Change Detection for Hyperspectral Images Via Convolutional Sparse Analysis and Temporal Spectral Unmixing

Qingle Guo, *Student Member, IEEE*, Junping Zhang, *Senior Member, IEEE*, Chongxiao Zhong, *Student Member, IEEE*, and Ye Zhang, *Member, IEEE*

***Abstract*—With the increase in the availability of multitemporal hyperspectral images (HSIs), HSIs change detection (CD) methods, including pixel-level and subpixel-level based methods, have attracted great attention in recent years. However, the widespread presence of mixed pixels in HSIs may make it difficult for pixel-level methods to detect subtle changes; meanwhile, the less utilization of spatial information may also lead to limitations in some subpixel-level methods. Therefore, a joint framework, which aims to combine the advantages of pixel-level in spatial utilization and subpixel-level in temporal and spectral exploration, is proposed to enhance the performance of HSIs CD. Two models, convolutional sparse analysis and temporal spectral unmixing, are introduced and presented to characterize different spatial structures and overcome the effects of spectral variability under this framework, respectively. In addition, a multiple CD-based on subpixel analysis is discussed as well. Experiments conducted on three bitemporal HSIs datasets indicate that the proposed framework is robust in capturing effective features and has achieved great detection accuracy.**

***Index Terms*—Convolutional sparse analysis, multitemporal hyperspectral images (HSIs) change detection (CD), pixel-level and subpixel-level combination, temporal spectral unmixing.**

I. Introduction

COMPARED to multispectral images, hyperspectral images (HSIs) contain a higher number of spectral channels, with the ability to detect both subtle and multiple changes, which can be applied to many fields [1]–[3]. However, the high data dimension and the widespread presence of mixed pixels make it difficult to obtain real changes in HSIs. Therefore, the search of more effective methods to address the above problems still remains a key concern.

Currently, the methods of HSIs change detection (CD) can be divided into two main categories: pixel-level and subpixel-level methods. Pixel-level methods are extended from multispectral images to obtain the changes by directly comparing the numerical relationships between each pixel; subpixel-level approaches consider the internal components of each pixel, starting from the characteristics of the HSIs, to detect more detailed changes [4]. Based on the above two ideas, various CD methods have been proposed. Multivariate alteration detection (MAD) adopts canonical correlation analysis to obtain detection result [5]. In addition, iterative reweighted MAD extends the applicability of MAD and improves detection performance by adding weights to unchanged regions [6]. Change vector analysis (CVA) is first applied to multispectral images as a traditional CD method [7], and is later extended to HSIs [8]. It is mainly used to obtain the magnitude and direction of changes by calculating the distance between multitemporal spectral vectors. However, the existence of spectral variability may affect the detection result by using CVA alone. Therefore, many researchers have provided improvement strategies for CVA. In [9], an enhanced principal component analysis CVA (PCA-CVA) is proposed to reconstruct HSIs to obtain more accurate changes by combining PCA transformation and inverse triangular function. Polar CVA (PCVA) projects the images to the polarization domain and acquires the changes using CVA [10]. An enhanced version, in [11], presents a compressed CVA (CCVA) method to alleviate the problem of PCVA when the number of spectral bands is too high. Additionally, a sequential spectral CVA is proposed to continuously analyze the change vector in an iterative manner to obtain fine changes [12]. To obtain a more accurate number of change classes, a hierarchical clustering-based CVA is also proposed [13]. However, although the above methods make full use of spectral information, their spatial utilization is relatively limited and the ambiguity may be increased due to abnormal spectral variations in isolated pixels and registration errors, leading to the increase of errors. Therefore, some spatial information injected methods are proposed. In addition to spectral information, the core step of these methods is spatial information extraction. The utilization of spatial information improving the performance of the methods has been validated in many fields, such as image classification [14] and object detection [15]. Likewise, the spatial information utilization has a significant impact on CD as well. In [16], a multiscale morphological CCVA method is proposed to describe the multiscale changes by the spectral and spatial combination. A maximum-likelihood correlation coefficient is presented to fuse spectral and spatial information [17]. In addition, a stacked autoencoders and segmentation-based strategy introduces block feature into CD to obtain subtle changes [18].

Manuscript received February 3, 2021; revised March 11, 2021 and April 2, 2021; accepted April 17, 2021. Date of publication April 22, 2021; date of current version May 10, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61871150. *(Corresponding author: Junping Zhang.)*

The authors are with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: gqle_hit@163.com; zhangjp@hit.edu.cn; zhongcx_hit@163.com; zhye@hit.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3074538

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

However, the obtained spatial features of these methods usually tend to be blurry that obscure some important structural information of the input image. This severely limits the effectiveness of joint analysis strategy on HSIs CD [19].

Whether analyzing spectral information alone or jointly analyzing spatial and spectral information, the above-mentioned methods start from the assumption of pure pixels. Mixed pixels are one of the typical characteristics in HSIs, which makes CD particularly complex due to their widespread presence [20]. Spectral unmixing, as a method to solve mixed pixels problem, determines the reference spectral features composing HSIs and the corresponding abundance fractions in each pixel, which can be used to obtain the subtle changes within a pixel. As a result, many subpixel-level methods based on spectral unmixing have been proposed, such as block-based unmixing [21], [22], multispectral joint unmixing [23], superpixels unmixing [24], sparse unmixing (SU) [25]–[27], and comprehensive unmixing [28]. The blocked-based unmixing method first stacks the images and divides images into blocks; then, the endmembers of each block can be extracted. Finally, the unmixing process is carried out based on this endmember set. The SU approach begins with the ideas of sparse analysis and matrix factorization to detect subpixel-level information through spectral libraries. In these methods, the spectral library is necessary, thus limiting the ranges of its application. Although subpixel-level methods can detect subtle changes, these methods are limited in spatial utilization and inevitably lead to fragmented detection results.

To cope with some issues of the aforementioned techniques, an effective CD framework that jointly analyzes temporal, spatial, and spectral features at the pixel-level and subpixel-level is designed. Two models, convolutional sparse analysis and temporal spectral unmixing, are proposed as the key techniques. The former can achieve the effective spatial utilization at the pixel-level; the later enables the combination of spectral and temporal information. The main contributions of this article can be summarized as follows.

1) A combination framework based on pixel-level and subpixel-level analysis is designed to address the shortcomings of existing CD methods in information exploitation.
2) A spectral unmixing method that considers temporal information, namely temporal spectral unmixing, is proposed to CD to alleviate the influence of spectral variability while obtaining robust subpixel information. During the processing, the perturbation endmembers and spatial constraint are adopted and their update rules are analyzed by iterative methods.
3) A convolution sparse analysis method is introduced to effectively extract multiscale spatial information, in which the analysis and synthesis sparse priors are cooperated in a convolution manner. The convolutional operation can exploit the intrinsic spatial characteristic of the spectral change difference (SCD) image and a global optimal solution.

The rest part of this article is organized as follows. In Section II, the framework of the proposed CD method is illustrated in detail. Section III reports experimental results using some datasets. The detail discussions are also presented in this section. Finally, conclusions are given in Section IV.

II. Proposed Framework

The proposed method aims to investigate an effective framework for integrating the advantages of detecting the abrupt changes at the pixel level and subtle changes at the subpixel level to improve change representation and discrimination. It first generates SCD by subtracting operation. Based on SCD, PCA transformation is conducted on SCD and the first three principal component maps can be obtained. Then, a convolutional sparse analysis method is introduced to obtain different spatial structures of these principal component maps. This strategy will integrate multiscale changes to improve the CD performance at pixel level. Due to the unavoidable spectral variability deriving from different acquisition times of multitemporal images, a novel temporal spectral unmixing method is proposed to obtain more robust subpixel abundance. Combining the advantages of the pixel-level and subpixel-level analyses, the binary changes can be extracted by Support Vector Machine (SVM); the multiple changes may be obtained by abundance combination. The flowchart of the proposed framework is illustrated in Fig. 1. Details of each step are introduced in the following sections.

A. Spectral Analysis at the Pixel-Level via PCA

Assuming that D_1 and D_2 are two temporal HSIs that are acquired at the same regions and have been preprocessed by radiometric correction and image registration. The SCD can be computed by

$$D = D_2 - D_1 \tag{1}$$

where D has the same dimension as those of D_1 and D_2. PCA, a widely used HSIs processing method, can be used to reconstruct the spectral features of D at the pixel level. The reason for reconstructing D is because PCA can achieve concentration of energy, reduce the dimensionality of the data, and decrease the computational effort of subsequent convolutional sparse analysis. PCA uses a projection matrix to transform D into a new image (principal component maps) in a linear way. The dimensionality of the new image remains the same as that of D; where the projection matrix consists of a feature vector of the covariance matrix of D. The variance between the bands in the principal component maps will gradually decrease. For specific implementation principles of PCA, the readers can refer to [29] and [30].

B. Spatial Structure Analysis at the Pixel-Level by Convolutional Sparse Analysis

The dimensionality of the principal component maps remains consistent with that of D, but the variance between the bands gradually decreases. It means that the first three bands of principal component maps may contain most of the image information. Therefore, in this section, we use the first three principal component maps as inputs for spatial feature analysis.

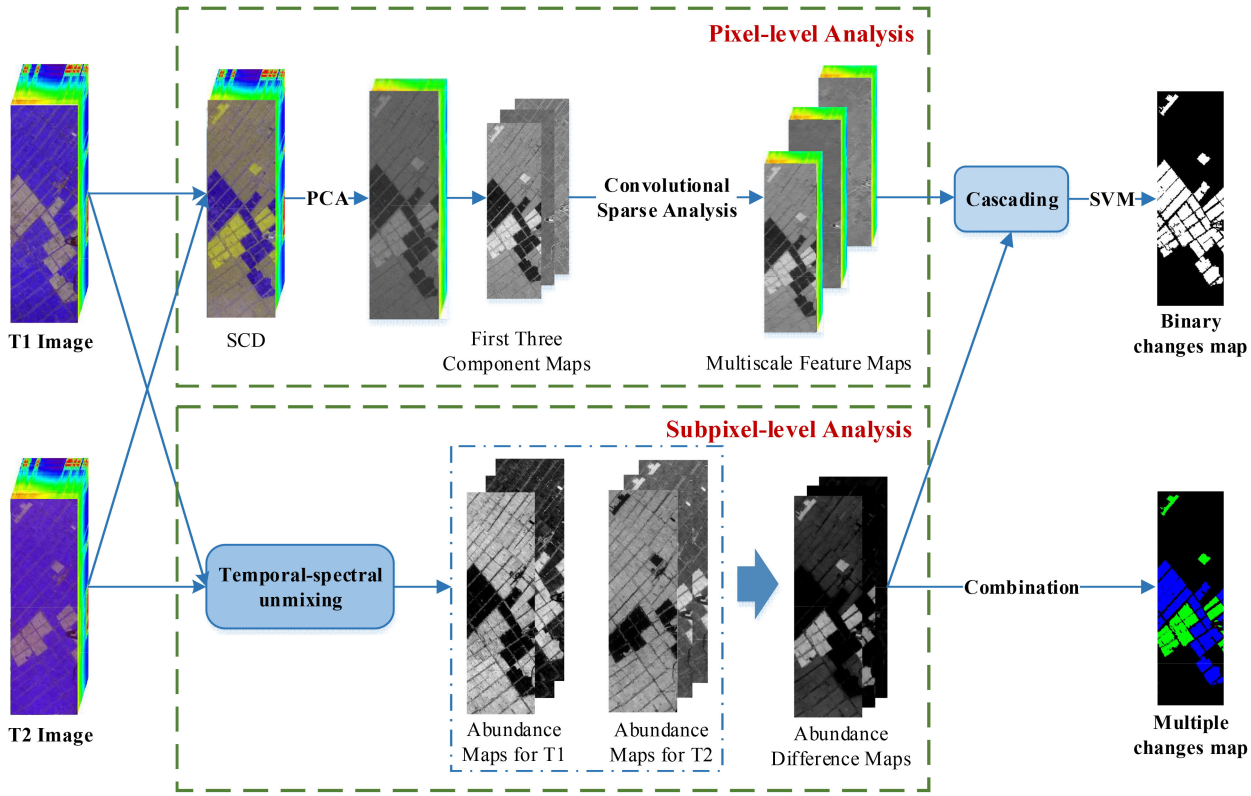


Fig. 1. Flowchart of the proposed method.

Assuming that an input image can be separated into two main components: a spatial structure component that is directly related to the spatial semantic of grounds and a corresponding texture component. It means that each principal component map P can be represented as the summation of these components [19], [31]

$$P = S + T \tag{2}$$

where S is the spatial structure component and T denotes the corresponding texture component. Therefore, the problem of spatial feature analysis is transformed into a problem of solving the variable S. Since the number of unknown variables in (2) is larger than the input, the solution processing is ill-posed. To solve this problem and successfully obtain the spatial structure, the analysis sparse prior [31] and synthesis sparse prior [32] can be jointly adopted as priori terms. While the former one is superior in extracting the spatial structure information, the later can effectively characterize the high-frequency texture components. In addition, considering the traditional approaches that partitioning the image into several independent block regions can only lead to local optimum solution, the convolution operation is introduced to further enhance the performance of sparse representation. Therefore, the convolution operation can be introduced to further enhance the image representation. The final spatial structure components can be obtained by solving the following objective function:

$$f = \left\| P - S - \sum_{m=1}^{M} d_{A,m} \otimes A_m \right\|_F^2 + \lambda \sum_{q=1}^{Q} \|d_{B,q} \otimes S\|_1 + \beta \sum_{m=1}^{M} \|A_m\|_1 \tag{3}$$

where λ and βrepresent the regularization parameters of two sparse priors, respectively, $d_{A,m}$ is the mth atom of the convolutional synthesis dictionary, A_m denotes the coefficient matrix, $d_{B,q}$ represents the qth atom of the analysis dictionary, and $\otimes$indicates a convolution operation. It can be seen that the operation of multiplying the coefficients with the dictionary is replaced by the convolution operation, and two sparse priors are adopted to constrain the structure and texture components, respectively. To achieve the minimization of the objective function to solve for S, the alternating direction method of multipliers (ADMM) can be used [19]. During the optimization process, the regularization parameters as key parameters have an impact on the results. As λ decreases, the finer the obtained structure component is. Thus, the adjustment of λ will obtain S in different scales. These different S can accurately depict the multiscale changes at the pixel level. In the specific experiments, we recommend the adjustment range of λ to (0, 0.5].

C. Abundance Analysis at the Subpixel-Level Via Temporal Spectral Unmixing

Due to the presence of the mixed pixels, the subtle changes can be usually ignored when only the pixel-level features are considered. Thus, spectral unmixing can be used to CD for subpixel analysis. However, some traditional separate unmixing models for multitemporal images treat a pixel as a summation of the multiplication of endmembers and abundances and do not consider the impact of spectral variability in the whole process, resulting in a poor performance. The spectral unmixing with temporal information, called temporal spectral unmixing, may provide a solution to the above problem. Temporal spectral unmixing consists of two inputs: the current temporal image and the endmembers of the previous temporal image. The endmembers of the current temporal image are obtained by optimizing and updating the previous temporal endmembers with the current temporal image information. The initial endmembers can be obtained by vertex component analysis. With such strategy, intertemporal correlation can be achieved and the abundance maps of each temporal images can be obtained. It is worth noting that only one temporal image and the endmembers of the previous temporal image need to be loaded into memory in the unmixing process, which greatly reduces the memory consumption. Moreover, a spectral perturbation term and spatial constraint term can be added into the unmixing process to overcome the influence of spectral variability.

Assuming that the number of endmembers is *K*, the number of images contained in the multitemporal image sequence is *T*, and the number of pixels of each temporal image is *N*. The unmixing model can be represented as

$$x_{it} = \sum_{k=1}^{K} a_{kit}\left(e_{kt} + de_{kt}\right) + b_{it} + \sum_{j\in\mathbb{N}(i)} \varpi_{jit} x_{it} \tag{4}$$

where x_{it} represents the *i*th pixel in the *t*th image and $i = 1, \ldots, N$, $t = 1, \ldots, T$, $k = 1, \ldots, K$. a_{kit} is the abundance value of the *k*th endmember in the *i*th pixel of the *t*th image. e_{kt} denotes the *k*th endmember of the *t*th image. de_{kt} is the spectral perturbation caused by spectral variability of the *t*th temporal image *k*th endmember. b_{it} represents the noise of the *i*th pixel in the *t*th image, due to data acquisition or model bias. $\mathbb{N}(i)$ denotes the neighborhood set of four pixels centered on pixel *i*. ϖ_{jit} is the *j*th weight in the four-pixel neighborhood of the *t*th temporal image. Specially, a positive value of ϖ_{jit} indicates that *i* and *j* are neighbors. The matrix form is

$$X_t = \left(E_t + dE_t\right) A_t + B_t + W_t X_t \tag{5}$$

where $X_t = [x_{1t}, \ldots, x_{Nt}]$ is a matrix that contains all the pixels of the *t*th temporal image. A_t is the abundance matrix. E_t is endmembers matrix where each column represents an endmember. dE_t represents the spectral perturbation matrix where each column represents a perturbation vector. B_t denotes the noise matrix. $W_t$represents the spatial weight matrix. In the temporal spectral unmixing, both abundance and endmembers have their own numerical limits that are mentioned in many unmixing methods [21]. Therefore, a similar idea is proposed to limit the perturbations. The constraints are formed as

$$\|dE_t\|_F^2 \le \sigma^2, \text{where } t = 1, \ldots, T \tag{6}$$

$$\left\| \sum_{t=1}^{T} dE_t \right\|_F^2 \le \kappa^2 \tag{7}$$

where σ and κ are constant parameters that control the energy value and instantaneous energy variation of the spectral variability, respectively. The values of these two parameters are closely related to the spectral variability between multitemporal images. When the spectral variability is severe, the two parameters should be increased properly. With the above-mentioned modeling form, the unmixing problem is transformed into a problem of solving endmembers space *M*, abundances space $\wp_A$, and perturbation space P_t, which can be regarded as an empirically minimized two-stage stochastic optimization problem. The objective functions to be optimized is expressed as

$$\min_{E\in M} \frac{1}{T} \sum_{t=1}^{T} H\left(X_t, E_t\right) \tag{8}$$

$$H\left(X_t, E_t\right) = \min_{(A,dE)\in \wp_A \times p_t} F\left(X_t, E_t, A_t, dE_t\right) \tag{9}$$

$$F(X_t, E_t, A_t, dE_t) = \frac{1}{2}\left\|X_t - \left(E_t + dE_t\right) A_t - W_t X_t\right\|_F^2 + \alpha\phi\left(A_t\right) + \gamma\varphi\left(E_t\right) + \mu\theta\left(dE_t\right) \tag{10}$$

where $F(\cdot)$ is the normalized difference measure function. $H(\cdot)$ is the loss function. α, γ , and μ are parameters that ensure the tradeoff between data fitting term and penalties. Different from the traditional unmixing methods, the perturbation endmembers are used as a penalty term to constrain the final solution. This processing way ensures that robust endmembers for spectral variability can be obtained. Additionally, the introduction of spatial weights will improve the performance of unmixing. Minimizing the function *F* and the loss *H*, the optimal solutions of the current temporal image can be obtained. Furthermore, the penalty terms can be represented as follows:

$$\phi\left(A_t\right) = \frac{1}{2}\left\|A_t - A_{t-1}\right\|_F^2 \tag{11}$$

$$\varphi(E_t) = \frac{1}{2} \sum_{i=1}^{K} \left(\sum_{\substack{j=1 \\ j\neq i}}^{K} \|e_i - e_j\|_F^2 \right) \tag{12}$$

$$\theta(dE_t) = \frac{1}{2}\left\|dE_t - dE_{t-1}\right\|_F^2 \tag{13}$$

where $\phi(\cdot)$ is to promote temporally smooth abundances in the l_2-norm sense between two temporal images. $\varphi(\cdot)$ represents the mutual distance between each endmember. $\theta(\cdot)$ denotes a prior spectral smooth between two temporal images. In this optimization process, proximal alternating linearized minimization and ADMM method can be adopted to obtain the final A_t, E_t, and dE_t [33]–[35].

Through the above-mentioned temporal spectral unmixing, the abundance maps for each temporal image can be obtained. Meanwhile, the abundance difference maps can be calculated

by subtracting the abundance maps corresponding to the same endmembers of different temporal image. Furthermore, a new feature cube can be cascaded by the pixel-level and subpixel-level features in the third dimension.

D. Binary Changes Extraction and Multiple Changes Discussion Via Abundance Combination

This step is to visualize and discriminate binary changes and multiple changes present in the feature cube. To achieve the purpose, clustering and supervised classifiers can be employed. In this article, the most widely used SVM classifier [36] is used to obtain the binary change map.

However, SVM cannot obtain the multiple changes in the absence of the multiple change truth map. Hence, obtaining relatively accurate multiple changes should be discussed in this case. One solution is to use unsupervised clustering methods; however, the individual clustering is difficult to obtain good result due to the issue of feature distance between change classes; another solution is abundance combination. In fact, in the temporal spectral unmixing process, the endmembers satisfy linear independence and the spectrum of different endmember corresponds to that of different change classes [21]. This means that the number of endmembers is the same as the number of change classes. Additionally, according to the correspondence between abundance and endmembers in the linear unmixing process, the number of abundance features is equal and the same as the number of change classes as well. Therefore, multiple changes can be obtained by combining abundance difference maps. For changes, assume that each change pixel has the same label on the same abundance difference map, and has different labels on different maps. We can determine the final label of each change pixel by comparing the pixel value at the same coordinates on all abundance difference maps. The determination criterion is that the label of each change pixel is the same as the label of the abundance difference map where the largest pixel value is located, which can be expressed as follows:

$$M_C(i,j) = \arg\max_{c_K \in C} \left(DA_{c_K}(i,j)\right) \tag{14}$$

where $k = 1, \ldots, K$ is the index and $C = (c_1, \ldots, c_K)$ represents the set of change types, c_K is the kth class. i and j are indexes of the abundance difference maps DA_{c_K} with different labels.

III. Experiments and Analysis

A. Experimental Dataset Descriptions

To evaluate the performance of the proposed method, several experiments are conducted on three real bitemporal datasets. Details of these three datasets are described as follows.

The first farmland dataset was taken by Earth Observing-1 (EO-1) Hyperion from Jiangsu, China, with a size of 450 × 140 pixels and 155 bands, which were obtained on May 3, 2006, and April 23, 2007. The images are characterized by a spectral resolution of 10 nm and a spatial resolution of 30 m. The main changes relate to crop types and areas and the number of change pixels is 18 277. Fig. 2(a)–(c) provides the color image that is composed by band 32, band 25, and band 12. The corresponding reference binary change map is labeled by visual interpretation.

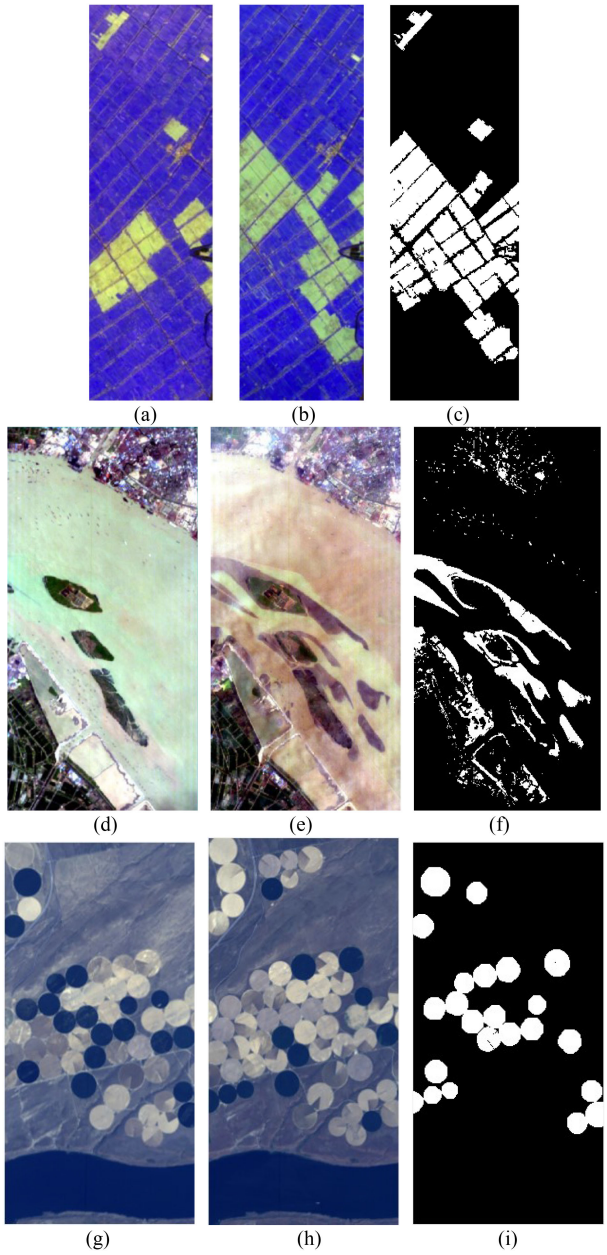


Fig. 2. Coregistered real HSIs datasets. (a) T1 color image of the farmland dataset acquired on May 3, 2006. (b) T2 color image of the farmland dataset acquired on April 23, 2007. (c) Ground truth of farmland dataset. (d) River dataset T1 acquired on May 1, 2004. (e) River dataset T2 acquired on December 3, 2013. (f) Ground truth of river dataset. (g) Agricultural dataset T1 acquired on April 10, 2014. (h) Agricultural dataset T2 acquired on May 8, 2007. (i) Ground truth of agricultural dataset.

The second bitemporal river dataset was acquired on May 3, 2013, and December 31, 2013 by EO-1 Hyperion from Jiangsu, China, with a size of 463 × 241 pixels and 198 bands. The number of change pixels is 12 570 and the main changes appear in a river. It contains many complicated situations, such as illumination and seasonal differences [28]. Fig. 2(d)–(f) displays

the color image and binary ground truth map annotated by visual interpretation of river dataset.

The third bitemporal dataset was obtained by Hyperion sensor mounted onboard the EO-1 satellite on May 1, 2004, and May 8, 2007 as well. The study area is an agricultural land of Hermiston city in Umatilla County, United States. The selected area, which has a size of 390×200 pixels and 242 bands, has the wavelength range from 350 to 2580 nm. Fig. 2(g)–(i) shows the color composite image and the binary ground truth annotated by visual interpretation. The number of change pixels is 9986 and the changes occurred in this dataset include land-cover transitions between crops, soil, water, and vegetation.

B. Comparative Methods and Evaluation Rules

Several representative and feature combination methods related to the proposed framework are selected to validate the performance of the proposed framework in the experiments, which include improved SU [25], GETNET [28], SCD_SVM (SCDS), Spectral_Spatial_SVM (SPS), Spectral_Abundance_SVM (SAS) and Spectral_Traditional sparse_Abundance_SVM (STAS). SU is a subpixel-level method to detect the subtle changes based on spectral libraries [25], which uses variable splitting augmented Lagrangian and total variation (SUnSAL-TV) method to acquire endmembers and abundances. In our experiments, the abundance features obtained by SU will replace the abundance features obtained by temporal spectral unmixing in the proposed framework. This improved way can be used to verify the validity of the adopted unmixing model; GETNET is an outstanding CD method based on an end-to-end deep learning network, which combines deep learning with spectral unmixing by designing a new form of data input to obtain binary changes. SCDS is a traditional pixel-level method that uses SVM to directly detect changes in SCD images. SPS is a spatial-spectral ensemble method that combines the first three principal components with the spatial features obtained by convolutional sparse analysis to form a feature cube, and then uses SVM to detect the changes. SAS is a pixel-level and subpixel-level joint method that directly exploits spectral information by combining the first three principal component maps and the abundance features after temporal spectral unmixing into a new feature cube, which then uses SVM to detect the changes. STAS is the method that adopts the traditional sparse representation to obtain the spatial features. Among these feature combination methods, SCDS is used to validate the improved performance of spatial feature and abundance feature analysis for CD. SPS is used to further validate the effectiveness of convolutional sparse analysis based on the advantages of spatial and spectral ensemble. SAS is used to confirm the effectiveness of temporal spectral unmixing. STAS is to prove the effectiveness of the adopted convolutional operation in spatial feature extraction.

To evaluate the performance of the competing methods comprehensively, commonly used precision, recall, and *F*-score are applied to record and evaluate the performance of different CD methods. Higher precision and recall values indicate that small number of false alarms and missed detections, respectively. Meanwhile, *F*-score is evenly balanced between false alarms and missed detections, revealing the overall detection performance [37], [38]. All experiments are implemented on a desktop computer with 2.90-GHz CPU and 32 GB of RAM.

TABLE I
QUANTITATIVE COMPARISON OF FARMLAND DATASET CD RESULTS BY DIFFERENT METHODS

Method	Precision	Recall	F-score
Improved SU	95.75%	94.35%	95.04%
SCDS	90.46%	93.83%	92.11%
SPS	90.09%	96.36%	93.12%
SAS	96.19%	97.10%	96.75%
STAS	96.89%	96.52%	96.85%
GETNET	90.02%	94.52%	92.21%
The proposed	**96.92%**	**97.30%**	**97.10%**

Bold is the maximum value of the current column.

C. Results and Discussions of Farmland Datasets

The first experiment is conducted on a farmland dataset. During the experiment, the optimal parameters of SU and GETNET are selected based on the references. Moreover, the related parameters of improved SU, SPS, SAS, STAS, and the proposed framework are set as follows: $\lambda = \{0.01, 0.02, 0.03\}$, $K = 3$, $\beta = 0.02$, $\sigma^2 = 1$, $\kappa^2 = 0.1$, $\alpha = 10^{-4}$, $\gamma = 10^{-3}$, and $\mu = 10^{-5}$. According to our iterative experiments, the key core parameters on the adopted dataset are *K* and λ. Therefore, we give default values for the other parameters, which remain unchanged in the subsequent experiments. By setting the above-mentioned parameters, we can obtain the number of channels of the new feature cube as 15, where the number of spectral features, abundance features, and spatial features are 3, 3, and 9, respectively. For the process of SVM, 0.1% samples are randomly chosen from the reference map to obtain the final changes. The obtained detection results of the proposed framework and the comparative methods are tabulated in Table I, and the corresponding change maps are shown in Fig. 3.

Some interesting findings are presented in Fig. 3. Comparing Fig. 3(a) and (g), it can be seen that although the improved SU and the proposed method adopt the same spatial feature constraints, however, the false alarm in (a) is more than that in (g). This point can be obtained in the quantitative accuracy as well (precision is equal to 95.75% and 96.92%, respectively.) The reason may be that the SU does not consider the effects of spectral variability in multitemporal images. In contrast, the introduced temporal spectral unmixing is more suitable for multitemporal image analysis. In addition, SCDS that directly detects changes in SCD cannot obtain satisfactory results because it does not take account of the constraints on spatial information. It can be verified by the *F*-score of SCDS and SPS. SAS, which introduces abundance features obtained by temporal spectral unmixing, can obtain a significant improvement (the *F*-score is reached to 96.75%). The possible reason is that the spectral utilization of SAS is higher than that of SPS and SCDS. This result indicates that subpixel analysis has good applicability for HSIs CD. However, SAS detects some false

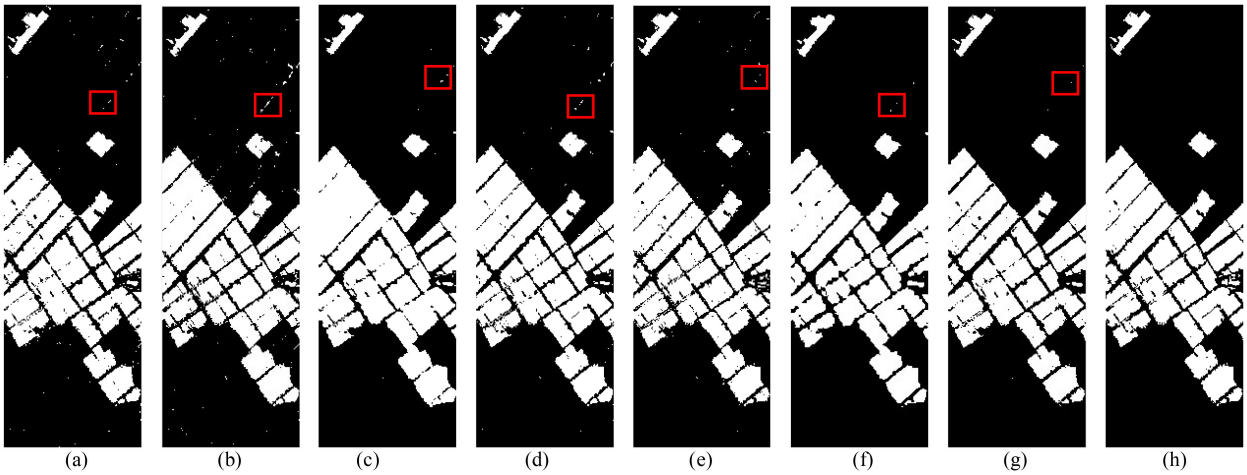


Fig. 3. Detection results of different methods on farmland dataset. (a) Improved SU. (b) SCDS. (c) SPS. (d) SAS. (e) STAS. (f) GETNET. (g) The proposed method. (h) Binary ground truth map.

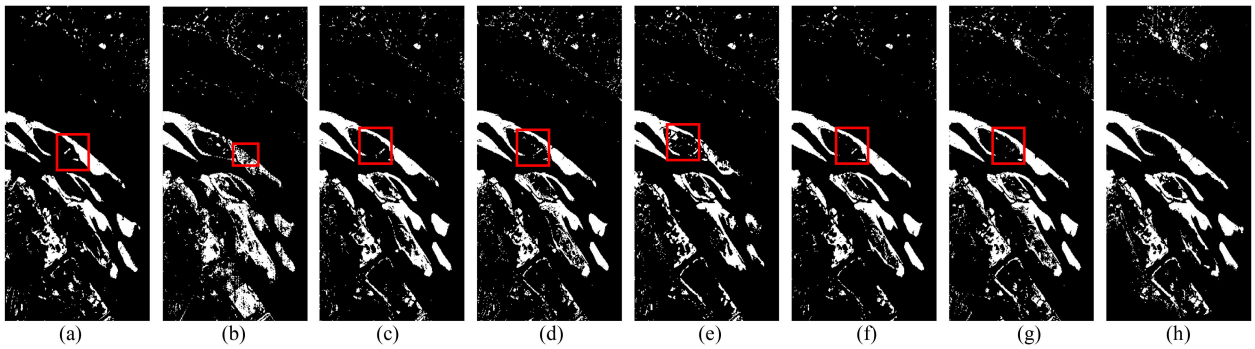


Fig. 4. Detection results of different methods on river dataset. (a) Improved SU. (b) SCDS. (c) SPS (d) SAS. (e) STAS. (f) GETNET. (g) Proposed method. (h) Binary ground truth map.

alarms as well [as shown in red box of Fig. 3(d)], which are mainly due to the excessive focus on spectral features. Therefore, further improvements in detection performance can be achieved through appropriate spatial information injection [see Fig. 3(e)]. GETNET uses abundance features to construct a new input form, which essentially takes advantage of the spectral information of HSIs and also yields relatively good detection results [see Fig. 3(f)]. Overall analysis, the F-score of the proposed method is highest (reached to 97.10%), which confirms the validity of the adopted spatial feature analysis method and the reliability and superiority of the proposed combination framework. Comparing Fig. 3(e) and (g) also demonstrates that the convolutional sparse form used is more efficient than the traditional one.

D. Results and Discussions of River Datasets

The second experiment is conducted on a river dataset. Different temporal images often exhibit spectral variability due to differences in acquisition time, season, and other factors, which can seriously affect the applicability of many algorithms, and makes it difficult to obtain satisfactory detection results. Therefore, this dataset can well validate the advantages of our proposed method in overcoming the spectral variability. During the experiment, the optimal parameters of SU and GETNET are selected. Meanwhile, the parameters of the related and the proposed method are set as follows: $K = 4$ and $\lambda = \{0.007, 0.008, 0.009, 0.01\}$, the rest of the parameters remain the same as the farmland dataset. The dimension of the new feature cube is 19, where the number of spectral, abundance, and spatial features are 3, 4, and 12, respectively. For changes extraction, 0.1% of samples are randomly chosen. The corresponding results of all mentioned methods are given in Fig. 4, meanwhile, the quantitative results are tabulated in Table II.

TABLE II
QUANTITATIVE COMPARISON OF RIVER DATASET CD RESULTS BY DIFFERENT METHODS

Method	Precision	Recall	F-score
Improved SU	81.34%	76.50%	78.85%
SCDS	60.36%	65.37%	62.77%
SPS	85.01%	81.30%	83.11%
SAS	85.84%	81.55%	83.64%
STAS	**87.81%**	78.16%	82.58%
GETNET	80.28%	62.66%	70.39%
The proposed	86.71%	**83.26%**	**84.95%**

Bold is the maximum value of the current column.

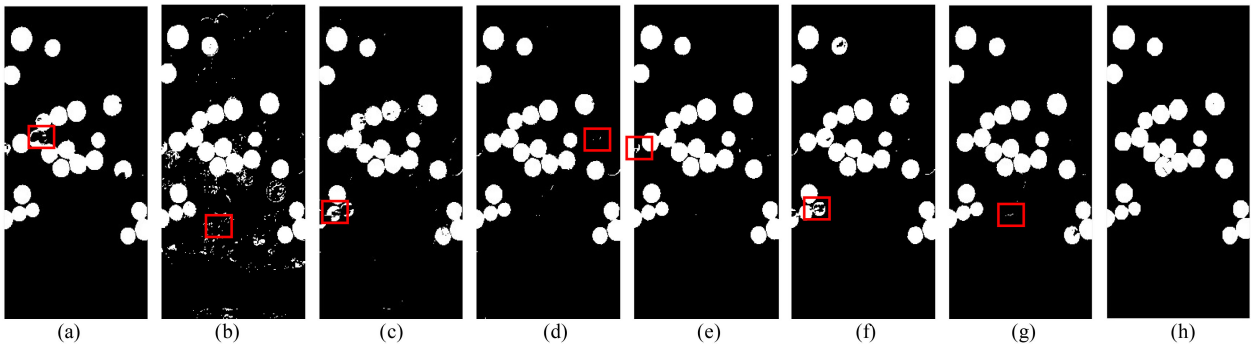


Fig. 5. Detection results of different methods on agricultural dataset. (a) Improved SU. (b) SCDS. (c) SPS. (d) SAS. (e) STAS. (f) GETNET. (g) Proposed method. (h) Binary ground truth map.

Some conclusions similar to those of the farmland dataset can be obtained. It can be observed that the results obtained by the proposed framework are better than the others. Specifically, the proposed framework gets the highest accuracy in all three indexes (reached to 86.71%, 83.26%, and 84.95, respectively). Compared to the SU and the proposed approach, the advantage of the temporal spectral unmixing strategy is even more significant, with a 6% increase in *F*-score. Additionally, the fact that spatial information injection can improve the detection accuracy is confirmed again in this dataset, comparing SCDS and SPS. The SPS is approximately 20 percentage points higher than the SCDS in terms of *F*-score, which is sufficient to justify the effectiveness of the adopted convolutional sparse analysis. Similarly, the result of SAS also confirms the advantages of subpixel features. It is worth noting the GETNET, which uses excellent ways to combine subpixel and pixel-level information. But it still does not yield good results, with an *F*-score only reaching to 70.39%. There are two possible reasons for this phenomenon. On the one hand, it uses the pseudotraining sets to train a deep network, and these samples may not be effective in training deep network without parameter adjustment; on the other hand, the input of GETNET is the feature combination of bitemporal stacking unmixing, not the original HSIs. Such stacking unmixing shows poor robustness in the presence of spectral variability, and feeding the convolutional neural network with such a nonrobust result may introduce unavoidable errors, resulting in a relatively low accuracy. On the contrary, the proposed temporal spectral unmixing obtains more robust abundance features and thus leading to a better detection result.

E. Results and Discussions of Agricultural Datasets

To illustrate the suitability of our proposed method further, we introduce an agricultural dataset. The parameters of SU and GETNET are optimal as well. For the proposed framework, the relevant parameters are set to: $\lambda = \{0.007, 0.008, 0.009, 0.01\}$, $K = 5$. Therefore, the dimension of the feature cube is 20, where the number of spectral, abundance, and spatial features are 3, 5, and 12, respectively. In SVM, 0.1% of samples are randomly selected as well. The corresponding qualitative detection results and the quantitative indicators of all methods are displayed in Fig. 5 and Table III, respectively.

TABLE III
QUANTITATIVE COMPARISON OF AGRICULTURAL DATASET CD RESULTS BY DIFFERENT METHODS

Method	Precision	Recall	F-score
Improved SU	85.28%	86.99%	86.13%
SCDS	84.37%	**95.18%**	89.45%
SPS	94.87%	89.68%	92.20%
SAS	96.27%	93.47%	94.85%
STAS	95.32%	94.53%	94.93%
GETNET	96.72%	90.67%	93.60%
The proposed	**96.75%**	94.72%	**95.72%**

Bold is the maximum value of the current column.

It can be observed that the proposed framework is an effective idea for CD, after considering precision, recall, and *F*-score, comprehensively. The accuracy is reached to 96.75%, 94.72%, and 95.72%. Temporal spectral unmixing, by adding perturbation endmembers to the unmixing process, is more suitable for subpixel feature analysis, as the detection accuracy of the proposed method is higher than the results of SU (only 86.13%). In addition, the adopted convolutional sparse analysis has shown to be an effective strategy as well, since the accuracy of SPS is higher than that of SCDS that uses the spectral information directly. The SAS method also achieved an *F*-score of 94.85%, which demonstrates that the subpixel analysis method is capable of detecting strong changes as well as obtaining subtle changes. GETNET comprehensively analyzes pixel and subpixel features has obtained a great result as well. The framework presented yields the best results, demonstrating that the new spatial and abundance feature analysis strategies are effective. Compared to GETNET, our proposed approach achieves more effective result with a low framework complexity and a simple and easy-to-use strategy.

F. Discussions of Multiple Changes

The higher spectral resolution of HSIs makes it possible to make full use of spectral information to detect multiple changes. However, when the ground truth map only labels for two classes, i.e., the changed and unchanged classes, we cannot obtain how

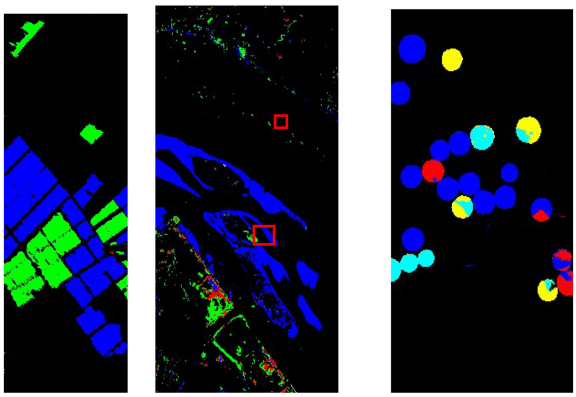


Fig. 6. Multiple change results of three HSI datasets obtained by abundance feature combination. (a) Three types of changes of the farmland dataset. (b) Four types of changes of the river dataset. (c) Five types of changes of agricultural dataset.

many change types are included in the change class through SVM. Therefore, in order to obtain relatively accurate information on multiple changes in this context, the use of subpixel abundance features in an unsupervised manner can be considered. In fact, in the spectral unmixing, the endmembers satisfy linear independence and the spectrum of different endmembers corresponds to that of different change classes. This means that the number of endmembers is the same as the number of change classes. Hence, according to the principle of linear expression, the number of abundance features is the same as the number of change classes. For example, the number of endmembers in the farmland dataset is 3, so its number of abundance features is 3, which corresponds to the number of change classes. Therefore, multiple changes can be detected by combining these abundance features. The results of the multiple CD of three experimental datasets are presented in Fig. 6, where the farm dataset contains three types of changes, the river dataset includes four types of changes, and the agricultural dataset has five types of changes.

Three, four, and five types of changes are detected for three datasets, respectively, where different classes are marked by different colors. The results show that the introduced unmixing method yields effective abundance features, and the adopted unsupervised combination way obtains relatively accurate multiple changes. It is worth noting that there is a certain amount of misinterpretation on the multiple change results (marked by red boxes), due to the fact that only subpixel features are considered in the combination process and pixel-level features are ignored. However, this is acceptable. On the one hand, subpixel features can yield a relatively satisfactory result, and on the other hand, the datasets are lack of pixel-level class labels that naturally prevent the introduction of such pixel-level information.

IV. Conclusion

In this article, based on the spectral, spatial, and temporal analysis strategy, we proposed a novel comprehensive analysis framework to obtain binary and multiple changes for HSIs. Overall, a convolutional sparse analysis, which cooperates two sparse priors and adopts convolution operation to replace the traditional method, enables the effective integration of multiscale spatial structure while significantly improving the information utilization. Furthermore, due to the introduction of spatial constraint and perturbation endmembers, the proposed temporal spectral unmixing overcomes the spectral variabilities between bitemporal images, while achieving the accurate analysis. The whole framework presents a new solution for CD in a simple but effective way. Three datasets are adopted to verify the proposed framework and the F-score is reached to 97.10%, 84.95%, and 95.72%, respectively. Meanwhile, the proposed strategy is also clearly superior to some typical and feature combination methods.

References

[1] P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Digital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, 2004.

[2] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, 2004.

[3] S. Liu, D. Marinelli, L. Bruzzon, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images," *IEEE Geosci. Remote Sens. Mag.,* vol. 7, no. 2, pp. 140–158, Jun. 2019.

[4] Y. Ban and O. Yousif, "Change detection techniques: A review" in *Multitemporal Remote Sensing: Methods and Applications*, Y. Ban, Ed. New York, NY, USA: Springer-Verlag, 2016, pp. 19–43.

[5] A. Nielsen and A. Muller, "Change detection by the MAD method in hyperspectral images data," in *Proc. 3th Workshop Imag. Spectrosc.,* 2003, pp. 13–16.

[6] 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. 463–378, Feb. 2007.

[7] H. Zhuang, K. Deng, H. Fan, and M. Y., "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.

[8] A. Ertu¨rk, "Fuzzy fusion of change vector analysis and spectral angle mapper for hyperspectral change detection," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2018, pp. 5045–5048.

[9] M. Baisantry, D. Negi, and O. Manocha, "Change vector analysis using enhanced PCA and inverse triangular function-based thresholding," *Def. Sci. J.*, vol. 62, no. 4, pp. 236–242, 2012.

[10] F. Bolovo and L. Bruzzon, "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.

[11] 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.

[12] 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.

[13] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Hierarchical change detection in multitemporal hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 244–260, Jan. 2015.

[14] S. Li, Q. Hao, X. Kang, and J. A. Benediktsson, "Gaussian pyramid based multiscale feature fusion for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3312–3324, Sep. 2018.

[15] J. Kaufman, M. Eismann, and M. Celenk, "Assessment of spatial -spectral feature-level fusion for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2534–2544, Jun. 2015.

[16] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzon, and F. Bolovo, "Multiscale morphological compressed change vector analysis for unsupervised multiple change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4124–4137, Sep. 2017.

[17] K. Vongsy and M. Mendenhall, "Integrating spatial & spectral information for change detection in hyperspectral imagery," in *Proc. 8th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2016, pp. 1–5.

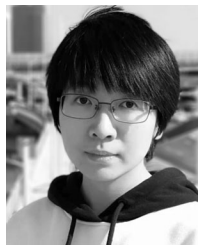
[18] J. Fandino, A. Garea, D. Heras, and F. Arguello, "Stacked autoencoders for multiclass change detection in hyperspectral images," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2018, pp. 1906–1909.
[19] C. Zhong, J. Zhang, and Y. Zhang, "Multiscale feature extraction based on convolutional sparse decomposition for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4960–4972, Aug. 2020.
[20] J. Bioucas-Dias, A. Plaza, G. Camp-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag,* vol. 1, no. 2, pp. 6–36, Jul. 2013.
[21] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Unsupervised multitemporal spectral unmixing for detection multiple changes in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2733–2748, May 2016.
[22] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Multitemporal spectral unmixing for change detection in hyperspectral images," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2015, pp. 4165–4168.
[23] G. Liu, X. Li, and Y. Dong, "Style transformation-based spatial-spectral feature learning for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3026099.
[24] S. Zhang *et al.* "Superpixel-guided sparse unmixing for remotely sensed hyperspectral imagery," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2019, pp. 2155–2158.
[25] A. Ertu¨rk, M. Iordache, and A. Plaza, "Sparse unmixing-based change detection for hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 708–719, Feb. 2016.
[26] A. Ertu¨rk, M. Iordache, and A. Plaza, "Sparse unmixing with dictionary pruning for hyperspectral change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 708–719, Jan. 2017.
[27] A. Ertu¨rk, M. Iordache, and A. Plaza, "An information change detection by unmixing for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1252–1256, Jun. 2015.
[28] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 1–11, Jun. 2018.
[29] X. Jia and J. Richards, "Segmented principal components transformation for efficient hyperspectral image display and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, Jan. 1999.
[30] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017.
[31] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.
[32] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1823–1831.
[33] P. Thouvenin, N. Dobigeon, and J. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.
[34] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 146, no. 1–2, pp. 459–494, Jul. 2013.
[35] D. Ralph and H. Xu, "Convergence of stationary points of sample aver- age two-stage stochastic programs: A generalized equation approach," *Math. Oper. Res.*, vol. 36, no. 3, pp. 568–592, Aug. 2011.
[36] A. Mathur and G. Foody, "Multiclass and binary SVM classification: Implications for training and classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 241–245, Apr. 2008.
[37] Q. Guo, J. Zhang, and Y. Zhang, "Multitemporal images change detection based on AMMF and spectral constraint strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3444–3457, Apr. 2021.
[38] Q. Guo, J. Zhang, and Y. Zhang, "Multitemporal hyperspectral images change detection based on joint unmixing and information coguidance strategy," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3045799.

Qingle Guo (Student Member, IEEE) received the B.S. degree in remote sensing science and technology and the M.S. degree in information and communication engineering, in 2016 and 2018, respectively, from Harbin Institute of Technology (HIT), Harbin, China, where he is currently working toward the Ph.D. degree in information and communication engineering.

His current research interests include image processing, multitemporal image analysis, change detection and application, and object detection and classification.

Junping Zhang (Senior Member, IEEE) received the B.S. degree in biomedical engineering and instrument from Harbin Engineering University and Harbin Medical University, Harbin, China, in 1993, and the M.S. and Ph.D. degrees in signal and information processing from Harbin Institute of Technology (HIT), Harbin, China, in 1998 and 2002, respectively.

She is currently a Professor with the Department of Information Engineering, School of Electronics and Information Engineering, HIT. Her research interests include hyperspectral data analysis and image processing, multisource information fusion, multitemporal change detection, and machine learning and its applications.



Chongxiao Zhong (Student Member, IEEE) received the B.S. degree in communication engineering from Northeastern University, Shenyang, China, in 2014, and the M.S. degree in electronics and communication engineering, in 2017, from Harbin Institute of Technology, Harbin, China, where she is currently working toward the Ph.D. degree.

Her current research interests include image processing, hyperspectral image classification, and transfer learning and application.

Ye Zhang (Member, IEEE) received the B.S. degree in communication engineering and the M.S. and Ph.D. degrees in communication and electronic system from Harbin Institute of Technology (HIT), Harbin, China, in 1982, 1985, and 1996, respectively.

Since 1985, he has been a teacher with HIT. Between 1998 and 1999, he was a Visiting Scholar with The University of Texas at San Antonio, San Antonio, TX, USA. His research interests include remote sensing hyperspectral image analysis and processing, image video compression and transmission, as well as multisource information collaboration processing and applications. He is currently a Professor and Doctoral Supervisor in information and communication engineering. He is the Director of Institute of Image and Information Technology with the School of Electronic and Information Engineering, HIT.