

A Novel Approach for Hyperspectral Change Detection Based on Uncertain Area Analysis and Improved Transfer Learning

Xiaohua Tong¹, Senior Member, IEEE, Haiyan Pan, Sicong Liu², Member, IEEE, Binbin Li, Xin Luo³, Huan Xie⁴, Senior Member, IEEE, and Xiong Xu⁵

Abstract—Although a number of change detection (CD) methods have been proposed during the past years, most of them are developed based on the assumption that there are either training samples or no training samples for both the pretime and posttime images. Few studies have addressed the scenario of only small amounts of training samples are available only in a single-time image. In this article, we propose a novel approach that can detect multiple changes in bitemporal hyperspectral images when only a few training samples are available in one of the images (the source image). The proposed method consists of four main steps: first, unsupervised CD based on uncertain area analysis to generate the binary change map; second, classification of the source image (X_1) according to active learning; third, classification of the target image (X_2) by the use of improved transfer learning; and fourth, generation of the multiclass change map by postclassification comparison. The proposed method was tested on one simulated dataset and two pairs of real bitemporal hyperspectral images. Experimental results demonstrate that: first, uncertain area analysis can improve the binary CD accuracy; while active learning and improved transfer learning can enhance the classification accuracy of the source and target images, the multiple CD accuracy is increased by the use of the proposed method; and second, compared with the state-of-the-art methods, the proposed method produced best results.

Index Terms—Change detection (CD), hyperspectral (HS) images, multiple changes, transfer learning, uncertain area analysis.

I. INTRODUCTION

CHANGE detection (CD) is the process designed to identify the changes that have occurred in the same geographical area at different observation times [1]. It has been widely used in various applications, such as urban expansion [2]–[4], land use/cover monitoring [5], [6], state and local planning [7], and disaster monitoring [8], [9]. Timely land-cover CD has grown

Manuscript received August 27, 2019; revised January 12, 2020, March 22, 2020, and April 21, 2020; accepted April 21, 2020. Date of publication April 28, 2020; date of current version May 22, 2020. This work was supported in part by the National Natural Science Foundation of China under Project 41631178, Project 41601354, Project 41601414, and Project 41971299 and in part by the National Key R&D Program of China under Project 2018YFB0505400 and Project 2018YFB0505000. (Corresponding author: Xiaohua Tong.)

The authors are with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: xhtong@tongji.edu.cn; hypan@tongji.edu.cn; sicong.liu@tongji.edu.cn; libinbin@tongji.edu.cn; huanxie@tongji.edu.cn; vxiong@tongji.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.2990481

in importance in recent years alongside our increasing concern over rapid urbanization and ecological change [10], and it can be used to provide timely and comprehensive information to decision-makers and for government management. In general, CD methods can be divided into two categories according to the availability of the ground reference samples: supervised and unsupervised CD [11]. Supervised CD generally involves using training samples to classify the images, and then postclassification comparison is conducted to generate the change map [12]. However, collection of training samples is often time-consuming. Thus, unsupervised CD is more appealing in real applications. According to the final applications, unsupervised CD can be divided into binary CD and multiple CD [13]. The aim of binary CD is to distinguish changed pixels from unchanged ones. However, the “from-to” information of the change is ignored. During the past decades, a number of binary CD methods have been proposed, e.g., change vector analysis (CVA) [14], [15], clustering methods [16], transformation based techniques [17], [18], object-based methods [19], deep learning methods [20]–[22] as well as the other specific approaches [23]–[36]. Among these methods, CVA is one of the most commonly used in binary CD. The most important step of CVA is determining an appropriate threshold to distinguish changed from unchanged on the magnitude of the spectral vectors [13]. A single-threshold method is simple and convenient to be implemented. However, sometimes it cannot accurately label the pixels whose magnitude values are close to the predefined threshold. Therefore, the concept of an uncertain region has been proposed [14], [15]. In addition, the single threshold method only considers the magnitude of the change vector, whereas other information such as texture and spatial information are ignored, which often leads to the misdetection of change information. As a result, a number of compound methods have been presented, e.g., by using a combination of texture information [6], spectral angle information [37], [38], spatial information [19], and CVA in probability domain [39].

Compared to binary CD, the detection of multiple changes is more challenging. A number of approaches have been developed over the past decades, e.g., postclassification comparison [39], [40], multi-index-based methods [41], unmixing-based methods [11], [42], CD in the CVA polar domain [1], [13]–[14], [43], hierarchical spectral CVA [15], sequential spectral CVA [13]. In [44], an adaptive band-dependent model was

proposed to represent and discriminate the multiple change information.

However, all of the above-mentioned methods have been developed based on the assumption that there is either some or no ground truth in both of the images. In real applications, there is also the scenario that only small amounts of ground-truth data are available in only one of the images. To date, there has been relatively few research into this situation. In order to solve this problem, transfer learning-based methods have been proposed. In [45], Demir *et al.* proposed a novel change detection driven transfer learning (CDTL) approach to update the land-cover map. This was done by applying an unsupervised CD method and transferring the class labels of the detected unchanged training samples from the source to the target domain to initialize the target domain training set. The training set was then optimized by an active learning procedure. Finally, the target image was classified. Xian *et al.* [46] and Xian and Homer [47] updated the 2001 National Land Cover Database and impervious surface products to 2006 using a CD method. Multithreshold method was used to distinguish changed from unchanged. The unchanged pixels were randomly selected then transferred into the target domain. Zurita-Milla *et al.* [42] presented a backdating approach for land-cover classification and change analysis, in which the unchanged objects were transferred. Finally, the types of change were identified according to certain features, e.g., the normalized difference crop index, the normalized difference water index, the shape index, the brightness, and the “relative border to forest with no change.”

According to the above-mentioned studies, some open problems still exist. First, most of the thresholding methods use a single threshold to generate the binary change map, ignoring the uncertainty of the pixels’ magnitude value around the threshold. Pixel’s magnitude value near the threshold has a higher probability of misclassification than those far from it [14], [15], [48]. Second, most of the transfer learning methods transfer the label of the unchanged pixels into the target domain by random selection, which will impact the classification accuracy. Finally, most of the above-mentioned multiple CD methods were mainly designed for multispectral images, only a few focused on hyperspectral (HS) CD.

Therefore, the objective of this article is to present a CD method that can detect multiple changes when only a small amount of training samples are available for a single-time HS image in the multitemporal domain. A CD method by taking advantages of uncertain area analysis, active learning, and improved transfer learning is proposed. First, unsupervised binary CD based on uncertain area analysis is performed to generate the binary change map. The source image is then classified according to active learning. After that, the classification map of the target image is obtained by the use of improved transfer learning method. Finally, the change map is generated by postclassification comparison. The proposed method was validated on a simulated dataset and two real HS datasets. Experiment results demonstrate the effectiveness of the proposed method.

The novelties of this article are that: it proposes a method to solve the multiple CD issue when only a small amount of training samples are available in single-time image. An uncertain area is

defined to improve the CD accuracy. In addition, an optimization mechanism is used in the transferring learning process to obtain reliable samples, and a satisfactory result can be obtained with the limited number of the transfer learning samples.

The rest of this article is organized as follows. Section II describes the methodology in detail. The study area and data descriptions are presented in Section III. Experimental results are provided in Section IV. In Section V, the parameter of uncertain area and the number of samples in improved transfer learning are analyzed, and the contributions of each component are evaluated and discussed. Finally, Section VI concludes this article.

II. PROPOSED METHOD

Fig. 1 shows the flowchart of the proposed CD approach. The proposed approach consists of the following four main components:

- 1) binary CD based on uncertain area analysis;
- 2) source image classification using active learning;
- 3) target image classification using the improved transfer learning method;
- 4) generation of the multiple change map by postclassification comparison.

A. Binary CD Based on Uncertain Area Analysis

Let us consider two HS images \mathbf{X}_1 and \mathbf{X}_2 with the size $I \times J \times B$ acquired in the same geographical area at times t_1 and t_2 , respectively. I and J are the rows and columns of the image and B is the number of spectral bands. The difference image \mathbf{X}_D of CVA is formulated by subtracting the multitemporal images from each other, pixel by pixel [44], i.e.,

$$\mathbf{X}_D = \mathbf{X}_2 - \mathbf{X}_1. \quad (1)$$

The change magnitude ρ can be calculated as follows [44]:

$$\rho = \sqrt{\sum_{b=1}^B (\mathbf{X}_D^b)^2}. \quad (2)$$

The magnitude of \mathbf{X}_D represents the total spectral difference between two images. The greater value of ρ , the higher probability of change. The most important step of CVA is to determine a threshold for the magnitude image to distinguish the changed pixels from unchanged ones. Otsu’s [49], expectation maximization (EM) [25] and other thresholding methods have all been proven to be effective in addressing this problem. However, sometimes the single threshold method might be inappropriate for pixels whose magnitude values were close to the predefined threshold [42]. These pixels belong to an uncertain region because they have a higher probability of misclassification than those far from the predefined threshold [14], [15]. For instance, in this uncertain region, if a pixel’s magnitude was greater than the threshold, it would be detected as a changed pixel by the use of a single threshold method. However, it may be actually unchanged because its spectral signatures were similar in bitemporal images [37]. This phenomenon might be caused by the radiometric difference of bitemporal images. In addition, there have been a number of methods that combine CVA and



Fig. 1. Flowchart of the proposed methodology.

other features to improve the CD accuracy, e.g., texture, spectral angle information. Most of them were applied to a whole image. Therefore, in this article, we proposed a binary CD method based on uncertain area analysis and spectral similarity measure.

Let us consider the distribution of the magnitude ρ in CVA. We assume that the magnitude value in the difference image \mathbf{X}_D can be modeled as a Gaussian mixed model [25]. Generally, a threshold value T was used to distinguish the changed from unchanged pixels. According to idea of uncertain area, the pixels were divided into the following three cases according to the magnitude value and threshold T :

Case A: $\rho < T_u$, unchanged pixels

Case B: $\rho > T_c$, changed pixels

Case C: $T_u \leq \rho \leq T_c$, uncertain pixels

$$T_u = (1 - \alpha) * T \quad (3)$$

$$T_c = (1 + \alpha) * T \quad (4)$$

where T_u and T_c were the thresholds of unchanged and changed pixels, respectively. The magnitude value located within $[(1 - \alpha) * T, (1 + \alpha) * T]$ was defined as uncertain area. The parameter α belongs to $[0, 1]$, which represents the percentage around threshold T . Parameter α plays an important role in identifying pixel's label and can result in different CD maps. A large α will expand the range of uncertain area, whereas a small α will narrow the scope. In a real application, the parameter α can be determined empirically or by the trial-and-error method.

First, the threshold value T was initially determined by the use of k -means based on the magnitude vector of CVA. The k -means

method is a special case of the Gaussian mixture model, which assumes that both the changed and unchanged classes follow a Gaussian distribution. Therefore, the parameter of the number of classes in k -means was set as 2. The standard deviations, mean values, and probability density functions of the changed and unchanged pixels can be calculated according to the initial binary change map, which is generated by the k -means method. Then, the initial threshold T can be calculated by the use of the following quadratic equation [25]:

$$(\sigma_n^2 - \sigma_c^2) T^2 + 2(\mu_n \sigma_c^2 - \mu_c \sigma_n^2) T + \mu_c \sigma_n^2 - \mu_n \sigma_c^2 - 2\sigma_n^2 \sigma_c^2 \ln \left[\frac{\sigma_c P(w_n)}{\sigma_n P(w_c)} \right] = 0 \quad (5)$$

where σ_n and σ_c were the standard deviations, μ_n and μ_c were the mean values, and $P(w_n)$ and $P(w_c)$ were the probability of unchanged and changed pixels, respectively. It should be noted that the other thresholding methods can also be used to determine the threshold. However, the parameter α might be selected differently. In general, the more accurate the threshold is, the smaller the uncertainty area, and therefore the smaller the parameter α . In addition, in order to investigate the impact of random initialization on the cluster result of k -means, we conducted k -means 20 times independently for each dataset, and the initial threshold for each experiment was calculated. The experiment result indicated that the impact of random initialization in k -means is very limited.

Second, the parameter α was determined. In this article, α was tested from 5% to 40% with a step of 5%. α located within this area may provide a reasonable range for analyzing. The

value α that produced the best binary CD accuracy was chosen. Then, a ruled-based method that combined magnitude and spectral angle information was applied to the uncertain pixels to determine whether a pixel was changed or not. Only the pixel with magnitude greater than T and spectral angle greater than T_θ was considered as changed. Otherwise, they were considered as no changed. T_θ was the threshold of spectral angle mapper, which can be determined by the trial-and-error method. The final binary map was then generated.

B. Source Image Classification Based on Active Learning

Obtaining training samples is a time-consuming task. It is often difficult to obtain satisfactory classification result with limited training samples, especially in HS remote sensing images [50]. In this work, only a small amount of training samples were available for the source image. In order to obtain desirable classification result, active learning was introduced in the proposed methodology. Active learning is a semi-supervised machine learning method, with the goal to select the most informative training samples. A general active learner can be modeled as a quintuple $(C, Q, S, L, \text{ and } U)$ [51], [52]. C is a classifier, which is trained on the labeled training samples in the training set L . Q is a query function used to select the most informative samples from the unlabeled sample pool U . S is a supervisor that can assign the true class label to unlabeled samples. Let us assume that D be a set of all pixels in the source image (X_1) , and D_L is a set of all the labeled training samples. $D_L = \{x_{i,j}, y_k\}$, where $x_{i,j}$ is the location of the pixel with label y_k . The goal of the active learning algorithm is to select $x_{i,j}$ from $D_{UL} = D \setminus D_L$ and add it to the training set $D_L^+ = D_L \cup (x, y)$. The process of identifying $x_{i,j}$ and adding it to D_L is repeated according to a user-specified number of iterations. Different criteria and query functions can be adopted to select the most informative samples, e.g., by uncertainty sampling strategy [53], [54], by selecting more informative data that optimizes the expected gain [55], by selecting the examples that minimize the expected error of the model [56]. In this article, an uncertainty sampling strategy was used, and it was a simple and popular query strategy. In the uncertainty sampling method, the most informative samples were considered to be those on which the classifier was the least certain of their classification. The objective function was defined as conditional entropy [53]

$$x^* = \arg \max_{x \in U} - \sum_{y \in Y} P(y|x_i) \log(P(y|x_i)) \quad (6)$$

where $P(y|x_i)$ was the probability that samples x_i with label y . When the probability of a sample being classified into different categories was the same, the entropy was the largest, which means the uncertainty of this sample was the largest. The active learning based on uncertainty sampling method mainly consists of three steps as follows. First, the original training samples and unlabeled samples were initialized. Then, a random forest classifier was used to train the classifier with the original samples D_L . After that, the uncertainties of the unlabeled samples were calculated and the most uncertain samples were selected according

TABLE I
DESCRIPTION OF ACTIVE LEARNING ALGORITHM

Input: original training samples D_L
Step 1. Initialize the original labeled training samples and unlabeled samples, where $D_L = \{x_{i,j}, y_k\}$, $D_{UL} = D \setminus D_L$
Step 2. Train the classifier C with the original training samples D_L
Step 3. Select $x_{i,j}$ from the unlabeled pool D_{UL} and calculated the uncertainty of $x_{i,j}$
Step 4. Select the most uncertain samples according to the score of uncertainty and the user-defined number of uncertainty samples
Step 5. Manual label the most uncertainty samples
Output: the most informative samples

to their entropy values. Table I lists the description of the active learning process.

C. Target Image Classification Based on Improved Transfer Learning

Transfer learning techniques or domain adaption (DA) methods in transfer learning; have been introduced in the CD literature [45]. DA methods define strategies that use the available information in the source domain to classify the target domain, in which no prior information is available [51], [57]. Transfer learning provides a good solution for the cases when only a small amount of training samples are available for one of the single-time image in multiple CD.

Traditional transfer learning based CD methods transfer the samples by random selection. However, the label of randomly selected samples might be wrong due to classification error of the source image. Therefore, in this article, we proposed an improved transfer learning method by considering pure pixel index (PPI) and the spectral similarity measure. By selecting the pure pixels that are the most similar to the training samples of the source image, the correctness of transfer learning samples would be increased, which may improve the classification accuracy of the target image and the final CD result. Fig. 2 shows the flowchart of the proposed transfer learning step. First, “noise whitening” and dimensionality reduction were performed by using minimum noise fraction (MNF) transform in the target image. The goal of MNF transformation was to reduce the dimensionality and ease computational complexity as well as to remove the noise of HS images. Then, PPI was applied to the transformed image to find the purest pixels. PPI algorithm is based on a convex geometry that finds the data vectors having minimal or maximal orthogonal projections toward certain directions. The pixel having more count on the extreme position, which is calculated as a digital number (DN), is considered as pure pixel. The larger DN value of the pixel, the more pure the pixel is. In the experiment, the PPI was calculated in ENVI 5.3 software with the setting of two parameters: number of iterations and threshold factor. Number of iterations was used to set the number of data projected to the random skewers. The threshold

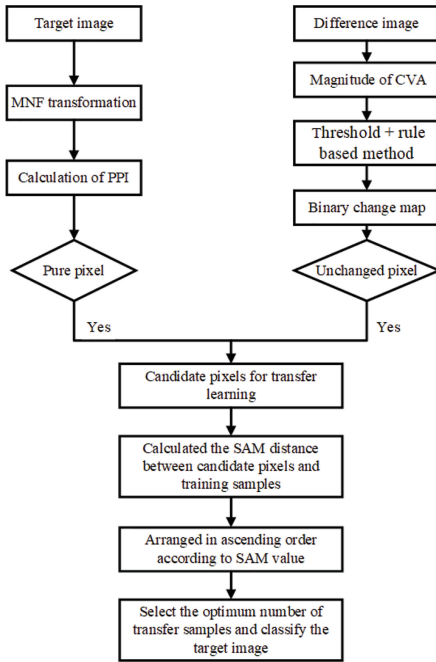


Fig. 2. Flowchart of improved transfer learning.

factor was used to distinguish the pure pixels from mixed ones. A relatively large threshold will generate more extreme pixels, which will contain more mixed pixels. In the experiment, the maximum number of iteration was set 1000 according to experience. The threshold was determined by manual selection. After that, the unchanged pure pixels were selected as the candidate pixels for transfer learning. The candidate pixels were arranged in ascending order according to the spectral angle distance to the training samples of the source image. Through this mechanism, the priorities were given to those the most similar to the training samples. The target image was then classified using a different number of candidate pixels and the accuracy was evaluated. The number that produced the highest classification accuracy was selected. By doing this the training samples of the target image can be optimized and the classification accuracy of the target image can be improved.

D. Postclassification Comparison Based Multiple CD

The training samples of the source image were expanded by the use of active learning, whereas the training samples of the target image were optimized by the proposed improved transfer learning method. Then, the source and target images were classified by SVM classifier. Finally, postclassification comparison was implemented to generate the final multiple CD map.

III. STUDY AREA AND DATASET

A. Simulated HS Dataset

The first dataset is made of a real single-time HS image acquired by the ROSIS optical sensor over the urban area of the University of Pavia, Italy. The size of this image is 610×340

pixels, with 103 spectral channels in the wavelength range of 430–860 nm [58]. The dataset contains nine classes including asphalt, meadows, gravel, trees, metal sheets, bare soil, bitumen, bricks, and shadows. Fig. 3(a) shows a false color composite of the ROSIS scene. By taking advantage of the available ground truth data, we simulated and generated a changed image (considered as X_2) based on the source image (considered as X_1) according to the following steps.

- 1) Nine tiles were extracted from the source image X_1 , which correspond to four land-cover classes.
- 2) The extracted tiles were assigned different land-cover classes to simulate the changes. The pixel values of extracted tiles were randomly selected according to the ground truth data.
- 3) A small constant bias value was applied to X_2 to simulate the difference of light condition.
- 4) White Gaussian noise was added to X_2 by setting an SNR equal to 20 dB [13].

Finally, eight change types were simulated, which were changes from meadows to asphalt, meadows to gravel, bare soil to meadows, meadows to bitumen, meadows to bare soil, gravel to bitumen, bitumen to meadows, and meadows to brick. The reference change map was then generated. The false color composite of X_2 is shown in Fig. 3(b). The ground truth of X_1 and the reference change map are shown in Fig. 3(c) and (d), respectively.

B. Real HS Dataset

The second dataset used in this study includes two real HS images acquired by the Hyperion sensor mounted on board the EO-1 satellite. The Hyperion images were acquired on May 3, 2006 and April 23, 2007, in an agricultural area in Yancheng, Jiangsu Province, China. The original images contain 242 spectral channels, with a wavelength range from 350 to 2580 nm. The spectral resolution is 10 nm and the spatial resolution is 30 m. The preprocessing applied to the images included stripe repair, atmospheric corrections, and radiometric calibrations. The uncalibrated bands, overlapped redundant bands, and water absorption bands were also removed. Finally, 179 preprocessed bands (bands 8–57, 79–120, 123–125, 128–166, 179–223) were selected for the CD task. Images were then coregistered with a residual error of less than 0.5 pixels. Two subsets were extracted from the whole images as the experimental data. The land cover and land-cover changes differ sharply between the two subsets, which presents a good case study for testing the proposed approach in different regions.

1) *Subset 1*: The first subset is located in a wetland area, with a size of 357×226 pixels. It contains four main land-cover classes: water, crop, bare land, and wet soil. Water and crop correspond to the dark and red areas in the images. The spectral signature of bare land exhibits high reflectance corresponding to the high brightness areas in the images. In total, 12 types of changes occurred during two time periods. The majority of changes are C_2 (water to bare land), C_3 (water to soil), C_8 (bare land to crop), and C_{10} (soil to water). Because there was no ground truth map for accuracy assessment, a random sampling

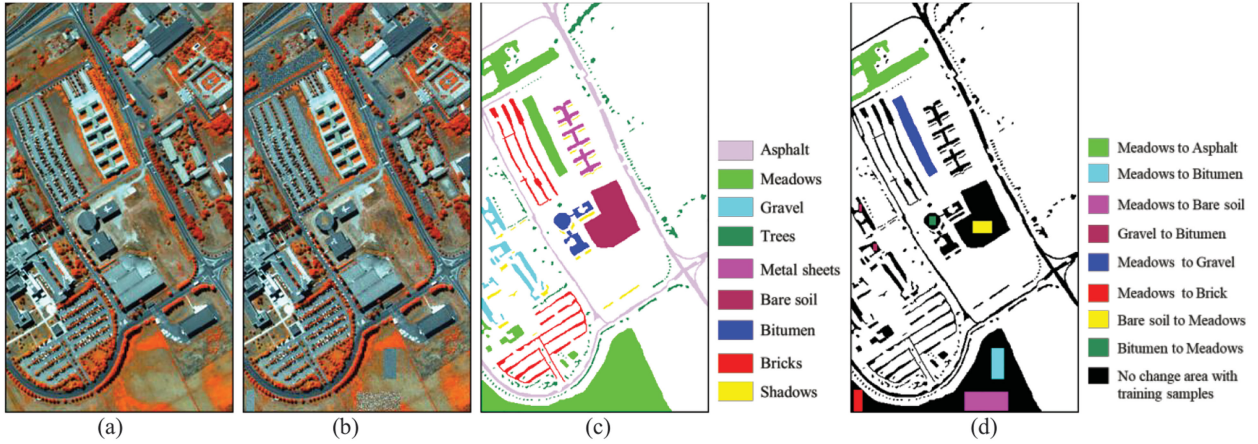


Fig. 3. False color composite of the ROSIS Pavia scene (R: band 80, G: band 70, B: band 60). (a) Real HS image acquired by the ROSIS sensor in Pavia, Italy. (b) Simulated HS dataset computed with an additive white Gaussian noise. (c) Ground truth of X_1 . (d) Change reference map (eight changes in different colors and no-change class in white color).

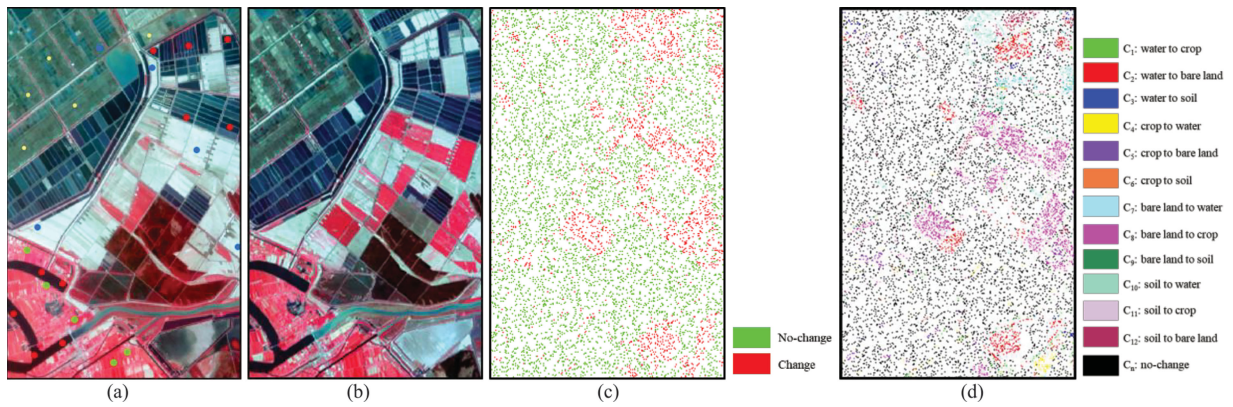


Fig. 4. Subset 1 in real HS dataset. (a) and (b) False-color composite images (R: 752.43 nm, G: 650.67 nm, B: 548.92 nm) acquired in 2006 and 2007, respectively. (c) and (d) Distribution check points for binary and multiple CD accuracy assessment, respectively.

and stratified random sampling method was applied to the binary and multiple change map to generate the check points. Totally 20% of the check points were selected and the change types were determined through visual interpretation. The distribution of training samples and check points for the binary and multiple CD are shown in Fig. 4(a), (c), and (d), respectively.

2) *Subset 2*: The second subset is in an agricultural area, with a size of 405×170 pixels. The area is characterized by three kinds of land-cover classes: water, crop, and bare land. The initial training samples of the source image can be seen in Fig. 5(a). The main changes are from crop to bare land (C_4) and bare land to crop (C_6). The reference classification maps of the source and target images were produced by visual inspection, and the multiple change map was then generated by direct comparison [see Fig. 5(c)].

IV. EXPERIMENTAL RESULTS

A. Binary CD Results

The threshold values of different binary CD methods are presented in Table II. T_{Prop} is the initial threshold of the proposed method, $T_{\text{Otsu's}}$ and T_{EM} are the thresholds of Otsu's

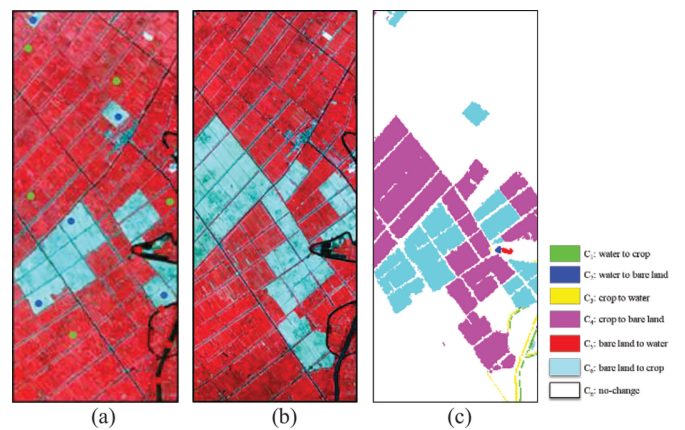


Fig. 5. Subset 2 in real HS dataset. (a) and (b) False-color composite (R: 752.43 nm, G: 650.67 nm, B: 548.92 nm) acquired in 2006 and 2007, respectively. (c) Multiple change reference map.

and EM algorithm, respectively. Fig. 6 shows the binary CD map obtained in the simulated HS dataset (row 1) and real HS datasets (rows 2 and 3). Fig. 6(a)–(e) indicates the results obtained by the proposed method, Otsu's thresholding method, EM algorithm,

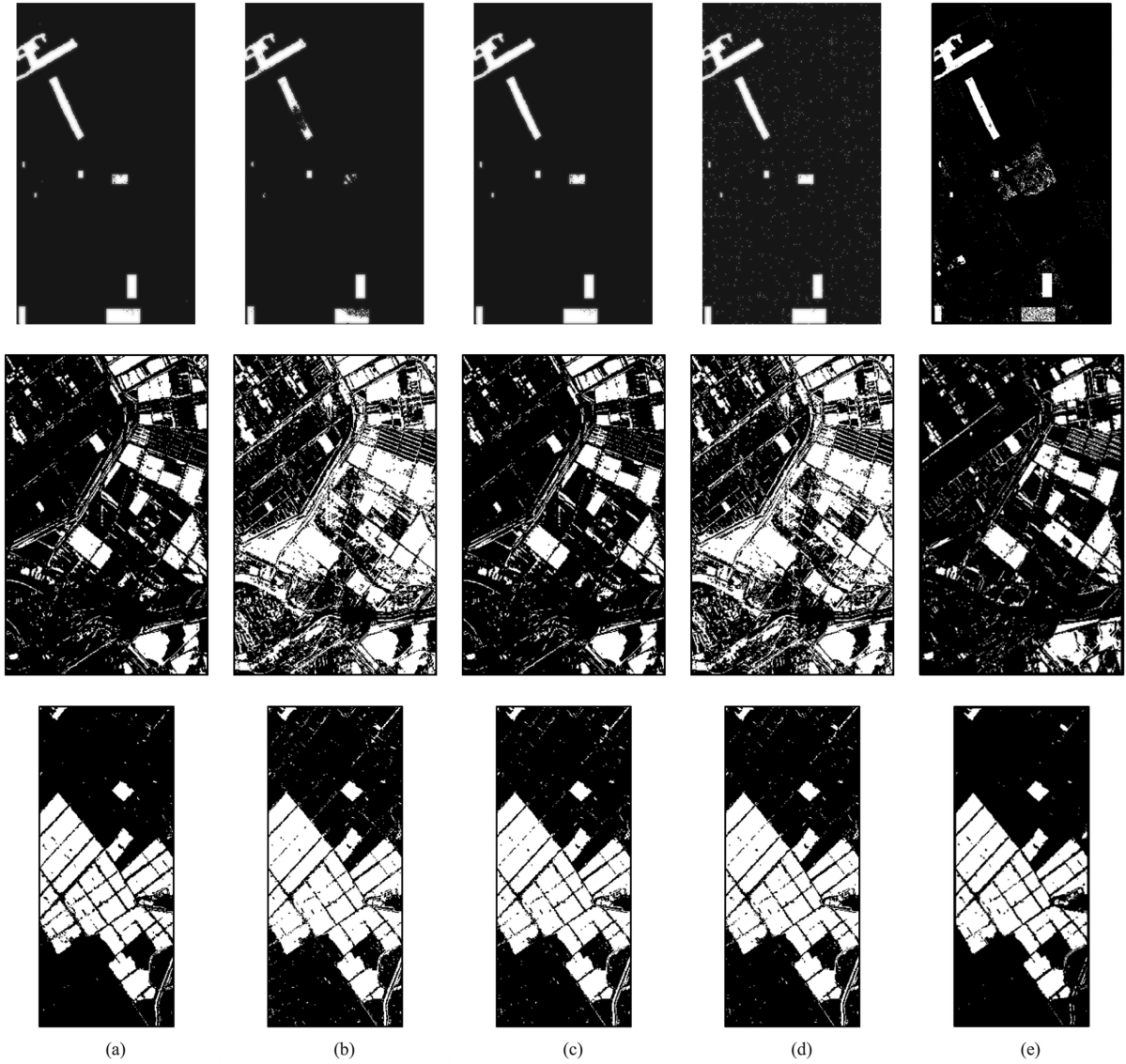


Fig. 6. Binary CD results in simulated HS dataset (row 1) and real Hyperion datasets (rows 2 and 3). (a)–(e) Results obtained by the proposed method, Otsu's, EM, K -means, and CVAPS, respectively.

TABLE II
THRESHOLDS OBTAINED BY USING DIFFERENT METHODS

Data sets	Thresholds of different methods		
	T_{Prop}	$T_{\text{Otsu's}}$	T_{EM}
Simulated HS data set	0.1749	0.4510	0.2203
Subset 1 in real HS data set	0.8939	0.7608	1.1120
Subset 2 in real HS data set	0.8157	0.7647	0.8061

K -means, and CVA in posterior probability space (CVAPS) method [39], respectively. In the CVAPS method, the bitemporal images were first classified by random forest classifier, and the posterior probabilities on date T_1 and T_2 were also calculated.

Then, the change magnitude based on posterior probability difference can be obtained. After that, a new image that included the magnitude of the posterior probability differences, the classified image on date T_1 and the classified image on date T_2 can be generated. By the use of the real change/no-change information of training samples, the new image can be classified to obtain the binary CD result. Table III lists the accuracy statistics of different binary CD methods.

The uncertain area parameter α was defined as 25% for three datasets according to parameter analysis in Section V. For the simulated HS dataset, the overall accuracy and Kappa of the proposed method were 99.73% and 0.9921. EM algorithm produced a similar result as the proposed method. However, Otsu's thresholding performed an obvious underextraction. The overall accuracy and Kappa by using Otsu's thresholding method were 96.35% and 0.8836, respectively. K -means exhibited similar accuracy with the Otsu's method. However, it suffered more noise. The overall accuracy of the CVAPS method was 92.27%, which

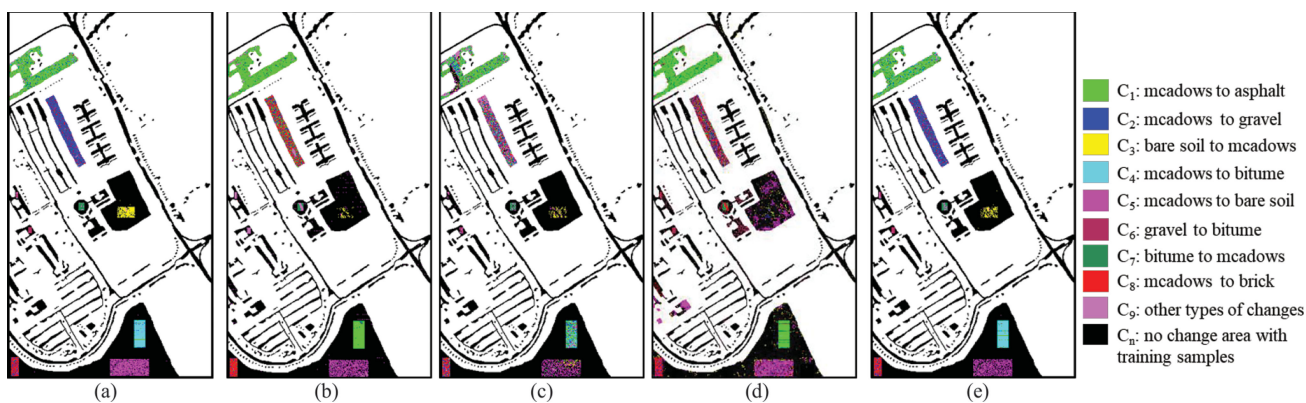


Fig. 7. Multiple CD map obtained on simulated HS dataset by: (a) the proposed method, (b) PCC_RF, (c) PCC_MTTD, (d) CVAPS+PCC, and (e) CDTL, respectively.

TABLE III
BINARY CD ACCURACY BY USING DIFFERENT METHODS

Binary CD method	Simulated HS data set		Subset 1 in real HS data set		Subset 2 in real HS data set	
	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa
Proposed method	99.73	0.9921	93.88	0.8379	96.89	0.9299
Otsu's	96.35	0.8836	75.35	0.4989	95.72	0.9063
EM	99.73	0.9921	93.97	0.8405	96.29	0.9183
K-means	96.77	0.9071	87.55	0.7075	96.35	0.9194
CVAPS	92.27	0.7806	89.62	0.7236	96.78	0.9273

was significantly lower than the other methods. An obvious over extraction can be observed in the central of the image. For subset 1, the accuracy was quantitatively evaluated by the check points that were generated by random sampling method. EM algorithm produced the best binary CD result. The overall accuracy of the proposed method was 93.88%, with the Kappa of 0.8379. The changes were over extracted by using Otsu's and *k*-means. However, by taking advantage of uncertain area analysis, the over extraction was effectively suppressed and the accuracy was improved 18.53% and 6.33% compared with Otsu's thresholding method and *k*-means, respectively. CVAPS method generated satisfactory result, with the overall accuracy of 89.62%. Some omission can be observed in the top right corner and lower right corner of the image. For subset 2, the proposed method produced the best binary CD result. The noise was effectively suppressed by using the proposed method and CVAPS, compared with the other three ones. The overall accuracy and Kappa of the proposed method were 96.89% and 0.9299. Compared to Otsu's, EM, *k*-means, and CVAPS method, the accuracy was improved about 1.17%, 0.60%, 0.54%, and 0.11% by using the proposed method, respectively.

TABLE IV
MULTIPLE CD ACCURACY BY USING THE PROPOSED METHOD

Data	OA (%)	Kappa
Simulated HS data set	92.14	0.8008
Subset 1 in real HS data set	76.10	0.6438
Subset 2 in real HS data set	95.15	0.9071

B. Multiple CD Results

Fig. 7 and Table IV show the results of multiple CD map on the simulated HS dataset. Eight types of changes were simulated. However, more than eight types of changes were detected by using the proposed method due to classification error. For the sake of simplicity, the changes that did not appear in the multiple reference change map were displayed in Orchid [see Fig. 7(a)]. The overall accuracy and Kappa were calculated by using the reference change map. The overall accuracy of the proposed method was 92.14% with the Kappa of 0.8008. Most of the changes were correctly detected, e.g., C_1 , C_3 , C_4 , C_5 , C_6 , and C_7 . The change type C_2 was confused with C_8 due to classification error.

Fig. 8 shows the multiple change map of subset 1. Table V lists the confusion matrix. The overall accuracy was 76.10% with the Kappa of 0.6438. In total, 12 types of changes were detected during two time periods. The main changes detected were C_2 (water to bare land), C_8 (bare land to crop), and C_{10} (soil to water). The change types C_1 , C_2 , C_5 , C_8 , C_{10} , C_{11} , and C_{12} obtained satisfactory results, with class detection accuracy of 94.05%, 90.57%, 97.22%, 98.91%, 83.97%, 84.78% and 93.43%, respectively. However, the accuracy of change types C_3 , C_6 , and C_9 was relatively low, which was 31.36%, 48.78%, and 52.22%, respectively. The low accuracy might be caused by the following reasons: 1) the changed pixels were detected as unchanged (C_6 and C_9), and 2) misclassification of the target images (C_3). However, the main change types (C_2 , C_8 , and C_{10})

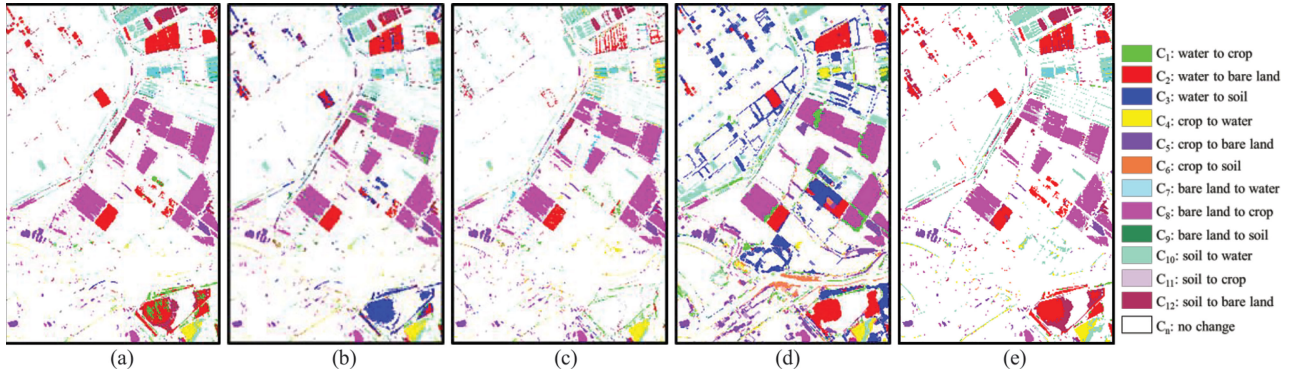


Fig. 8. Multiple CD map obtained on subset 1 by: (a) the proposed method, (b) PCC_RF, (c) PCC_MTTD, (d) CVAPS+PCC, and (e) CDTL, respectively.

TABLE V
CONFUSION MATRIX OF MULTIPLE CD IN SUBSET 1

Detected data (pixels)	Reference data (pixels)													Total (pixels)
	C _n	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	C ₁₂	
C _n	4404	2	0	0	5	0	7	3	6	28	49	2	2	4508
C ₁	235	79	39	3	0	0	0	0	0	0	0	1	1	358
C ₂	124	2	663	53	0	0	0	0	2	0	0	0	13	857
C ₃	270	0	0	31	0	0	3	0	0	0	0	0	0	304
C ₄	16	0	0	0	90	0	3	0	0	0	0	0	0	109
C ₅	140	0	1	0	3	140	2	0	0	0	0	0	0	286
C ₆	7	0	0	0	1	0	20	0	0	0	0	0	0	28
C ₇	1	0	0	0	7	0	0	74	1	13	17	0	0	113
C ₈	62	0	0	0	0	0	2	2	1176	1	0	8	1	1252
C ₉	24	0	0	0	0	0	2	9	0	47	2	0	0	84
C ₁₀	404	0	0	0	17	0	2	7	1	1	393	0	0	825
C ₁₁	190	1	1	0	0	0	0	0	2	0	0	78	1	273
C ₁₂	482	0	28	11	2	4	0	0	1	0	7	3	256	794
Total	6359	84	732	98	125	144	41	95	1189	90	468	92	274	9791
Accuracy (%)	69.26	94.05	90.57	31.63	72.00	97.22	48.78	77.89	98.91	52.22	83.97	84.78	93.43	—
OA (%)	76.10													

were all effectively detected and high accuracy was achieved. Fig. 9 shows the multiple change map of subset 2. The confusion matrix of the proposed method was listed in Table VI. All six change types were effectively detected by using the proposed method, even the small changes, e.g., C₁ (water to crop), C₂ (water to bare land), C₃ (crop to water), and C₅ (bare land to water). The overall accuracy was 95.15% with the Kappa of 0.9071. The accuracy of the main change types C₄ and C₆ was 98.56% and 98.16%, respectively. The change type C₂ showed relatively low accuracy. There were 68 pixels changed from water to bare land (C₂). However, only 43 pixels were correctly detected. Most of the pixels were detected as unchanged (C_n)

and change from crop to bare land (C₄) due to the contamination of vegetation.

The proposed method was also compared with four state-of-the-art methods, e.g., postclassification comparison based on random forest (PCC_RF) [2], postclassification comparison based on multithreshold and decision tree (PCC_MTTD) [46], the method of fusion CVA in posterior probability space and postclassification comparison (CVAPS+PCC) [39] and CDTL method [45]. These methods were described as follows.

- 1) In PCC_RF, the training samples for the source image were the initial training samples of the proposed method. The training samples of the target image were randomly

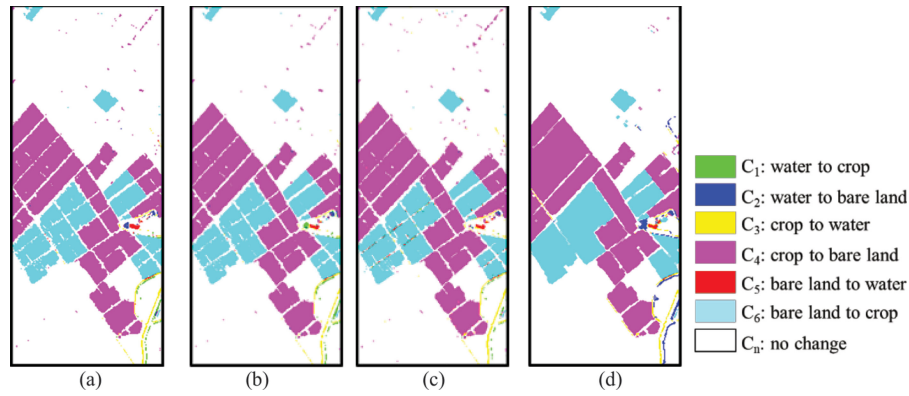


Fig. 9. Multiple CD map obtained on subset 2 by: (a) the proposed method, (b) PCC_RF, (c) PCC_MTTD, and (d) CVAPS+PCC, respectively.

TABLE VI
CONFUSION MATRIX OF MULTIPLE CD IN SUBSET 2

Detected data (pixels)	Reference data (pixels)							Total (pixels)
	C_n	C_1	C_2	C_3	C_4	C_5	C_6	
C_n	42617	22	18	16	170	6	128	42977
C_1	106	122	0	0	0	0	5	233
C_2	15	5	43	0	9	0	0	72
C_3	607	3	1	236	34	8	3	892
C_4	1401	1	6	2	14784	0	5	16199
C_5	118	3	0	2	0	67	2	192
C_6	634	1	0	0	3	7	7640	8285
Total	45498	157	68	256	15000	88	7783	68850
Accuracy (%)	93.67	77.71	63.24	92.19	98.56	76.14	98.16	—
OA (%)	95.15							

selected and the labels were interpreted by visual interpretation. Then the images were classified by random forest classifier and postclassification comparison was conducted to generate the multiple change map.

- 2) In PCC_MTTD, the thresholds of each land-cover types were determined by manual selection. The training samples of the target image were selected through randomly stratifying samples from the unchanged area. Different numbers of transferred samples were tested and the one that produced the best accuracy was finally selected. Then, only the changed pixels of the target image were classified by using Data Mining Tools See5 and C5.0 (See5/C5.0). Thus for comparison purpose, the unchanged pixels in PCC_RF, CVAPS+PCC, CDTL, and the proposed method were kept the same and the overall accuracy was calculated.
- 3) In CVAPS+PCC, the bitemporal images were classified by random forest classifier and then Markov random field (MRF) was applied to the classified images to eliminate the “salt and pepper” noise. Then, CVAPS and PCC were combined to generate the multiple change map.
- 4) In CDTL, the EM algorithm was first applied to generate the binary change map. Then, the unchanged training

samples of the source image were transferred to the target image, and the active learning was used to expand the training set. Finally, the bitemporal images were classified by the SVM method, and the postclassification comparison was conducted to generate the multiple change map. For subset 2 in real HS dataset, bare land in the source image were all changed into crop, which resulted in that no bare land sample can be transferred. As a result, the CDTL method was only conducted on the simulated HS dataset and subset 1.

The multiple CD maps by using the above methods are shown in Figs. 7–9. Table VII lists the accuracy comparison. From the figures and table, we can find that the following statements hold.

- 1) For the simulated HS dataset, all of the methods produced more than eight types of changes due to classification error. Similarly, the change types that did not appear in the reference multiple change map were displayed in Orchid (see Fig. 7). The overall accuracy and Kappa were calculated by using the reference change map. The proposed method produced the best multiple CD result with the overall accuracy of 95.83%, with the Kappa of 0.8818. Most of the changes were correctly detected. The accuracy of CDTL was slightly lower than that of the proposed method, with

TABLE VII
MULTIPLE CD ACCURACY BY USING DIFFERENT METHODS

Method	Simulated HS data set		Subset 1 in real HS data set		Subset 2 in real HS data set	
	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa
PCC_RF	90.90	0.7293	90.45	0.8284	96.79	0.9347
PCC_MTD	89.88	0.6599	85.85	0.7261	96.73	0.9341
CVAPS+PCC	85.01	0.6177	86.94	0.7486	96.44	0.9299
CDTL	93.62	0.8138	91.64	0.8507	—	—
Proposed method	95.83	0.8818	92.95	0.8741	96.97	0.9386

the overall accuracy of 93.62%. The multiple CD accuracy of PCC_RF was 90.90%. Some of the change type C_2 (meadows to gravel) were detected as C_8 (meadows to bricks) and C_1 (meadows to asphalt). In addition, some of the change type C_4 (meadows to bitumen) were detected as C_1 (meadows to asphalt). The multiple CD accuracy of PCC_MTD was 89.88%. The change type C_2 (meadows to gravel) was detected as C_5 (meadows to bare soil). The multiple CD accuracy of CVAPS+PCC method was 85.01%. Most of the change type C_2 (meadows to gravel) and C_4 (meadows to bitumen) were wrongly detected due to classification error. In addition, the change type C_3 was detected as unchanged.

- For subset 1, the proposed method produced the best multiple CD result. The overall accuracy was 92.95%, which was significantly higher than that of PCC_RF, PCC_MTD, and CVAPS+PCC methods. The overall accuracy of CDTL was 91.64%, with the kappa coefficient 0.8570. The significant difference of the final multiple CD was located in the lower right of the image. The actual change for the left part of this area was changes from water to bare land. Most of these changes were correctly detected by using CVAPS+PCC, CDTL, and the proposed method. PCC_MTD has the problem of under extraction, whereas the fusion of CVAPS and PCC methods has the problem of over extraction.
- For subset 2, all of the methods produced comparable multiple CD results. The overall accuracy of the proposed method was 96.97%, a little higher than that of the other three methods. PCC_RF, PCC_MTD, and the proposed method suffered more noise than CVAPS+PCC method. This was because CVAPS+PCC method used MRF to eliminate the “salt and pepper” noise.

The efficiency of the comparison methods was also compared (see Table VIII). The experiment was carried out by using MATLAB R2016b, on an Intel Core i7-4770 Quad-core 3.40-GHz personal computer with 16 GB of RAM. The comparison results indicated that PCC_RF and PCC_MTD were more efficient than CVAPS+PCC, CDTL, and the proposed method. However, the accuracy of CDTL and the proposed method was significantly higher than the other three ones. CVAPS+PCC method

TABLE VIII
TIME COST OF DIFFERENT METHODS

Methods	Time costs (Seconds)		
	Simulated HS data set	Subset 1 in real HS data set	Subset 2 in real HS data set
PCC_RF	19	6	4
PPC_MTD	17	15	10
CVAPS+PCC	15301	123	68
CDTL	4972	516	—
Proposed method	6442	519	475

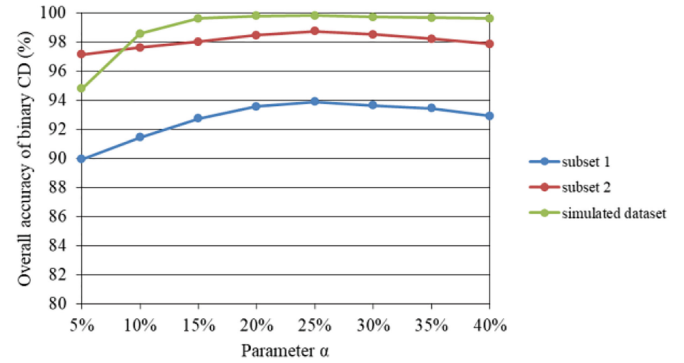


Fig. 10. Overall accuracy of binary CD by defining different values of α .

showed low both in efficiency and accuracy. The proposed method took the longest time, but with the best accuracy. The time cost of the proposed method was mainly occupied by the active learning and classification of the target image. The other factor affecting the time cost of the proposed method was the SVM classifier. Compared with the random forest and decision tree classifiers, the SVM process takes more time to train well the model for obtaining better classification result.

V. PARAMETER AND PERFORMANCE ANALYSIS

A. Parameter Analysis

1) *Uncertain Area Analysis*: In order to select the optimal parameter α of uncertain area that would produce the best binary CD accuracy, different α was tested and the accuracy was assessed. In the experiment, the parameter α was selected as 5%, 10%, 15%, 20%, 25%, 30%, 35%, and 40% for three datasets. The overall accuracy of binary CD with respect to different α is presented in Fig. 10. As shown in the figure, the accuracy of three datasets exhibited similar tendencies: it first increased and then gradually decreased with the increase of α . The best accuracy was achieved when $\alpha = 25%$ for the three datasets. This indicated that the magnitude value located within $[0.75 * T, 1.25 * T]$ has the highest probability of misdetection. By using the proposed method, the accuracy of binary CD can be improved.

2) *Number of Samples in Improved Transfer Learning*: The number of samples N in transfer learning is another important parameter in the proposed approach. We therefore carried out experiment to select the optimal number of N . Different numbers

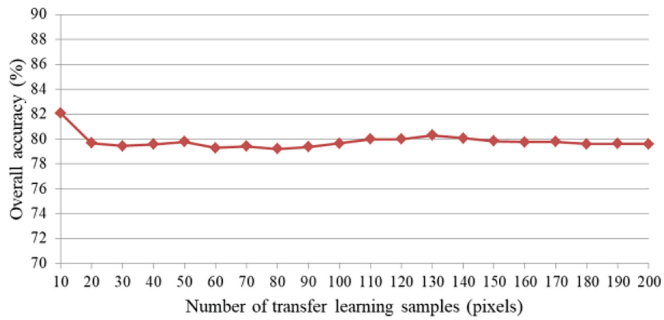


Fig. 11. Target image classification accuracy in subset 1.

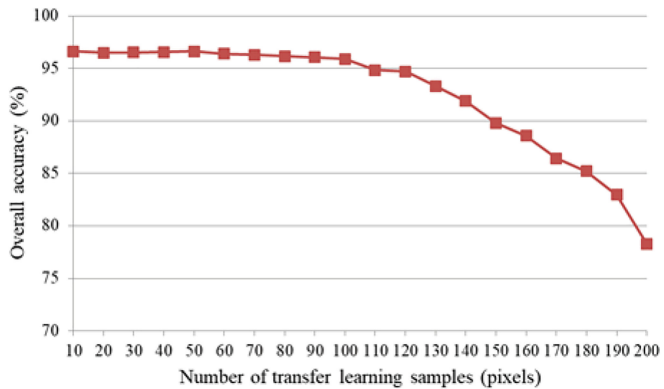


Fig. 12. Target image classification accuracy in subset 2.

of N were tested and the accuracy was assessed. For the simulated HS datasets, there were small amounts of pure pixels with reference labels for most of the classes. Therefore, the transfer samples were selected from the unchanged pixels according to the spectral angle distance to training samples of the source image. For the real Hyperion dataset, the numbers of transfer learning samples N were varied from 10 to 200 with a step size of 10. Figs. 11 and 12 show the relationship between classification accuracy and the number of transfer learning samples for subsets 1 and 2, respectively. For subset 1, the classification accuracy tends to be stable with the increasing number of transfer learning samples. However, for subset 2, the classification accuracy kept stable at first and then gradually decreased with the increase of N . This was because in the proposed improve transfer learning method, priority was given to the pixels that the most similar to the training samples. The best classification accuracy was achieved when $N = 10$ for subsets 1 and 2, with an accuracy of 82.10% and 96.64%, respectively. This indicated that by using the optimization mechanism of the proposed improved transfer learning method, a satisfactory result can be obtained with the limited number of samples.

B. Performance Analysis

1) *Contribution of Uncertain Area Analysis*: In order to investigate the contribution of uncertain area, uncertain area analysis was excluded from the binary CD task for comparison. Table IX lists the comparison result. As can be seen from the table, the accuracy of binary CD was significantly improved

TABLE IX
CONTRIBUTION OF UNCERTAIN AREA ANALYSIS

Dataset	Binary CD with uncertain area analysis		Binary CD without uncertain area analysis	
	OA (%)	Kappa	OA (%)	Kappa
Simulated HS data set	99.73	0.9921	96.77	0.9071
Subset 1 in real HS data set	93.88	0.8379	87.55	0.7075
Subset 2 in real HS data set	96.89	0.9299	96.35	0.9194

TABLE X
CONTRIBUTION OF ACTIVE LEARNING AND IMPROVED TRANSFER LEARNING

Data sets	OA (%)			
	Proposed method	Without active learning	Without improved transfer learning	Without all
Simulated HS data set	92.14	91.78	90.17	90.09
Subset 1 in real HS data set	76.10	73.48	72.18	70.91
Subset 2 in real HS data set	95.15	95.38	88.58	76.10

for all the three datasets, especially for simulated HS datasets and subset 1. The accuracy of the binary CD was improved from 96.77% to 99.73% and 87.55% to 93.88% for simulated HS datasets and subset 1, respectively. Overextraction has been effectively eliminated by using uncertain area analysis.

2) *Contribution of Active Learning*: Active learning was used to improve the classification accuracy of the source image, which would have an impact on transfer learning and the final multiple CD results. In order to investigate the contribution of active learning, active learning was excluded from the proposed method, whereas uncertain area analysis and improved transfer learning were retained. Table X lists the comparison result. Compared with the method without active learning, the accuracy was improved by 0.36% and 2.62% for the simulated HS datasets and subset 1, respectively. For subset 2, the multiple CD accuracy without active learning was a little higher than that of the proposed method. This was because the land-cover classes were relatively simple in subset 2. Satisfactory result can be obtained with limited samples.

3) *Contribution of Improved Transfer Learning*: The contribution of improved transfer learning was also evaluated by comparison with transfer of the unchanged pixels by random selection. The experiment was conducted 10 times and the average accuracy was calculated as the final result. Table X lists the comparison results. Compared with the method without improved transfer learning, the accuracy was improved by 1.97%,

3.92%, and 6.57% for the simulated HS dataset, subset 1, and subset 2, respectively.

We also conducted experiment that active learning and improved transfer learning were all excluded from the proposed method. Compared with the method without all, the accuracy was improved by 2.05%, 5.19%, and 19.05% for the simulated HS datasets, subset 1, and subset 2, respectively. The improvement was significant for all of the datasets. In addition, as can be seen from Table X, for all of the three datasets, the best accuracy was obtained by using the proposed method. The comparison results indicated the effectiveness of the proposed method.

VI. CONCLUSION

In this article, a novel approach for multiple CD based on uncertain area analysis, active learning, and improved transfer learning has been proposed. The method consists of four main steps. First, unsupervised binary CD based on uncertain area analysis was performed to generate the binary change map. The source image was then classified according to active learning. After that, the classification map of the target image was obtained by using the improved transfer learning method. Finally, the multiple change map was generated by postclassification comparison. In order to obtain the best results, the parameter of the uncertain area α and the number of improved transfer learning samples N were analyzed and the contribution of uncertain area, active learning, and improved transfer learning were evaluated. In addition, four state-of-the-art methods were compared with the proposed method. Experiment results demonstrated the effectiveness of the proposed method. A number of conclusions can be made as follows.

- 1) This article proposed a novel approach for solving the multiple CD problem when only small amounts of training samples were available for a single-time image.
- 2) The parameter of uncertain area α for all of the three datasets was 25%, which indicated that the magnitude value within this area may have a higher probability of misextraction. By using uncertain area analysis, the accuracy of binary CD can be improved.
- 3) Active learning and improved transfer learning can improve the accuracy of single-time classification results, which will finally improve the multiple CD accuracy.

However, the limitation of the proposed method is that the multitemporal images are assumed to have the same land-cover types, where new land-cover types in the target image cannot be detected. This issue will be addressed in our future work.

ACKNOWLEDGMENT

The authors would like to thank the editors and the anonymous reviewers for their valuable comments on the improvement of this article. In addition, the authors also would like to thank P. Gamba for providing the Pavia University hyperspectral image and RuleQuest Research for providing the license of Data Mining Tools See5 and C5.0 (See5/C5.0).

REFERENCES

- [1] 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.
- [2] X. Li, P. Gong, and L. Liang, "A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data," *Remote Sens. Environ.*, vol. 166, pp. 78–90, 2015.
- [3] C. 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] P. Xiao, X. Zhang, D. Wang, M. Yuan, X. Feng, and M. Kelly, "Change detection of built-up land: A framework of combining pixel-based detection and object-based recognition," *ISPRS J. Photogramm. Remote Sens.*, vol. 119, pp. 402–414, 2016.
- [5] J. Deng, K. Wang, Y. Deng, and G. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, 2008.
- [6] C. He, A. Wei, P. Shi, Q. Zhang, and Y. Zhao, "Detecting land-use/land-cover change in rural–urban fringe areas using extended change-vector analysis," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 4, pp. 572–585, 2011.
- [7] X. Zhang, R. Sun, B. Zhang, and Q. Tong, "Land cover classification of the North China Plain using MODIS EVI time series," *ISPRS J. Photogramm. Remote Sens.*, vol. 63, pp. 476–484, 2008.
- [8] Z. Li, W. Shi, P. Lu, L. Yan, Q. Wang, and Z. Miao, "Landslide mapping from aerial photographs using change detection-based Markov random field," *Remote Sens. Environ.*, vol. 187, pp. 76–90, 2016.
- [9] S. Liu, M. Chi, Y. Zou, A. Samat, J. A. Benediktsson, and A. Plaza, "Oil spill detection via multitemporal optical remote sensing images: A change detection perspective," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 324–328, Mar. 2017.
- [10] Z. Qi, A. Yeh, X. Li, and X. Zhang, "A three-component method for timely detection of land cover changes using polarimetric SAR images," *ISPRS J. Photogramm. Remote Sens.*, vol. 107, pp. 3–21, 2015.
- [11] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Unsupervised multitemporal spectral unmixing for detecting multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2733–2748, May 2016.
- [12] L. Bruzzone and S. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.
- [13] S. Liu, L. Bruzzone, F. Bovolo, M. Zanetti, and P. Du, "Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4363–4378, Aug. 2015.
- [14] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.
- [15] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Hierarchical unsupervised change detection in multitemporal hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 244–260, Jan. 2015.
- [16] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.
- [17] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, Feb. 2007.
- [18] C. Wu, B. Du, and L. Zhang, "A subspace-based change detection method for hyperspectral images," *IEEE J. Sel. Topics. Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 815–830, Apr. 2013.
- [19] J. Lu, J. Li, G. Chen, L. Zhao, B. Xiong, and G. Kuang, "Improving pixel-based change detection accuracy using an object-based approach in multitemporal SAR flood images," *IEEE J. Sel. Topics. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3486–3496, Jul. 2015.
- [20] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.
- [21] X. Li, Z. Yuan, and Q. Wang, "Unsupervised deep noise modeling for hyperspectral image change detection," *Remote Sens.*, vol. 11, no. 3, pp. 258–274, 2019.
- [22] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, 2016.

- [23] A. Ertürk and A. Plaza, "Informative change detection by unmixing for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1252–1256, Jun. 2015.
- [24] Y. Bazi, F. Melgani, and H. 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.
- [25] L. Bruzzone and D. 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.
- [26] P. Du, S. Liu, P. Gamba, K. Tan, and J. Xia, "Fusion of difference images for change detection over urban areas," *IEEE J. Sel. Topics. Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1076–1086, Aug. 2012.
- [27] P. Du, S. Liu, J. Xia, and Y. Zhao, "Information fusion techniques for change detection from multi-temporal remote sensing images," *Inf. Fusion*, vol. 14, no. 1, pp. 19–27, 2013.
- [28] S. Le Hégarat-Masclé and R. Seltz, "Automatic change detection by evidential fusion of change indices," *Remote Sens. Environ.*, vol. 91, no. 3, pp. 390–404, 2004.
- [29] H. Nemmour and Y. Chibani, "Multiple support vector machines for land cover change detection: An application for mapping urban extensions," *ISPRS J. Photogramm. Remote Sens.*, vol. 61, no. 2, pp. 125–133, 2006.
- [30] C. Tarantino, M. Adamo, R. Lucas, and P. Blonda, "Detection of changes in semi-natural grasslands by cross correlation analysis with WorldView-2 images and new Landsat 8 data," *Remote Sens. Environ.*, vol. 175, pp. 65–72, 2016.
- [31] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. Fisher, "A critical synthesis of remotely sensed optical image change detection techniques," *Remote Sens. Environ.*, vol. 160, pp. 1–14, 2015.
- [32] D. Wen, X. Huang, L. Zhang, and J. Benediktsson, "A novel automatic change detection method for urban high-resolution remotely sensed imagery based on multi-index scene representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 609–625, Jan. 2016.
- [33] Y. Yuan, H. Lv, and X. Lu, "Semi-supervised change detection method for multi-temporal hyperspectral images," *Neurocomputing*, vol. 148, pp. 363–375, 2015.
- [34] F. Bovolo, "A multilevel parcel-based approach to change detection in very high resolution multitemporal images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 1, pp. 33–37, Jan. 2009.
- [35] P. He, W. Shi, H. Zhang, and M. Hao, "A novel dynamic threshold method for unsupervised change detection from remotely sensed images," *Remote Sens. Lett.*, vol. 5, no. 4, pp. 396–403, 2014.
- [36] F. Melgani and Y. Bazi, "Markovian fusion approach to robust unsupervised change detection in remotely sensed imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 457–461, Oct. 2006.
- [37] H. Zhuang, K. Deng, H. Fan, and M. Yu, "Strategies combining spectral angle mapper and change vector analysis to unsupervised change detection in multispectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 681–685, May 2016.
- [38] O. Carvalho, R. Guimarães, A. Gillespie, N. Silva, and R. Gomes, "A new approach to change vector analysis using distance and similarity measures," *Remote Sens.*, vol. 3, no. 11, pp. 2473–2493, 2011.
- [39] F. Zakeri, B. Huang, and M. Saradjian, "Fusion of change vector analysis in posterior probability space and post classification comparison for change detection from multispectral remote sensing data," *Remote Sens.*, vol. 11, no. 13, pp. 1151–1524, 2019.
- [40] L. Castellana, A. D'Addabbo, and G. Pasquariello, "A composed supervised/unsupervised approach to improve change detection from remote sensing," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 405–413, 2007.
- [41] W. Yu, W. Zhou, Y. Qian, and J. Yan, "A new approach for land cover classification and change analysis: Integrating backdating and an object-based method," *Remote Sens. Environ.*, vol. 177, pp. 37–47, 2016.
- [42] R. Zurita-Milla, L. Gómez-Chova, L. Guanter, J. Clevers, and G. Camps-Valls, "Multitemporal unmixing of medium-spatial-resolution satellite images: A case study using MERIS images for land-cover mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4308–4317, Nov. 2011.
- [43] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2196–2212, Jun. 2012.
- [44] F. Marinelli, F. Bovolo, and L. Bruzzone, "A novel change detection method for multitemporal hyperspectral images based on binary hyperspectral change vectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4931–4928, Jul. 2019.
- [45] 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.
- [46] G. Xian, C. Homer, and J. Fry, "Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1133–1147, 2009.
- [47] G. Xian and C. Homer, "Updating the 2001 National Land Cover Database impervious surface products to 2006 using Landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 114, no. 8, pp. 1676–1686, 2010.
- [48] F. Bovolo, L. Bruzzone, and M. Marconcini, "A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2070–2082, Jul. 2008.
- [49] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 285–296, 1975.
- [50] S. Rajan, J. Ghosh, and M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, Apr. 2008.
- [51] 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.
- [52] M. Li and I. Sethi, "Confidence-based active learning," *IEEE Trans. Pattern Anal.*, vol. 28, no. 8, pp. 1251–1261, Aug. 2006.
- [53] M. Sharma and M. Bilgic, "Evidence-based uncertainty sampling for active learning," *Data Mining Knowl. Discovery*, vol. 31, no. 1, pp. 1–39, 2016.
- [54] H. Xu, X. Wang, Y. Liao, and C. Zheng, "An uncertainty sampling-based active learning approach for support vector machines," in *Proc. Int. Conf. Artif. Intell. Comput. Intell.*, Washington, DC, USA, 2009, pp. 208–213.
- [55] D. MacKay, "Information-based objective functions for active data selection," *Neural Comput.*, vol. 4, no. 4, pp. 590–604, 1992.
- [56] Q. Gu, T. Zhang, J. Han, and C. Ding, "Selective labeling via error bound minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 323–331.
- [57] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [58] 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.