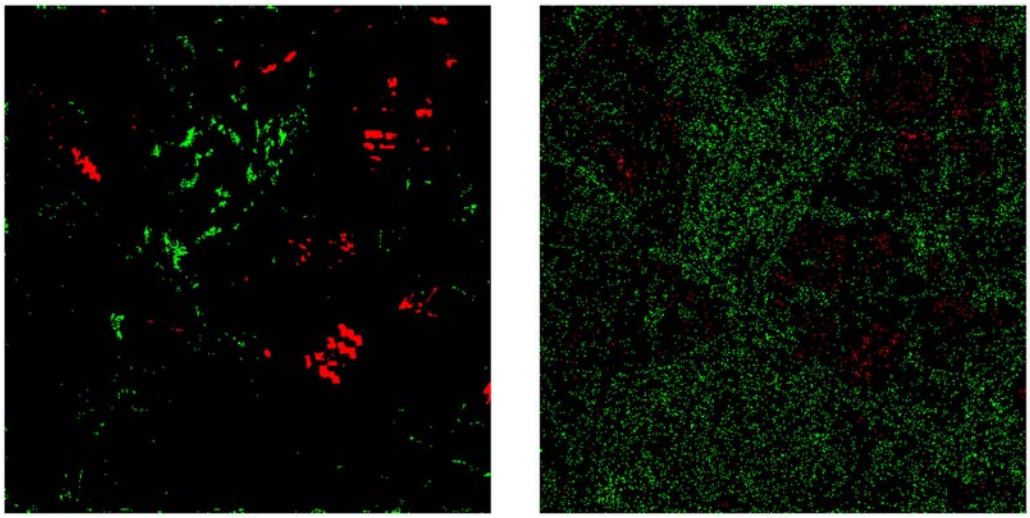


Fig. 6. Sample sets for the first dataset which were selected by different sample selection methods (red samples represent changed class and green pixels represent unchanged class). (a) Sample set for $\alpha = 0.01$ with original method. (b) Sample set for $\alpha 1 = 0.3$ with new method.

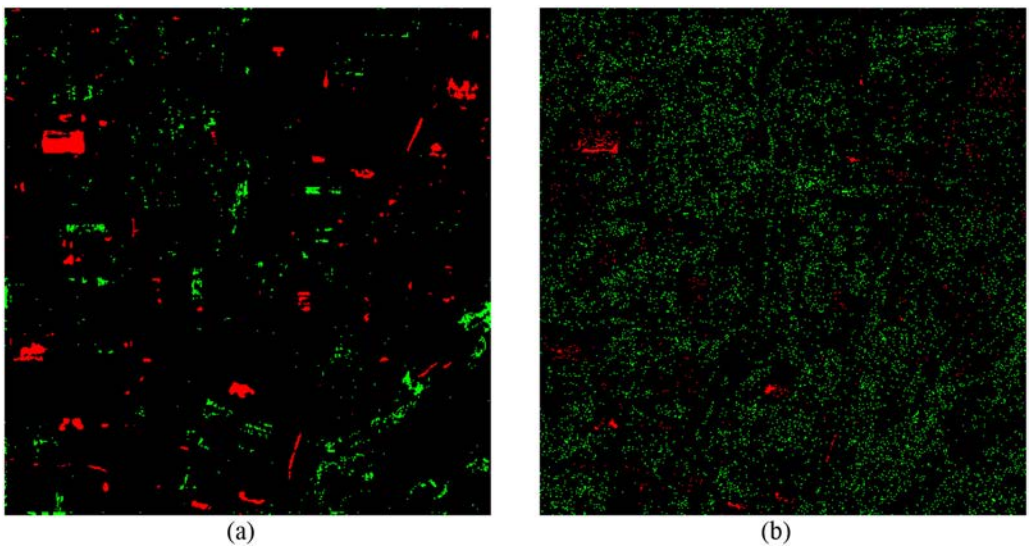


Fig. 7. Sample sets for the second dataset, which were selected by different sample selection methods (red samples represent changed class and green pixels represent unchanged class). (a) Sample set for with original method. (b) Sample set for with new method.

ELM. When $n\_c > n\_nc$, the pixel is labeled as changed class. Otherwise, the pixel is labeled as unchanged class.

2) Pixels that cannot be labeled by ELM are further classified by MLR and RSM, which are also used for producing base classifiers. Then, a part of pixels can be labeled following the strategy described in (1).
3) KNN employs the same ensemble strategy and rules to label the remaining pixels as the other two learning algorithms.
4) The pixels that could not be labeled by the three learning algorithms separately are subsequently labeled by taking into account the results of all the three learning algorithms with voting. Then, the change information is extracted initially by the MCS.

D. Change Mapping With the Constraint of Spatial Information

In order to make full use of spatial information in high-resolution remote sensing images, we employ objects and spatial-contextual information to constrain the initial change information, which can eliminate some salt-and-pepper noise and reduce the omission ratio and commission ratio. First, we segment the multitemporal images with the mean shift algorithm. Then, the objects for CD are obtained by split and merge, according to the rules proposed by Huo *et al.* [18]. The change ratio for each object is computed, and if the ratio is greater than a threshold (R), then the changed pixels are identified; otherwise, they are regarded as unchanged pixels. After that, the labels of each pixel and its eight spatial neighbors are counted. If the number of labels represented in the changed

class is greater than four, the pixel is identified as a changed one; otherwise, it will be labeled as unchanged.

III. Description of Datasets

To establish the effectiveness of the proposed methodology, we used two multitemporal and high-resolution remote sensed images collected by ZY-3 for experiments. ZY-3 satellite is a civilian high-resolution remote sensing satellite, which was launched in China on January 2012. The multispectral images collected by ZY-3 have four bands and their spatial resolution is 5.8 m. The first dataset consists of two high-resolution images captured by a multispectral camera over an area of Xuzhou (Jiangsu Province, China) on November 5, 2012 and November 4, 2013, respectively. The test site comprises 450 × 450 pixels and includes part of Yunlong district. During the two acquisition dates, some new buildings were constructed and some grassland turned into barren. Fig. 2(a) and (b) shows the two true color images of the first dataset. The two high-resolution images of second dataset were acquired in November 2013 and October 2014, respectively. A section of 450 × 450 pixels is selected as test site. The test site covers part of Tongshan District, which is also located in Xuzhou, Jiangsu Province in China. Some changes are some new buildings and crop growth in farmland. Fig. 3(a) and (b) shows the true color images of 2013 and 2014, respectively.

IV. Results and Analysis

To evaluate the effectiveness of the proposed method, experiments were conducted on two multitemporal and high-resolution remote sensing images. The performance of the proposed method is compared with those of some supervised methods based on ELM, MLR, KNN, ELM-RSM, SVM, the MCS (MCS), three unsupervised methods based on CVA (including the pixel-based CVA, the object-based CVA and CVA combined with EM algorithm), and Markov random field-based method. For the supervised methods, the training samples are the same sample set selected by the newly developed method for sample selection.

First, geometric correction and atmospheric correction were performed for the multitemporal image dataset. Then, texture features and morphological features were extracted to make up the input dataset with spectral features, as mentioned in Section II-A. For the first dataset, the window size for computing GLCM is set to 5×5 and the direction is set to $0°$, which are selected according to the resolution of objects represented in the sense and empirical experience. For the second dataset, the texture features are extracted by computing the GLCM with a 7×7-pixel window and $0°$ orientation. Morphological operations are defined with disk-shaped structuring elements and radius 5 pixels for the first dataset, and with disk-shaped structuring elements and radius 7 pixels for the second dataset.

In order to obtain appropriate training samples, we set the parameter values with a series of experiments. To keep the size of the training samples in a reasonable range, parameter α in Huo's method for sample selection was limited in the range

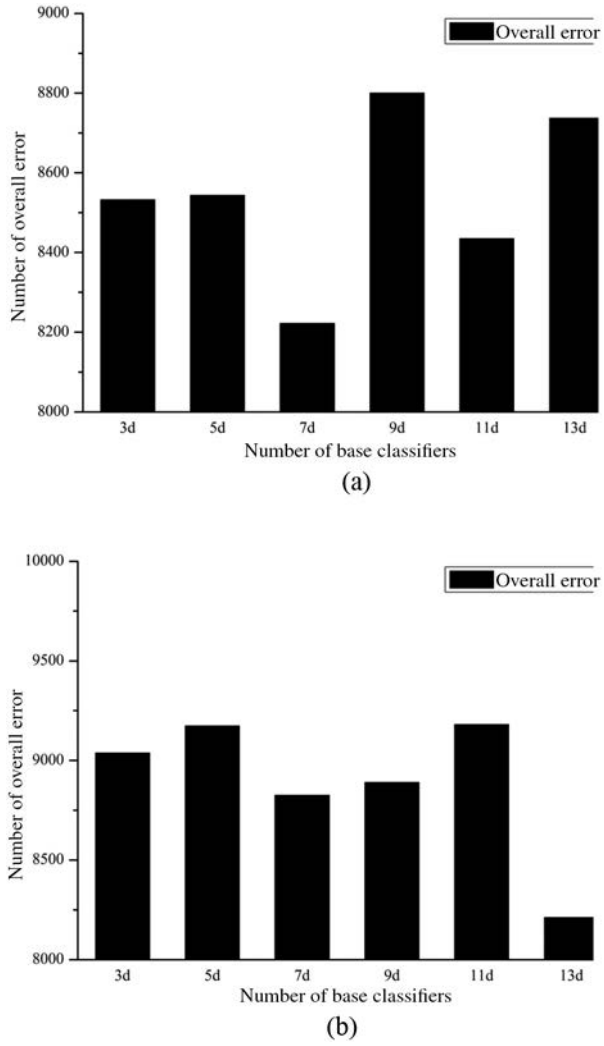


Fig. 8. Plots for different numbers of base classifiers. (a) First dataset. (b) Second dataset.

[0.01, 0.09], and parameter $\alpha 1$ in the new method was limited in the range [0.1, 0.9]. ELM was applied to find the values of parameters for each experiment because of its high speed of training. As illustrated in Fig. 4, using the image pair in Fig. 2, the new method can obtain good performance when $\alpha 1$ was set at 0.3 and Huo's method performs best when α was set at 0.01. So, we set $\alpha 1 = 0.3$ and $\alpha = 0.01$ in the experiments with the first dataset. Fig. 5 shows the overall error for different values of $\alpha 1$ and α in the experiments with the second dataset. From Fig. 5, we can see that the value of $\alpha 1$ and α should be set at 0.15 and 0.02 separately.

Fig. 6 shows the sample sets for the first dataset that were selected by different sample selection methods. The sample sets that were selected by the proposed method (distributed more evenly) are shown in Fig. 6(b), while the samples were selected by the original method were concentrated in some regions as shown in Fig. 6(a). Moreover, the number of positive samples (changed pixels) and negative samples (unchanged pixels) were almost the same in Fig. 6(a) and the numbers of samples in Fig. 6(b) are proportional to the actual distribution of pixels in the first dataset. From Fig. 4, the overall error for

TABLE I
ACCURACY OF CD BY DIFFERENT METHODS FOR THE FIRST DATASET

Method	ELM	MLR	KNN	SVM	ELM-RSM	MCS	CVA	CVA-OB	MRF	EM	Proposed method
Kappa	0.6037	0.6158	0.6002	0.6175	0.6566	0.6950	0.5740	0.6226	0.5339	0.6010	**0.7107**
OA	0.9418	0.9334	0.9292	0.9369	0.9475	0.9517	0.9171	0.9454	0.8975	0.9291	**0.9587**
Commission ratio	0.3314	0.4204	0.4423	0.3942	0.3115	0.3017	0.4966	0.3008	0.5580	0.4426	0.1991
Omission ratio	0.0358	0.0239	0.0332	0.0273	0.0291	0.0234	0.0190	0.0352	0.0132	0.0236	0.0292

TABLE II
ACCURACY OF CD BY DIFFERENT METHODS FOR THE SECOND DATASET

Method	ELM	MLR	KNN	SVM	ELM-RSM	MCS	CVA	CVA-OB	MRF	EM	Proposed method
Kappa	0.6978	0.6986	0.6970	0.7467	0.7607	0.7963	0.6266	0.6702	0.5546	0.7298	**0.8230**
OA	0.9543	0.9495	0.9496	0.9640	0.9657	0.9724	0.9130	0.9550	0.9059	0.9658	**0.9776**
Commission ratio	0.3590	0.4046	0.4031	0.2824	0.2762	0.2039	0.4930	0.3227	0.5684	0.2072	0.1143
Omission ratio	0.0137	0.0060	0.0066	0.0139	0.0122	0.0134	0.0040	0.0223	0.0018	0.0224	0.0160

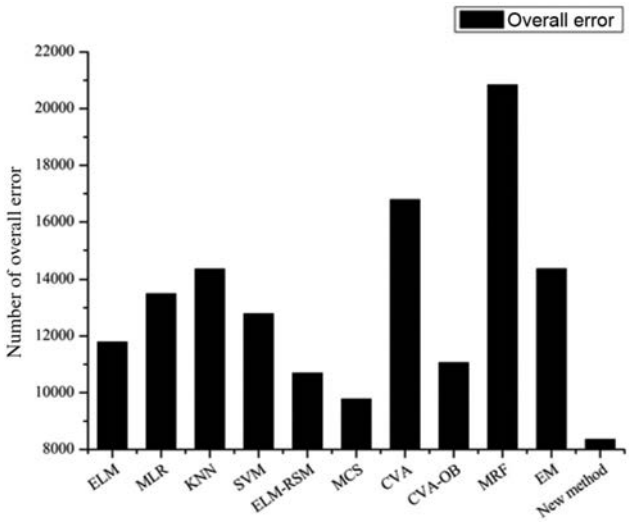


Fig. 9. Bar charts for the first dataset showing the overall error for different methods.

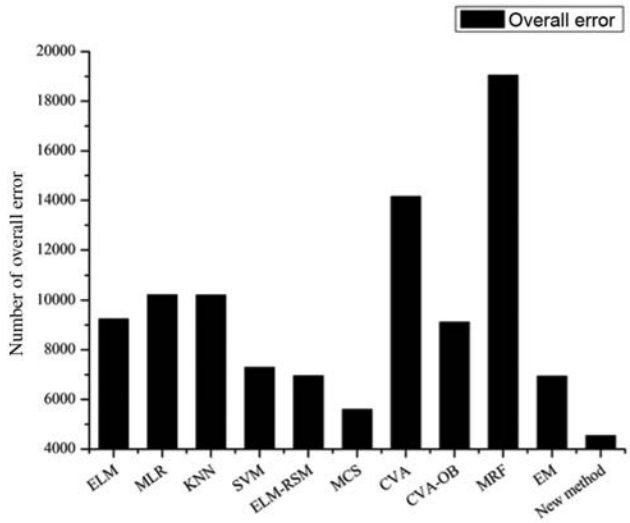


Fig. 10. Bar charts for the second dataset showing the overall error for different methods.

sample set is 12 456 in Fig. 6(a) and just 11 600 in Fig. 6(b). Fig. 7 displays the sample sets for the second dataset that were selected by different sample selection methods. It is shown that the new method could obtain more evenly distributed sample sets. Similar to the experiment with the first dataset, the overall error for the sample set is 8963 in Fig. 7(a) and is less than 7700 in Fig. 7(b). This implies that the new sample selection method presented in this work performs better than the original one. In the following experiments, the sample sets were always selected by the newly proposed method.

As mentioned in Section II-C, the RSM was employed for the three learning algorithms to produce **N** base classifiers. Considering the computational complexity of the method, we set N in the range of [3], [13] and we experimentally get the value of N. From Fig. 8, it can be observed that the overall error for the two datasets was the smallest separately when the numbers of base classifiers (N) were set to 7 and 13 separately. So, we set $N = 7$ for the first dataset and $N = 13$ for the second dataset in the following experiments.

The object-based CVA was used for comparison with the proposed method and the objects information was extracted and used to constrain the change information, so objects were obtained with the method described Section II-D. The threshold r in Section II-D was empirically set to 0.25 for the first dataset and to 0.3 for the second dataset. The parameters of the other methods such as the number of neural networks for ELM and the number of base classifiers for the ELM-RSM were determined by several groups of empirical experiments. The value k of KNN was set to three because the number of available training samples is large evenly distributed.

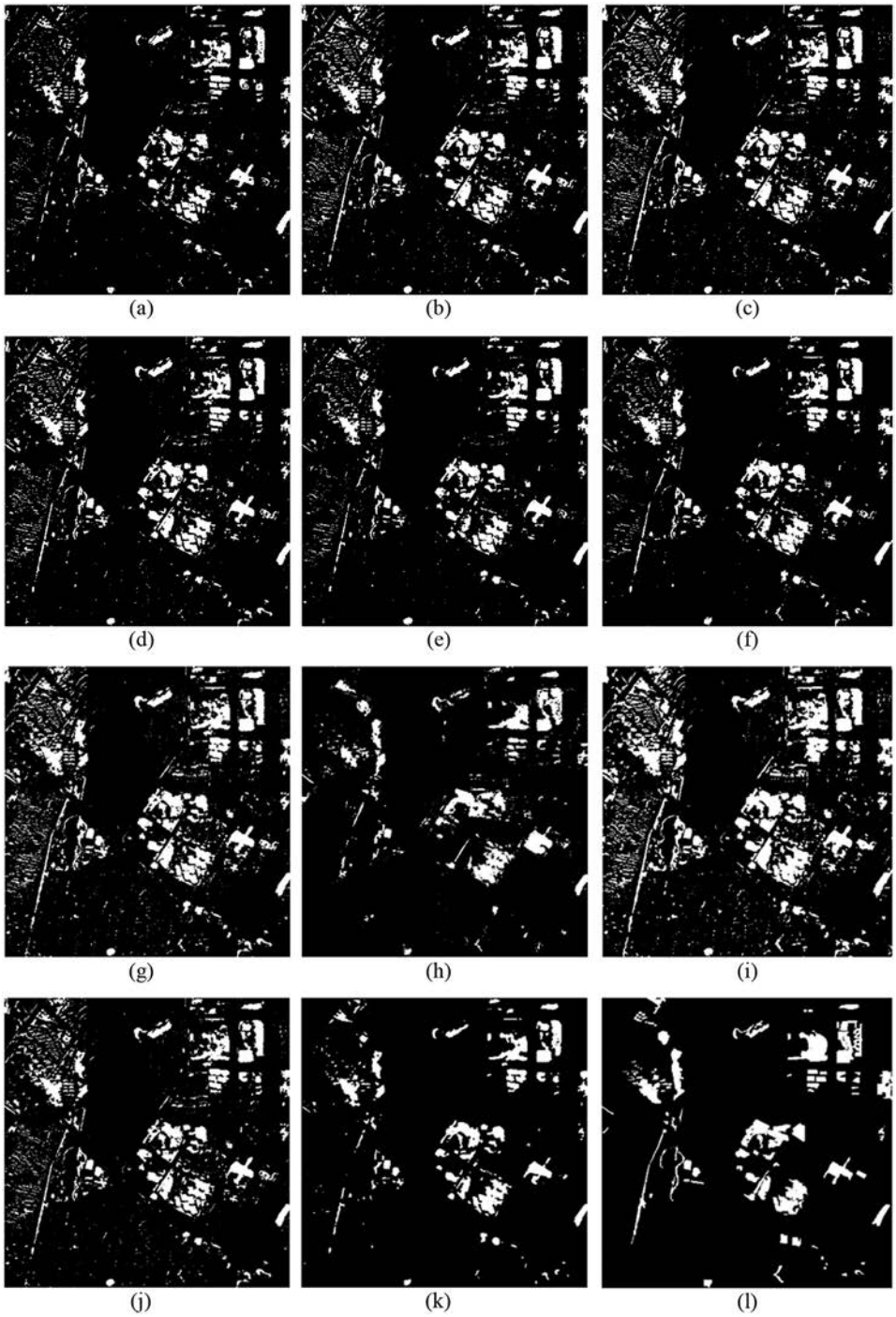


Fig. 11. CD maps obtained for the first dataset by: (a) ELM; (b) MLR; (c) KNN; (d) SVM; (e) ELM-RSM; (f) MCS; (g) CVA; (h) CVA-OB; (i) MRF; (j) EM; (k) proposed method; and (l) reference map.

To assess the effectiveness of the proposed methodology, various performance metrics were considered as follows: overall accuracy (OA); kappa coefficient; commission ratio; and omission ratio. For ELM, ELM_RSM, MCS, and the proposed method, each experiment was run 10 times because of the instability of ELM. The accuracy of those methods is reported as the average of ten experiments. The accuracy of CD for the two datasets is shown in Tables I and II, respectively. The charts of the number of overall error for CD are also shown in Figs. 9 and 10. As shown in Table I, the proposed method outperforms the other methods for the first dataset. The mean kappa and mean OA computed are equal to 0.7107 and 0.9587, which are the highest among the methods. The mean commission ratio and mean omission ratio of the proposed method are 0.1991 and 0.0292, respectively. For the first dataset, overall accuracies are ranging from 0.9580 to 0.9598 at the 95% confidence interval and kappa coefficients are ranging from 0.7024 to 0.7219 at the 95% confidence interval. For the second dataset, the proposed method has better performance as illustrated in Table II. The overall accuracies of the proposed

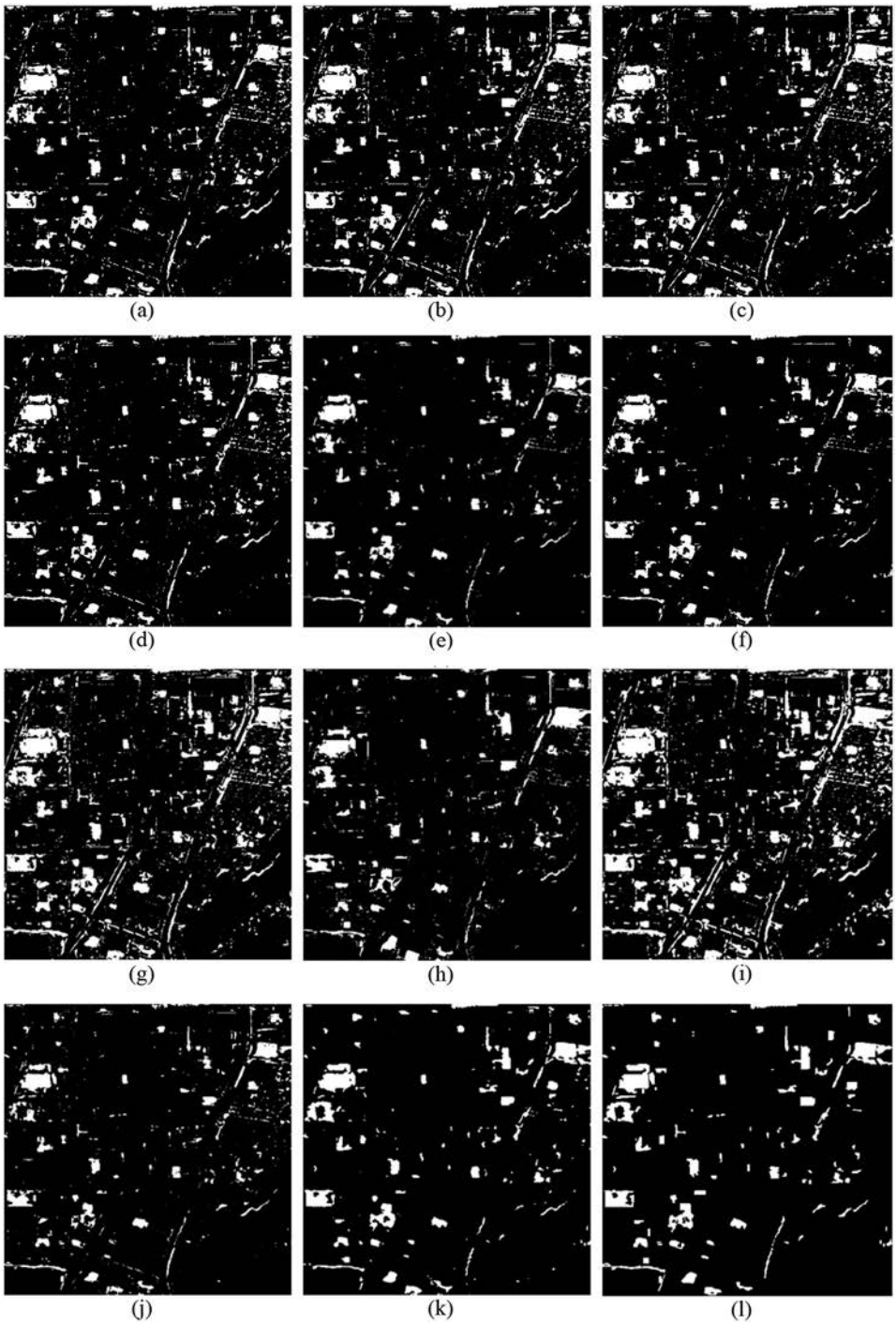


Fig. 12. CD maps obtained for the second dataset by: (a) ELM; (b) MLR; (c) KNN; (d) SVM; (e) ELM-RSM; (f) MCS; (g) CVA; (h) CVA-OB; (i) MRF; (j) EM; (k) proposed method; and (l) reference map.

method are ranging from 0.9801 to 0.9828 at the 95% confidence interval and kappa coefficients are ranging from 0.8205 to 0.8582 at the 95% confidence interval. The mean kappa and mean OA are 0.8230 and 0.9776, which are the highest, while the mean commission ratio and mean omission ratio are 0.1143 and 0.0160. From Tables I and II, it can be noticed that the performance of the individual base classifiers is improved by the ensemble algorithm. For example, the kappa and OA of ELM are 0.6037 and 0.9418 for the first dataset, but the same factors for the ensemble ELM are 0.6566 and 0.9475. The kappa and OA of ELM-RSM for the second dataset are increased by 0.0629 and 0.0114, respectively, compared with ELM. In addition, the performance of MCS is significantly better than that of ELM, MLR, and KNN. The unsupervised method based on CVA exhibits poor performance when compared with other supervised methods. The commission ratio of CVA is equal to 0.4966 and 0.4930 for the first dataset and second dataset respectively. The unit for analysis in the

object-based CVA is an image object, which contains abundant spatial information. Therefore, the accuracy of CVA-OB is higher than the pixel based CVA. However, due to the impact of image segmentation error, a large number of pixels are omitted or falsely detected. The EM algorithm can provide the change threshold for difference image to extract change information. The OA of EM-based method is increased by 0.012 for the first dataset, and by 0.0528 for the second dataset when compared with CVA. The result of MRF-based method is the worst and its commission ratio is the highest among all methods. From Figs. 9 and 10, it can be observed that the overall error of the proposed method when analyzing the two datasets is less than that obtained by other methods. The proposed method exhibits superiority in CD.

For visual comparison purposes, the change maps obtained for the two considered datasets are displayed in Figs. 11 and 12, respectively. The changes shown in Fig. 11 are mainly about the increase in impervious (road and building) areas and decrease of vegetation and barren. The main reason is the urban sprawl and construction of Xuzhou in recent years. From Fig. 11, it can be observed that a lot of unchanged pixels present in the leftmost part of the reference map [in Fig. 11(l)] are wrongly identified as changed pixels in the maps obtained by the supervised methods such as ELM [in Fig. 11(a)], MLR [in Fig. 11(b)], KNN [in Fig. 11(c)], SVM [in Fig. 11(d)], ELM-RSM [in Fig. 11(e)], and MCS [in Fig. 11(f)], whereas the pixels belonging to false detection class are significantly less in the map [in Fig. 11(k)] obtained by the proposed method. Meanwhile, as compared with the reference map [in Fig. 11(l)], some changed regions of the maps obtained by object-based CVA [Fig. 11(g)] are wrongly identified as unchanged regions, whereas the regions can be detected by the proposed method accurately. Compared with other change maps, a large number of unchanged pixels are erroneously detected as changed pixels by the methods of pixel based CVA [see Fig. 11(g)], EM [see Fig. 11(j)], and MRF [see Fig. 11(i)]. The change map obtained by the proposed method [in Fig. 11(k)] is more consistent across the reference map [in Fig. 11(l)] compared with the change map obtained by MCS [in Fig. 11(f)], which benefits from the spatial constraint. For the second dataset, many buildings were constructed and some crops were growing out in farmland during the acquisition date. From the Fig. 12, it could be seen that small changed regions are scattered all over the maps [in Figs. 12(a)–(j)]. These scattered and wrongly identified changed regions in the change map are comparatively less than those obtained by the other seven methods. In addition, it has also been found that the changed region in the middle left of Fig. 12(k) is totally absent in the maps obtained by some methods such as ELM, MLR, KNN, SVM, and CVA-OB, whereas this changed region is identified in Fig. 12(k) obtained by the proposed method. The MRF-based method in this paper uses the result of CVA as the initial change map, and the relatively bad performance of CVA [see Fig. 12(f)] affects the result of MRF-based method [see Fig. 12(h)], which considers the spatial-contextual information. All these results illustrate that the proposed method can obtain more accurate change maps as compared with other methods.

V. Conclusion and Future Lines

In this paper, we have developed a new automatic CD method that combines a new method of sample selection and an MCS. In order to obtain more evenly distributed training samples, the new method of sample selection uses two groups of threshold ranges to replace one threshold value. Then, ELM, MLR, and KNN are used to build a distributed MCS. The MCS performs a classification of the input dataset using spectral features, texture features, and morphological features. After that, the objects and spatial-contextual information are utilized to constraint the change map. Experiments are carried out on two multitemporal and high-resolution images to confirm the effectiveness of the proposed method. From the results obtained, it has been demonstrated that the proposed method has a better performance on CD for high-resolution remote sensing images. However, parameter settings still depend on trial and error in the current version. In future, we will perform a detailed study about the relationship between the parameters and the dataset to find more intelligent methods for parameter settings.

References

[1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, pp. 989–1003, 1989.

[2] J. Zhou, B. Yu, and J. Qin, "Multi-level spatial analysis for change detection of urban vegetation at individual tree scale," *Remote Sens.*, vol. 6, pp. 9086–9103, 2014.

[3] C. M. Mertes, A. Schneider, D. Sulla-Menashe, A. Tatem, and B. Tan, "Detecting change in urban areas at continental scales with MODIS data," *Remote Sens. Environ.*, vol. 158, pp. 331–347, 2015.

[4] K. S. Willis, "Remote sensing change detection for ecological monitoring in United States protected areas," *Biol. Conserv.*, vol. 182, pp. 233–242, 2015.

[5] G. Cortés, M. Girotto, and S. A. Margulis, "Analysis of sub-pixel snow and ice extent over the extratropical Andes using spectral unmixing of historical Landsat imagery," *Remote Sens. Environ.*, vol. 141, pp. 64–78, 2014.

[6] X. Li, Y. Zhang, X. Liu, and Y. Chen, "Assimilating process context information of cellular automata into change detection for monitoring land use changes," *Int. J. Geog. Inf. Sci.*, vol. 26, pp. 1667–1687, 2012.

[7] M. Boldt, A. Thiele, and K. Schulz, "Object-based urban change detection analyzing high resolution optical satellite images," in *Proc. SPIE Remote Sens.*, 2012, pp. 85380E-1–85380E-9.

[8] B. Demir, F. Bovolo, and L. Bruzzone, "Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1930–1941, May 2012.

[9] G. Cao, Y. Li, Y. Liu, and Y. Shang, "Automatic change detection in high-resolution remote-sensing images by means of level set evolution and support vector machine classification," *Int. J. Remote Sens.*, vol. 35, pp. 6255–6270, 2014.

[10] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, "Supervised change detection in VHR images using contextual information and support vector machines," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 20, pp. 77–85, 2013.

[11] F. Bovolo, G. Camps-Valls, and L. Bruzzone, "A support vector domain method for change detection in multitemporal images," *Pattern Recognit. Lett.*, vol. 31, pp. 1148–1154, 2010.

[12] S. Marchesi, F. Bovolo, and L. Bruzzone, "A context-sensitive technique robust to registration noise for change detection in VHR multispectral images," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1877–1889, Jul. 2010.

[13] Y. Bazi, F. Melgani, and H. D. Al-Sharari, "Unsupervised change detection in multispectral remotely sensed imagery with level set methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3178–3187, Aug. 2010.

[14] B. Faiza, S. S. Yuhaniz, S. M. Hashim, and A. Kalema, "Detecting floods using an object based change detection approach," in *Proc. Int. Conf. Comput. Commun. Eng. (ICCCE)*, 2012, pp. 44–50.

[15] T.-A. Teo and T.-Y. Shih, "Lidar-based change detection and change-type determination in urban areas," *Int. J. Remote Sens.*, vol. 34, pp. 968–981, 2013.

[16] B. Demir, F. Bovolo, and L. Bruzzone, "Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 300–312, Jan. 2013.

[17] M. Roy, S. Ghosh, and A. Ghosh, "A novel approach for change detection of remotely sensed images using semi-supervised multiple classifier system," *Inf. Sci.*, vol. 269, pp. 35–47, 2014.

[18] C. Huo, Z. Zhou, H. Lu, C. Pan, and K. Chen, "Fast object-level change detection for VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 118–122, Jan. 2010.

[19] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, "A critical synthesis of remotely sensed optical image change detection techniques," *Remote Sens. Environ.*, vol. 160, pp. 1–14, 2015.

[20] S. Patra, S. Ghosh, and A. Ghosh, "Histogram thresholding for unsupervised change detection of remote sensing images," *Int. J. Remote Sens.*, vol. 32, pp. 6071–6089, 2011.

[21] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[22] M. Hao, H. Zhang, W. Shi, and K. Deng, "Unsupervised change detection using fuzzy c-means and MRF from remotely sensed images," *Remote Sens. Lett.*, vol. 4, pp. 1185–1194, 2013.

[23] Y. Chen and Z. Cao, "An improved MRF-based change detection approach for multitemporal remote sensing imagery," *Signal Process.*, vol. 93, pp. 163–175, 2013.

[24] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 609–630, Mar. 2013.

[25] N. Chehata, C. Orny, S. Boukir, D. Guyon, and J. Wigneron, "Object-based change detection in wind storm-damaged forest using high-resolution multispectral images," *Int. J. Remote Sens.*, vol. 35, pp. 4758–4777, 2014.

[26] Y. Tang, X. Huang, and L. Zhang, "Fault-tolerant building change detection from urban high-resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1060–1064, Sep. 2013.

[27] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91–106, 2013.

[28] H. Murray, A. Lucieer, and R. Williams, "Texture-based classification of sub-Antarctic vegetation communities on Heard Island," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 12, pp. 138–149, 2010.

[29] D. Tuia, F. Pacifici, M. Kanevski, and W. J. Emery, "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3866–3879, Nov. 2009.

[30] A. Jurek, Y. Bi, S. Wu, and C. Nugent, "A survey of commonly used ensemble-based classification techniques," *Knowl. Eng. Rev.*, vol. 29, pp. 551–581, 2014.

[31] L. Breiman, "Heuristics of instability and stabilization in model selection," *Ann. Stat.*, vol. 24, pp. 2350–2383, 1996.

[32] M. Skurichina, L. I. Kuncheva, and R. P. Duin, "Bagging and boosting for the nearest mean classifier: Effects of sample size on diversity and accuracy," in *Multiple Classifier Systems*. New York, NY, USA: Springer, 2002, pp. 62–71.

[33] S. Datta, V. Pihur, and S. Datta, "An adaptive optimal ensemble classifier via bagging and rank aggregation with applications to high dimensional data," *BMC Bioinformat.*, vol. 11, p. 1, 2010.

[34] L. He, Q. Song, J. Shen, and Z. Hai, "Ensemble numeric prediction of nearest-neighbor learning," *Inf. Technol. J.*, vol. 9, pp. 535–544, 2010.

[35] N. Murrugarra-Llerena and A. de Andrade Lopes, "An adaptive graph-based K-nearest neighbor," in *Proc. Eur. Conf. Mach. Learn. Principles Pract. Knowl. Discov. Databases*, 2011, pp. 1–11.

[36] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4768–4786, Sep. 2015.

[37] L. Zhang and W.-D. Zhou, "Sparse ensembles using weighted combination methods based on linear programming," *Pattern Recognit.*, vol. 44, pp. 97–106, 2011.

[38] H. Parvin and H. Alizadeh, "Classifier ensemble based class weightening," *Amer. J. Sci. Res.*, vol. 19, pp. 84–90, 2011.

[39] C. De Stefano, F. Fontanella, G. Folino, and A. S. Di Freca, "A Bayesian approach for combining ensembles of GP classifiers," in *Multiple Classifier Systems*. New York, NY, USA: Springer, 2011, pp. 26–35.

[40] Y. Bi, S. Wu, H. Wang, and G. Guo, "Combination of evidence-based classifiers for text categorization," in *Proc. IEEE 23rd Int. Conf. Tools Artif. Intell. (ICTAI'11)*, 2011, pp. 422–429.

[41] A. Jurek, Y. Bi, S. Wu, and C. Nugent, "Classification by cluster analysis: A new meta-learning based approach," in *Multiple Classifier Systems*. New York, NY, USA: Springer, 2011, pp. 259–268.

[42] B. Yang, C. Cao, Y. Xing, and X. Li, "Automatic classification of remote sensing images using multiple classifier systems," *Math. Prob. Eng.*, vol. 2015, 2015, pp. 1–10.

[43] M. A. Bencherif, Y. Bazi, A. Guessoum, N. Alajlan, F. Melgani, and H. AlHichri, "Fusion of extreme learning machine and graph-based optimization methods for active classification of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 527–531, Mar. 2015.

[44] M. R. Daliri, "Combining extreme learning machines using support vector machines for breast tissue classification," *Comput. Methods Biomech. Biomed. Eng.*, vol. 18, pp. 1–7, 2013.

[45] G. Huang, S. Song, J. N. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2405–2417, Dec. 2014.

[46] P. Sentas and L. Angelis, "Categorical missing data imputation for software cost estimation by multinomial logistic regression," *J. Syst. Software*, vol. 79, pp. 404–414, 2006.

[47] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.

[48] K. Tan, J. Hu, J. Li, and P. Du, "A novel semi-supervised hyperspectral image classification approach based on spatial neighborhood information and classifier combination," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 19–29, 2015.

[49] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*. New York, NY, USA: Springer, 2003, pp. 986–996.

[50] J. Calvo-Zaragoza, J. J. Valero-Mas, and J. R. Rico-Juan, "Improving kNN multi-label classification in prototype selection scenarios using class proposals," *Pattern Recognit.*, vol. 48, pp. 1608–1622, 2014.

[51] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[52] J.-M. Yang, B.-C. Kuo, P.-T. Yu, and C.-H. Chuang, "A dynamic subspace method for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2840–2853, Jul. 2010.

[53] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, pp. 3–17, 2014.

[54] D. J. Marceau, P. J. Howarth, J.-M. M. Dubois, and D. J. Gratton, "Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 513–519, Jul. 1990.

[55] C. E. Honeycutt and R. Plotnick, "Image analysis techniques and gray-level co-occurrence matrices (GLCM) for calculating bioturbation indices and characterizing biogenic sedimentary structures," *Comput. Geosci.*, vol. 34, pp. 1461–1472, 2008.

[56] E. M. Sigurðsson, S. Valero, J. A. Benediktsson, J. Chanussot, H. Talbot, and E. Stefánsson, "Automatic retinal vessel extraction based on directional mathematical morphology and fuzzy classification," *Pattern Recognit. Lett.*, vol. 47, pp. 164–171, 2014.

[57] A. Serna and B. Marcotegui, "Detection, segmentation and classification of 3D urban objects using mathematical morphology and supervised learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 243–255, 2014.

[58] K. Tan, E. Li, Q. Du, and P. Du, "Hyperspectral image classification using band selection and morphological profiles," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 40–48, Jan. 2014.

[59] L. Wei, Y. Zhong, L. Zhang, and P. Li, "Adaptive change method of remote sensing image fusion," *Yaogan Xuebao- J. Remote Sens.*, vol. 14, pp. 1196–1211, 2010.

[60] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, 2006.

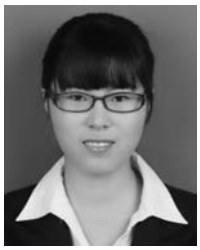
[61] H. Bourennane *et al.*, "Comparative performance of classification algorithms for the development of models of spatial distribution of landscape structures," *Geoderma*, vol. 219, pp. 136–144, 2014.

[62] P. Zhong and R. Wang, "Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1319–1334, Jul. 2014.

[63] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E^{2}LMs: Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014.
[64] M. Xia, W. Lu, J. Yang, Y. Ma, W. Yao, and Z. Zheng, "A hybrid method based on extreme learning machine and k-nearest neighbor for cloud classification of ground-based visible cloud image," *Neurocomputing*, vol. 160, pp. 238–249, 2015.
[65] K. Tan, J. Hu, J. Li, and P. Du, "A novel semi-supervised hyperspectral image classification approach based on spatial neighborhood information and classifier combination," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 19–29, 2015.

Kun Tan received the B.S. degree in information and computer science from Hunan Normal University, Hunan, China, in 2004. He was a Joint Ph.D. candidate in remote sensing at Columbia University, New York, NY, USA during September 2008 to September 2009. He received the Ph.D. degree in photogrammetric and remote sensing from China University of Mining and Technology, Jiangsu, China, in 2010.

Since 2010, he has been with the Department of Surveying, Mapping and Geoinformation, China University of Mining and Technology, Xuzhou, China. He is currently a Professor with the Department of Surveying, Mapping, and Geoinformation. His research interests include hyperspectral image classification and detection, spectral unmixing, quantitative inversion of land surface parameters, and urban remote sensing.



Xiao Jin received the Bachelor's degree in geographic information science from the Institute of Surveying and Landl Information Engineering, Henan Polytechnic University, Henan, China, in 2013. She is currently pursuing the master degree in photogrammetric and remote sensing at the Institute of Surveying and Spatial Information Engineering, China University of Mining and Technology, Xuzhou, China.

Her research interests include spectral unmixing and change detection.

Antonio Plaza (M'05–SM'07–F'15) received the Ph.D. degree in computer engineering from the University of Extremadura, Caceres, Spain, in 2002.

He is the Head of the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, Badajoz, Spain. He has authored or coauthored more than 500 publications, including more than 170 journal papers (more than 120 in IEEE journals), 20 book chapters, and more than 250 peer-reviewed conference proceeding papers. He has guest edited 9 special issues on Hyperspectral Remote Sensing for different journals. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza is an Associate Editor for IEEE ACCESS, and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012), and the *IEEE Geoscience and Remote Sensing Magazine* (2013). He was also a member of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012, and is currently serving as President of the Spanish Chapter of IEEE GRSS. He has reviewed more than 500 manuscripts for over 50 different journals. He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal. He was the recipient of the recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009) and the recognition of Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), for which he served as Associate Editor in 2007–2012. He was also the recipient of the Best Column Award of the IEEE SIGNAL PROCESSING MAGAZINE in 2015, the 2013 Best Paper Award of the JSTARS journal, and the most highly cited paper (2005–2010) in the Journal of Parallel and Distributed Computing. He was the recipient of best paper awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology.

Xuesong Wang received the Ph.D. degree in control science and engineering from China University of Mining and Technology, Xuzhou, China, in 2002.

She is currently a Professor with the School of Information and Electrical Engineering, China University of Mining and Technology. Her research interests include machine learning, artificial intelligence, and hyperspectral image processing.

Dr. Wang was the recipient of the New Century Excellent Talents in University from the Ministry of Education of China, in 2008.

Liang Xiao (M'11) was born in Hunan, China, in 1976. He received the B.S. degree in applied mathematics and the Ph.D. degree in computer science from Nanjing University of Science and Technology (NJUST), Nanjing, China, in 1999 and 2004, respectively.

From 2006 to 2008, he was a Postdoctoral Research Fellow with the Pattern Recognition Laboratory, NJUST. From 2009 to 2010, he was a Postdoctoral Fellow with the Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently a Professor with the School of Computer Science and Engineering, NJUST, where he has been serving as the second Director of the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information, Ministry of Education, since 2014, and as the Deputy Director of the Key Laboratory of Spectral Imaging Intelligent Perception, Jiangsu Province of China, since 2013. His research interests include signal processing, image processing and computer vision, machine learning, and remote sensing.

Peijun Du received the Ph.D. degree in photogrammetric and remote sensing from China University of Mining and Technology, Jiangsu, China, in 2001.

He is a Professor of Photogrammetry and Remote Sensing with the Department of Geographic information Sciences, Nanjing University, Nanjing, China. He is the Deputy Director of the Key Laboratory for Satellite Surveying Technology and Applications, National Administration of Surveying and Geoinformation, Beijing, China. His research interests include remote sensing image processing and pattern recognition; remote sensing applications; hyperspectral remote sensing information processing; multisource geospatial information fusion and spatial data handling; integration and applications of geospatial information technologies; and environmental information science (environmental informatics).