

# Multilayer Cascade Screening Strategy for Semi-Supervised Change Detection in Hyperspectral Images

Lian Liu <sup>1</sup>, Danfeng Hong <sup>1</sup>, *Senior Member, IEEE*, Li Ni, and Lianru Gao <sup>1</sup>, *Senior Member, IEEE*

**Abstract**—Change detection (CD) is an important application of remote sensing, which provides information about land cover changes on the Earth’s surface. Hyperspectral image (HSI) can show more spectral information, which greatly improves the ability of remote sensing to identify change features. The challenge is how to overcome the scarcity of labeled samples and extract the change information of high-dimensional spectra in HSI. To solve the previous problem, a semi-supervised CD with multilayer cascade screening strategy (MCS<sup>4</sup>CD) that uses both the spatial information and active learning is proposed to select highly reliable unlabeled samples to increase the training sets. The MCS<sup>4</sup>CD method can effectively use unlabeled samples to improve accuracy. Additionally, a subspace CD method based on iterative slow feature analysis is designed to extract the most temporally invariant component from the high-dimensional space. Experimental results on four hyperspectral datasets show that with a small number of labeled samples, the proposed method achieves a much better performance than existing CD methods.

**Index Terms**—Active learning, change detection, hyperspectral image (HSI), iterative slow feature analysis (ISFA), semi-supervised learning.

## I. INTRODUCTION

THE surface ecosystem and human social activities are dynamically developing and evolving [1], [2]. Accurate acquisition of land surface change information is of great significance to better protect the ecological environment, manage natural resources, study social development, and understand the relationship and interaction between human activities and the natural environment [3]–[5]. Change detection (CD) is the process of determining the change of the land cover state based on multiple observations at different times. As an advanced and mature technical means, remote sensing earth observation can quickly, macroscopically, and dynamically obtain surface images, which provides important data support for solving the CD of land cover [38], [40]. Therefore, using multitemporal or bitemporal remote sensing data to obtain the CD of surface

features has become one of the most widely used research fields of remote sensing technology [6], [7]. The purpose of CD research is to find interesting change information and filter out irrelevant change information as an interference factor [8]–[11].

Over the past decades, land-use and land-cover CD tasks of optical remote sensing imagery has received increasing attention in the supervised, unsupervised, and semi-supervised algorithms. Supervised learning (SL) relies on prior knowledge to supply the learning algorithm with labeled class data, unsupervised learning (UL) classifies data samples based on the features inherent within the data, and semi-supervised learning (SSL) uses a combination of both [11]–[13]. The main idea of SSL is to use the sample distribution information carried by unlabeled samples to select certain unlabeled samples to join the labeled training set, so as to improve the classification accuracy. Sufficient evidence demonstrated CD ability to provide remarkable performances on a wide range of SL tasks when trained on extensive collections of labeled data. In contrast, the classifying ability in UL still remains limited, due to missing label information [34], [35], [39]. However, obtaining such large labeled training samples requires a considerable amount of resources, time, and effort [14]–[18]. In many practical situations, the scarcity of labeled training samples is caused by bad weather areas (such as Gobi, desert, and other unmanned areas), which limits the adoption and application of many SL learning methods [19], [20]. In a search for more efficient unlabeled samples to overcome the need for large labeled datasets, there is a rising research interest in SSL and its applications to CD to reduce the amount of labeled data required, by either developing novel classification algorithms or adopting existing SSL frameworks [21].

At the same time, various CD methods are used to identify change information in the literature. There are mainly three CD feature extraction algorithms of spectral characterization for classification.

- 1) Simple algebraic operation: including image differencing, image ratioing, image regression, and change vector analysis (CVA) [22], absolute average difference (AAD) [36].
- 2) Image transformation: including multivariate alteration detection (MAD) [23], iteratively reweighted MAD (IR-MAD) [24], differential principal component analysis (DPCA) [25], and iterative slow feature analysis (ISFA) [26].

Manuscript received September 7, 2021; revised December 8, 2021 and January 22, 2022; accepted January 30, 2022. Date of publication February 11, 2022; date of current version March 2, 2022. This work was supported by the National Natural Science Foundation of China under Grant 62161160336 and Grant 42030111. (*Corresponding author: Lianru Gao.*)

The authors are with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: liulian0603@126.com; hongdf@aircas.ac.cn; nili@aircas.ac.cn; gaolr@aircas.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3150571

- 3) Classification detection method: including post classification comparison and compound classification. Among all these CD algorithms, image transformation methods have been extensively studied and applied.

The basic idea of image transformation is to project the original data into the transformed space to better separate the changed and unchanged pixels.

Slow feature analysis (SFA) is an image transformation CD algorithm that can extract constant and slowly varying features from input signals. Wu *et al.* [26] proposed a new ISFA algorithm to detect the presence of real changes in bitemporal multispectral datasets. The ISFA algorithm is an extension of the SFA algorithm. Through the iterative weighting method, the unchanged pixels can be highlighted, and the influence of the changed pixels can be suppressed in the calculation, so as to learn the best projection matrix. The ISFA algorithm has obtained good experimental results on two groups of real multispectral datasets. This method has better performance than other CD methods. Although ISFA methods can make use of spectral information, they are not suitable for the continuous high-dimensional spectral features derived from hyperspectral image (HSI) data. The high dimensionality of HSI data usually contain redundant information, which reduce the separability between changed and unchanged areas and increases the amount of computation [27].

Based on the above analysis of the current hyperspectral CD problems, it is obvious that we need to explore CD algorithms by focusing on two main points. First, it is often difficult for ISFA algorithm to separate the changed and unchanged pixels in HSI classification especially with limited small training samples. This observation has fostered the idea of SSL, which adds unlabeled samples to the initial training samples, without significant cost, to improve the capability of the classifier. In this article, a novel semi-supervised classification algorithm based on multilayer cascade screening strategy (MCS<sup>4</sup>CD) is put forward. In the semi-supervised process, the spatial neighborhood of labeled training samples is combined with active learning (AL) algorithm to select the most helpful unlabeled samples, which is used as the pseudo labeled set to retrain the support vector machine (SVM) classifier. Second, the performance of ISFA algorithm is degraded due to the band redundancy of HSI. To solve this problem, we designed a new CD algorithm for HSI called subspace iterative slow feature analysis (SISFA). The basic idea is projecting the original HSI into a new feature subspace to better separate changed and unchanged pixels.

The main contributions of this article are summarized as follows.

- 1) The reliability of selected unlabeled samples is increased with the proposed MCS<sup>4</sup>CD strategy that utilizes the spatial information and AL algorithm. The MCS<sup>4</sup>CD strategy reduces the tedious workload of labeling by introducing a semi-supervised method, and overcomes the limitations caused by too few label samples.
- 2) The MCS<sup>4</sup>CD strategy takes into account the positive effect of neighborhood spatial information in the semi-supervised classification process. We combine spatial information and traditional SVM algorithm to further

determine the label of unlabeled samples, which plays a positive role in improving the classification accuracy.

- 3) The SISFA method is designed for extracting the unchanged features of bi-temporal HSI data. The original HSI is projected into the subspace to better separate changed and unchanged pixels.

The rest of this article is organized as follows. In Section II, the proposed method is explained in detail, including MCS<sup>4</sup>CD strategy and SISFA method. In Section III, the HSI datasets and experimental assessment indexes are described. In Section IV, the experimental results are shown and discussed. Finally, Section V concludes this article.

## II. METHODOLOGY

### A. Subspace Iterative Slow Feature Analysis

1) *Subspace Analysis for Hyperspectral Imagery*: One of the main problems in hyperspectral CD is the presence of high dimensionality, which can be an important negative impact in the data interpretation process [28]. Feature extraction [e.g., principal component analysis (PCA)] can be used as a solution to address this issue, which is achieved by finding a new set of vectors that represents an observation, while reducing the dimensionality by transforming the input data to another low-dimensional space (referred to as subspace) and extracts informative features in the new domain [29].

2) *SISFA Change Detection*: Consider two co-registered HSI data acquired over the same geographical area at two different times. PCA is implemented on each HSI data, and the HSI cube are projected into the subspace. This process reduces data dimensions and views some principal components (PCs) as the inputs of ISFA algorithm to extract CD feature.

Given a bitemporal subspace spectral vector pair  $\mathbf{x}^i = [x_1^i, \dots, x_{\mathcal{H}}^i]$  and  $\mathbf{y}^i = [y_1^i, \dots, y_{\mathcal{H}}^i]$ , where  $\mathcal{H}$  represents the dimension of spectral subspace and  $i$  indicates the pixel number in each band. The two input data are normalized as follows:

$$\hat{x}_j^i = \frac{x_j^i - \mu_{x_j}}{\sigma_{x_j}} \text{ and } \hat{y}_j^i = \frac{y_j^i - \mu_{y_j}}{\sigma_{y_j}} \quad (1)$$

where  $\mu_{x_j}$  and  $\sigma_{x_j}$  are the mean and variance for band  $j$  of image  $x$ , respectively. SISFA is to minimize the difference in the unchanged pixels by finding a transformation function, so as to better highlight the changed regions, which can be formulated as follows:

$$\min \frac{1}{P} \sum_{i=1}^P (\mathcal{L}_j(\hat{\mathbf{x}}^i) - \mathcal{L}_j(\hat{\mathbf{y}}^i))^2. \quad (2)$$

Assuming that the transformation function  $\mathcal{L}$  is linear projection, the function can be expressed as  $\mathcal{L}_j(\mathbf{x}) = \mathbf{w}_j^T \mathbf{x}$ . And the following constraints must be satisfied:

$$s.t. \begin{cases} \frac{1}{2P} \left[ \sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{x}}^i) + \sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{y}}^i) \right] = 0 \\ \frac{1}{2P} \left[ \sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{x}}^i)^2 + \sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{y}}^i)^2 \right] = 1 \\ \frac{1}{2P} \left[ \sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{x}}^i) \mathcal{L}_l(\hat{\mathbf{x}}^i) + \sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{y}}^i) \mathcal{L}_l(\hat{\mathbf{y}}^i) \right] = 0. \end{cases} \quad (3)$$

A total of three constraints are included in formula (3). The function of constraint  $\frac{1}{2P}[\sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{x}}^i) + \sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{y}}^i)] = 0$  is to minimize the degree of spectral change over time, which is mainly measured by the mean value. Constraint condition  $\frac{1}{2P}[\sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{x}}^i)^2 + \sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{y}}^i)^2] = 1$  forces that each transformed feature should contain some information and avoid the false solution. Constraint  $\frac{1}{2P}[\sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{x}}^i)\mathcal{L}_i(\hat{\mathbf{x}}^i) + \sum_{i=1}^P \mathcal{L}_j(\hat{\mathbf{y}}^i)\mathcal{L}_i(\hat{\mathbf{y}}^i)] = 0$  means each feature to be decorrelated and carries different types of CD information. The constraint (3) is integrated into the optimization objective (2), and then the new optimization problem is required

$$\frac{\mathbf{w}_j^T \mathcal{O}_A \mathbf{w}_j}{\mathbf{w}_j^T \mathcal{O}_B \mathbf{w}_j} = \frac{\mathbf{w}_j^T \Sigma \Delta \mathbf{w}_j}{\mathbf{w}_j^T [\frac{1}{2}(\Sigma_x + \Sigma_y)] \mathbf{w}_j}. \quad (4)$$

Projection matrix  $\mathbf{w}$  can be solved by generalized eigenvalue problem  $\mathcal{O}_A \mathbf{w} = \mathcal{O}_B \mathbf{w} \Lambda$ . Where  $\Lambda$  represents diagonal matrix of the eigenvalue, the covariance matrices  $\mathcal{O}_A$  and  $\mathcal{O}_B$  are shown as follows:

$$\mathcal{O}_A = \frac{1}{P} \sum_{i=1}^P (\hat{\mathbf{x}}^i - \hat{\mathbf{y}}^i)(\hat{\mathbf{x}}^i - \hat{\mathbf{y}}^i)^T = \Sigma \Delta \quad (5)$$

$$\mathcal{O}_B = \frac{1}{2P} \left[ \sum_{i=1}^P (\hat{\mathbf{x}}^i)(\hat{\mathbf{x}}^i)^T + \sum_{i=1}^P (\hat{\mathbf{y}}^i)(\hat{\mathbf{y}}^i)^T \right] = \frac{1}{2} (\Sigma_x + \Sigma_y). \quad (6)$$

In slow feature analysis process, the most crucial work is to find an effective projection vector to extract unchanged information. Finally, the difference  $\text{SISFA}_j$  between the projected features is calculated as follows:

$$\text{SISFA}_j = \mathbf{w}_j^T \hat{\mathbf{x}} - \mathbf{w}_j^T \hat{\mathbf{y}}. \quad (7)$$

The mean value of  $\text{SISFA}_j$  variables is zero, independent of each other and obeys the standard normal distribution. Therefore, it can be considered that the square sum of the  $\text{SISFA}$  difference conforms to the  $\chi^2$  distribution with  $\mathcal{H}$  degrees of freedom

$$T^i = \sum_{j=1}^k \left( \frac{\text{SISFA}_j^i}{\sigma_{\text{SISFA}_j}} \right)^2 \in \chi^2(\mathcal{H}) \quad (8)$$

where  $T$  is the chi-square distance of pixel  $i$ ,  $\sigma_{\text{SISFA}_j}$  is the variance of  $\text{SISFA}$ ,  $\chi^2$  denotes the chi-square distribution. To highlight the spectral difference of the slowly varying unchanged pixels, the intensity of the change is weighted by the chi-square distance. Finally, the iterative weighted distance  $\mathcal{Q}$  can be formulated as follows:

$$\mathcal{Q}^i = P\{T^i > t\} = P\{\chi^2(\mathcal{H}) > t\}. \quad (9)$$

In the next iteration, the mean value and variance for the bitemporal spectral are updated as follows:

$$\mu_{x_j} = \frac{\sum_{i=1}^P \mathcal{Q}^i x_j^i}{\sum_{i=1}^P \mathcal{Q}^i} \text{ and } \sigma_{x_j} = \frac{\sum_{i=1}^P \mathcal{Q}^i (x_j^i - \mu_{x_j})^2}{\sum_{i=1}^P \mathcal{Q}^i}. \quad (10)$$

The weight is added into the whole calculation process of  $\text{SISFA}$ . We calculate the covariance matrices  $\mathcal{O}_A$  and  $\mathcal{O}_B$  with

the weight

$$\mathcal{O}_A = \frac{1}{\sum_{i=1}^P \mathcal{Q}^i} \sum_{i=1}^P \mathcal{Q}^i (\hat{\mathbf{x}}^i - \hat{\mathbf{y}}^i)(\hat{\mathbf{x}}^i - \hat{\mathbf{y}}^i)^T \quad (11)$$

$$\mathcal{O}_B = \frac{1}{2 \sum_{i=1}^P \mathcal{Q}^i} \left[ \sum_{i=1}^P \mathcal{Q}^i (\hat{\mathbf{x}}^i)(\hat{\mathbf{x}}^i)^T + \sum_{i=1}^P \mathcal{Q}^i (\hat{\mathbf{y}}^i)(\hat{\mathbf{y}}^i)^T \right]. \quad (12)$$

## B. Semi-Supervised Classification Based on Multilayer Cascade Screening Strategy

SSL can use a large number of “cheap” unlabeled data to improve the classification accuracy on few labeled training examples. The proposed MCS<sup>4</sup>CD strategy combines spatial neighborhood information (SNI) extraction strategy with AL algorithm to select the most informative unlabeled samples as the pseudo labeled set to further improve classification performance. The specific procedures of MCS<sup>4</sup>CD are shown in Fig. 1.

1) *Semi-Supervised Classification Based on Spatial Neighborhood Information*: If the label categories of unlabeled samples are determined only by the primary SVM classification map, it is difficult to ensure satisfactory accuracy. This is because the primary classification accuracy is not high, the labels of the candidate set are misclassified. Therefore, the subsequent SSL process will be affected by the error labels, resulting in error accumulation. In this article, MCS<sup>4</sup>CD strategy is constructed to help the SVM classifier label the selected unlabeled samples.

Step 1): Circular neighborhood (CN): We empirically find out that four or eight neighborhood are usually used to obtain the SNI. It only covers a small area within a fixed radius, which obviously cannot meet the needs of different sizes. Four or eight neighborhood windows are too small for searching useful unlabeled samples. In this article, we adopt a CN window, which can adjust the search radius  $d$  to find the optimal size (including four- and eight-neighborhood).

Step 2): SNI extraction strategy: “First Law of Geography” reveals that the attribute values of all objects on the earth are related to each other, but near values are more correlated than distant ones. This law gives us an important assumption that the label category of unlabeled samples should be consistent with the existing training sample categories in the spatial neighborhood area. However, in the process of determining unlabeled samples, the positive effect of SNI on SSL method is often ignored. Based on the initial classification results (the SVM classification map is the first sample screening) and SNI-CN (the second spatial screening), the labels of unlabeled samples are screened for the third time. The third screening strategy eliminates the negative effects of low quality and wrong labels.

2) *Selecting the Most Informative Unlabeled Samples Based on Active Learning*: In the SSL process, a large number of unlabeled data points are selected as the candidate sets. To further simplify the samples, an AL strategy is used to select the

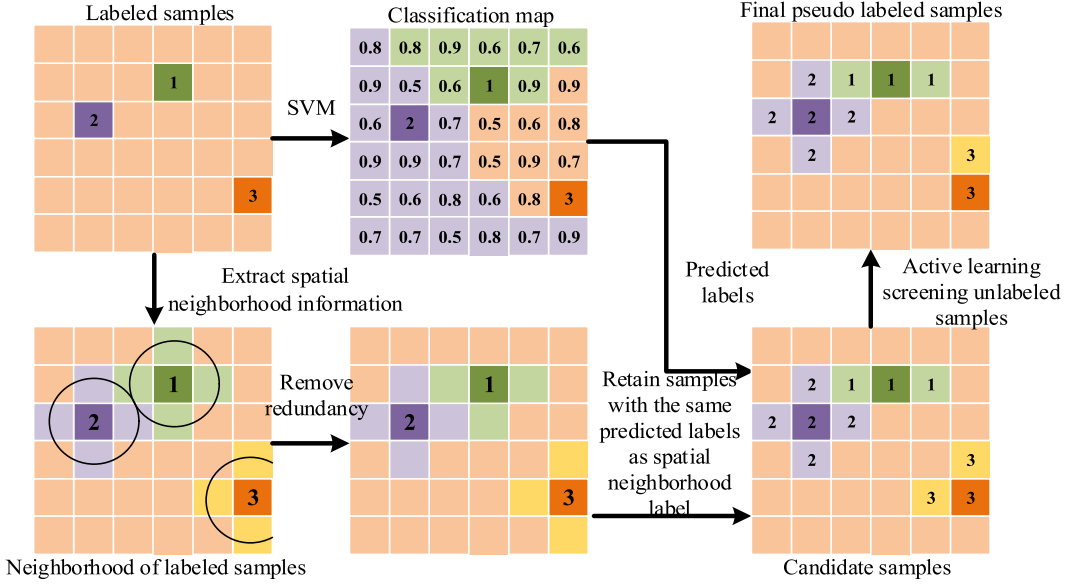


Fig. 1. Flowchart of the proposed  $MCS^4CD$  method (Take the radius  $d = 0.5$  as an example). The entire label screening process is cascaded, and the output of each step is used as the input of the next step.

most informative unlabeled samples from the candidate sets. AL aims to carefully choose the samples to be labeled to achieve a higher accuracy while using as few requests as possible, thereby minimizing the cost of obtaining labeled data. After the third screening step, the unlabeled samples with high confidence or most informative are selected as the final pseudo samples through AL (the fourth screening).

In this article, breaking ties (BT) query strategy of AL algorithm is used to collect the most informative unlabeled samples. The decision criterion of BT is as follows:

$$\hat{x}_i^{BT} = \arg \min_{x_i \in x_U} \left\{ \max_{k \in \mathcal{K}} p(y_i = k | \mathbf{x}_i) - \left\{ \max_{k \in \mathcal{K} \setminus \{k^+\}} p(y_i = k | \mathbf{x}_i) \right\} \right\} \quad (13)$$

$$p(y = k | \mathbf{x}_i) = \frac{1}{1 + \exp(\mathbf{A}\mathbf{f}(\mathbf{x}) + \mathbf{B})} \quad (14)$$

where  $k^+ = \arg \max_{k \in \mathcal{K}} p(y_i = k | \mathbf{x}_i)$  represents the class label corresponding to the largest posterior probability for sample  $x_i$ , and  $k \in \mathcal{K} \setminus \{k^+\}$  represents the interested class labels excluding  $k^+$ .  $p$  is provided the probabilistic outputs by the probability model-based SVM.

### C. Procedure of the Proposed $MCS^4CD$

To increase the reliability of selected unlabeled samples, an SSL method that is based on the  $MCS^4CD$  strategy is adopted. The detailed strategy is described as follows.

## III. EXPERIMENTS

### A. Datasets

To demonstrate the effectiveness of the proposed method, we will display the numerical results on four HSI datasets.

---

### Algorithm: $MCS^4CD$ .

---

Inputs: bi-temporal HSI images

Parameters of the model: spectral dimension of subspace feature  $\mathcal{H}$ , radius  $d$  of CN, the number of training samples for each class  $p$ .

Output: the CD classification result for  $D_C \cup D_L$

Steps:

- 1) Initialize training samples  
 $D_L = \{(x_1, y_1), \dots, (x_1, y_l)\}$ ;
  - 2) Extract changed and unchanged features  $x \in \mathbb{R}^s$  using the SISFA method;
  - 3) Train SVM probability model to predict the label  $\hat{y} = \{\hat{y}_i, i = 1, \dots, n_{pre}\}$  of unlabeled samples;
  - 4) Select a circle neighborhood which takes the selected  $D_L$  as the center and remove redundancy samples (including background information and repeated selection of training samples). Retain samples with the same  $\hat{y}$  labels as spatial neighborhood label. The extracted candidate samples are denoted as  $D_U = \{(x_1, y_1), \dots, (x_1, y_{n_d})\}, i = 1, \dots, n_d$ ;
  - 5) Simplify  $D_U$  to  $D_C$  by using the AL algorithm. The  $D_L$  is expanded with each iteration of BT. Then, update the labeled sample sets  $D_{C+L} = \{D_C, D_L\}$ ;
  - 6) Concatenate  $D_C$  and  $D_L$ , and then test the classification performance of the final pseudo labeled samples using the SVM classifier.
- 

- 1) *The Hermiston Dataset*: This dataset was collected by Hyperion sensor over the Hermiston city area (Oregon) on the years 2004 and 2007. The scene illustrates round farmland, whose spatial dimensions are  $390 \times 200$  pixels and includes 242 spectral bands. After data preprocessing,

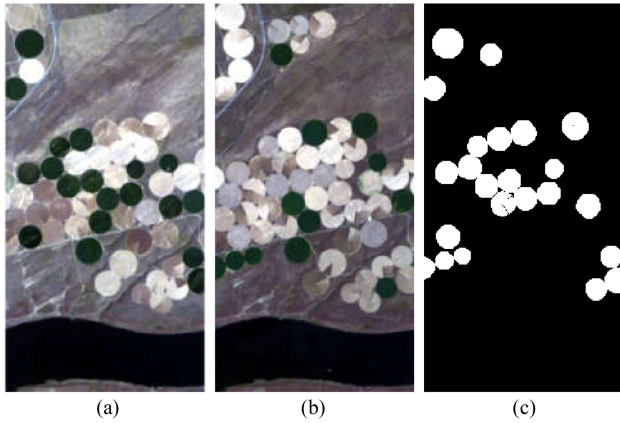


Fig. 2. Hermiston Dataset and land-cover ground truth map. (a) Image acquired in 2004. (b) Image acquired in 2007. (c) Ground Truth.

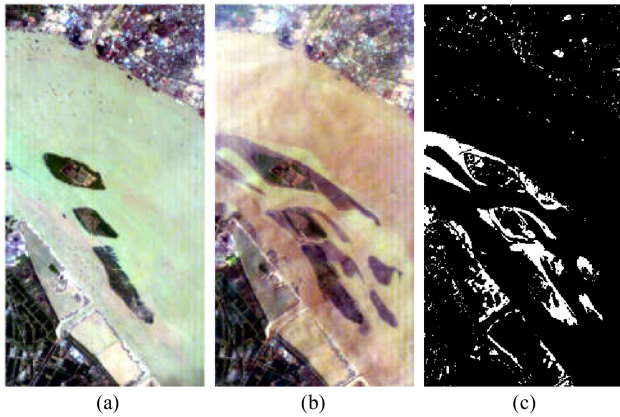


Fig. 3. River Dataset and land-cover ground truth map. (a) Image acquired on May 3, 2013. (b) Image acquired on December 31, 2013. (c) Ground Truth.

the spectral bands affected by noise and water absorption (1–7, 58–76, 121–126, 167–180, 222–224, 225–242) are removed, and the remaining 175 spectral bands are involved in the following experiment. The true color composite image of the Hermiston dataset and its corresponding land-cover CD map are shown in Fig. 2(a)–(c). The reference ground truth of the CD map contains 2 classes, which white and black colors indicate the change and nonchange areas, respectively.

- 2) *The River Dataset*: This dataset was collected by Hyperion sensor on May 3, 2013 and December 31, 2013 in Jiangsu Province, China. The spatial resolution of this image is 30 m covering spectral bands from 0.4 to 2.5  $\mu\text{m}$ . The image is consisting of 463 by 241 pixels. Some spectral bands were removed due to the noise, and the remaining 198 spectral dimensions were used. The main type of change in this dataset is the reduction of river channel. The true color composite image and the corresponding land-cover CD map are shown in Fig. 3.
- 3) *The USA Dataset*: This dataset illustrates an irrigated agricultural field of Hermiston city in Umatilla County, Oregon, OR, the USA, which was collected on

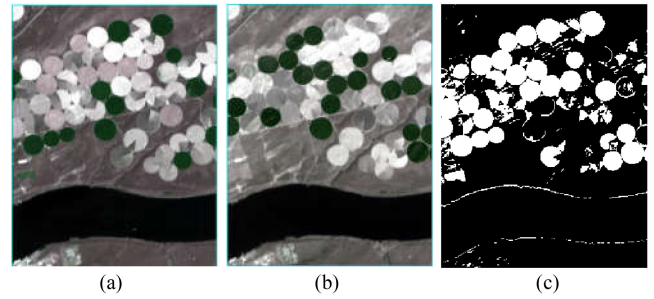


Fig. 4. USA Dataset and land-cover ground truth map. (a) Image acquired on May 1, 2004. (b) Image acquired on May 8, 2007. (c) Ground Truth.

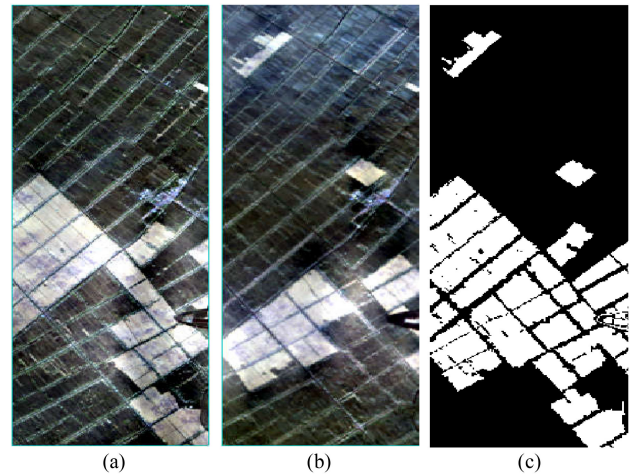


Fig. 5. China Dataset and land-cover ground truth map. (a) Image acquired on May 3, 2006. (b) Image acquired on April 23, 2007. (c) Ground Truth.

May 1, 2004, and May 8, 2007, respectively. The size of this dataset is 307 lines by 241 samples, with 154 spectral bands. The land cover types include soil, irrigated fields, building, river, cultivated land and grassland. For this dataset, all changes related to the type of land cover and river. The true color composite image of the USA dataset and its corresponding land-cover CD map are shown in Fig. 4.

- 4) *The China Dataset*: This dataset illustrates a farmland near the city of Yancheng Jiangsu province in China, which was collected on May 3, 2006, and April 23, 2007, respectively. The size of this dataset is 420 lines by 140 samples, with 154 spectral bands. This scene is mainly a combination of soil, road, and agricultural field. For this dataset, all changes related to the type of land cover and river. The composite image of the China dataset and its corresponding CD map are shown in Fig. 5.

#### B. CD Accuracy Assessment Indexes

In order to evaluate the performance of the CD methods, the accuracy of the experimental results is evaluated by the confusion matrix. The related assessment criteria include overall accuracy (OA), kappa coefficient (Kappa), missing alarm (MA)

TABLE I  
CLASSIFICATION ACCURACY IN RELATION TO DIFFERENT CHANGE DETECTION METHODS FOR THE HERMISTON DATASET

Method	CVA	AAD	TDI-GC	MAD	IR-MAD	ISFA	DPCA	SISFA	MCS <sup>4</sup> CD
OA (%)	92.48	82.79	87.44	74.36	91.64	88.17	93.43	93.98	<b>97.84</b>
Kappa	0.7257	0.2429	0.2762	0.3427	0.6978	0.6163	0.7544	0.7714	<b>0.9075</b>
MA	0.0823	0.0962	0.1025	0.2757	0.0891	0.1335	0.0717	0.0655	<b>0.0198</b>
FA	0.0042	0.1018	0.0337	0.0248	0.0074	<b>0.0026</b>	0.0039	0.0038	0.0050

rate, and false alarm (FA) rate. Previous studies have provided the meanings and calculation method for these assessment indexes [30]–[33].

#### IV. RESULTS AND DISCUSSION

##### A. Comparison With Other Change Detection Methods

To verify the effectiveness of the proposed method, we compare MCS<sup>4</sup>CD with other CD methods under the limited labeled training samples. In this experiment, the initial number of labeled samples per class  $p$  is fixed. On each HSI dataset, we take  $p$  samples from each class for first training, and the remaining pixels are used for testing. Comparison methods include CVA [22], AAD [36], TDI-GC [37], MAD [23], IR-MAD [24], ISFA [25], and DPCA [26], which employ different techniques to detect changes, such as simple algebraic operation and image transformation. The first three algorithms (i.e., CVA, AAD, and TDI-GC) are unsupervised CD methods, which do not require prior information of samples. The last four algorithms (i.e., MAD, IR-MAD, ISFA, and DPCA) and the proposed SISFA are designed in a supervised mode, so the ground reference data are needed in practical application. Supervised method uses labeled training data to achieve accurate CD results. Therefore, note that these methods are supervised and require HSI reference samples to train SVM classifiers.

The proposed MCS<sup>4</sup>CD is based on the CD feature extraction algorithm of SISFA followed by a semi-supervised classification of the difference image using multilayer cascade screening strategy. In this framework, the binary CD map is obtained in the semi-supervised classification step, and richer change information is captured in the SISFA feature extraction step. In the case of limited training samples, the SISFA algorithm can show better performance, so the screening strategy is implemented based on the SISFA algorithm.

In the experimental setup, in order to verify the effectiveness of each step of the proposed method, we conducted an ablation study. ISFA without spectral subspace constraint, DPCA without slow feature extraction, and SISFA without cascade screening strategy were compared with our proposed method. Moreover, MCS<sup>4</sup>CD algorithm was implemented by using the SISFA algorithm and cascade screening strategy. The comparison results of Hermiston, River, USA, and China datasets are shown in Tables I–IV and Figs. 6–9, respectively. The bold values indicate the best performance among the methods in each case. Take the average of ten independent operation results.

For the Hermiston dataset, we report the results of the proposed two methods and the comparison methods in Table I. About 20 samples were selected for each class labeled

training samples. For the proposed MCS<sup>4</sup>CD algorithm, we added 100 samples, which constitute the unlabeled set, to update the pseudo labeled set. MCS<sup>4</sup>CD is the same as other algorithms, and the number of initial training sample labels is 20. Statistical analysis shows that the OA and Kappa for the MCS<sup>4</sup>CD improve by 9.67% and 0.2912 compared with ISFA method, by 4.41% and 0.1531 compared with the DPCA method. The performance of SISFA is better than ISFA, and its OA and Kappa are increased by 5.81% and 0.1551, respectively. The lowest FA (FA = 0.0026) is achieved when using ISFA, and the best results appear in boldface. As can be observed in Fig. 6, MCS<sup>4</sup>CD is capable of building better classification performance that compensates for the lack of labeled training data. MCS<sup>4</sup>CD projects the original data into a new transformed space to better separate the changed and unchanged pixels.

For the River dataset, the CD accuracy achieved by these methods are illustrated in Table II. The parameter  $p$  was set to 50, because this dataset required a relatively larger number of training samples. If the parameter  $p$  were set as above, the OA of the MAD method dropped below 15%. Therefore, we selected 50 samples for each class, and then selected 40 unlabeled samples from  $\mathbf{D}_L$  as the pseudo labeled set. Most of existing CD method may not work well for the situation of limited samples. The visual classification results of the River dataset are shown in Fig. 7.

From Fig. 7(d)–(f), it can be seen that MAD, IR-MAD, ISFA-based methods are distributed a large number of white patches. These three algorithms are affected by limited samples and high dimensionality, which suppresses the separation between the change and the background. Statistical analysis shows that the OA and Kappa for the MCS<sup>4</sup>CD improve by 5.08% and 0.1294 compared with the DPCA method, by 7.14% and 0.1747 compared with the SISFA method. Although both DPCA and SISFA are based on dimensionality reduction, their experimental results are different. The performance of SISFA is lower than DPCA, and its OA and Kappa are reduced by 2.06% and 0.0453, respectively. The proposed SISFA algorithm is slightly unsatisfactory in terms of both OA and kappa. This poor performance is mainly due to the implementation of dimensionality reduction, SISFA loses part of the information, resulting in worse performance than DPCA.

For the USA and China datasets, parameters were set to the same values in the experiment. The initial labeled samples of these two datasets selected 100 samples for each class, and then selected 40 unlabeled samples as the pseudo labeled set. The classification accuracy of each CD method is listed in Tables III and IV. Statistics show that the OA and Kappa of MCS<sup>4</sup>CD

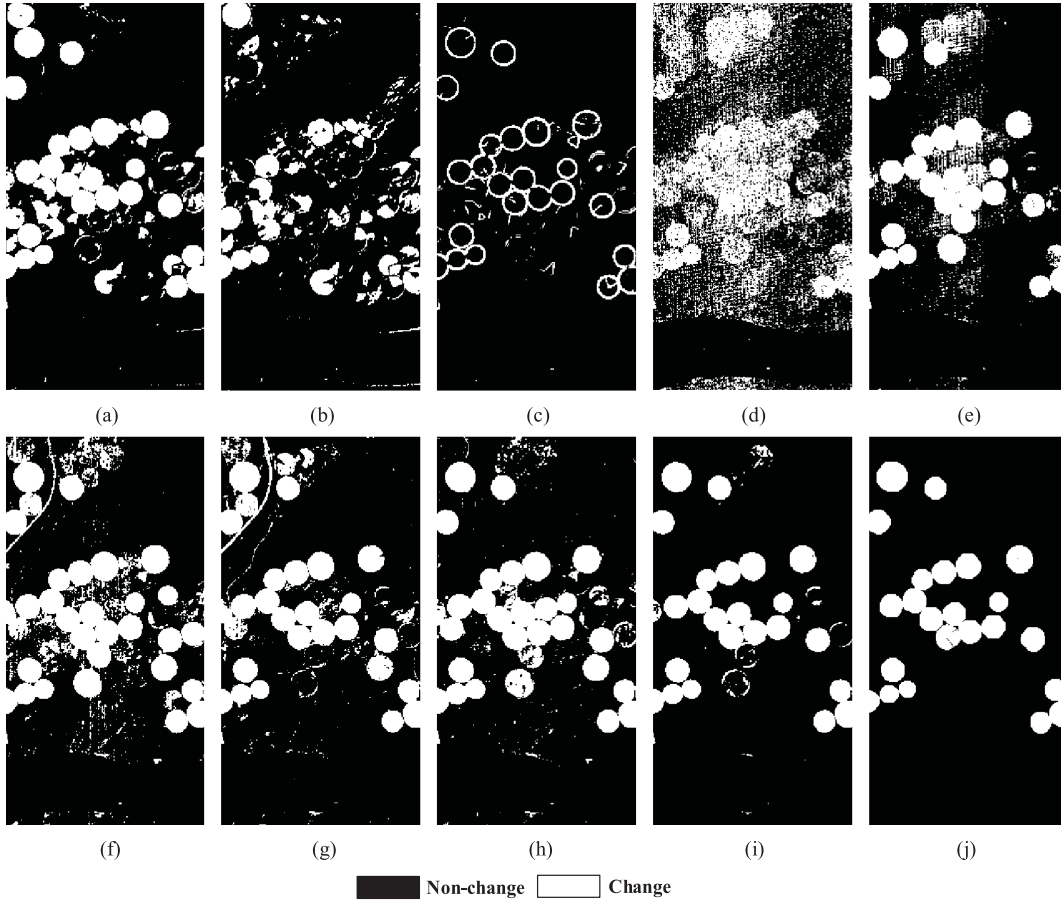


Fig. 6. Classification results of different change detection methods for the Hermiston dataset. (a) CVA. (b) AAD. (c) TDI-GC. (d) MAD. (e) IR-MAD. (f) ISFA. (g) DPCA. (h) SISFA. (i) MCS<sup>4</sup>CD. (j) Ground truth.

TABLE II  
CLASSIFICATION ACCURACY IN RELATION TO DIFFERENT CHANGE DETECTION METHODS FOR THE RIVER DATASET

Method	CVA	AAD	TDI-GC	MAD	IR-MAD	ISFA	DPCA	SISFA	MCS <sup>4</sup> CD
OA (%)	90.82	89.53	80.20	48.64	65.71	75.51	88.38	86.32	<b>93.46</b>
Kappa	0.6011	0.5040	0.2222	0.0638	0.1178	0.2161	0.5155	0.4702	<b>0.6449</b>
MA	0.0975	<b>0.0243</b>	0.0514	0.5397	0.3402	0.2360	0.1175	0.1409	0.0526
FA	<b>0.0033</b>	0.0921	0.1719	0.0472	0.0508	0.0405	0.0110	0.0103	0.0196

is 13.41% and 0.2969 greater than that of ISFA for the USA dataset, and 7.39% and 0.1541 greater than that of DPCA for the China dataset. Figs. 8 and 9 detail the visual comparison of seven different methods. We can see obviously that Fig. 8(i) is more accurate than Fig. 8(d)–(f) for the USA dataset, and Fig. 9(g)–(i) is visually more precise than Fig. 9(a)–(f) for the China dataset, which indicates that the proposed MCS<sup>4</sup>CD algorithm is more accurate. It can be concluded that the proposed MCS<sup>4</sup>CD are valid for change and superior to the CD methods, in terms of small-size training set. Few PCs capture most of the change information, and use slow feature analysis to further enhance the change, so that the classifier can identify the change in different components.

*B. Effect of Labeled Training Sample Number*

In this section, four real HSI images are used to evaluate the effect on using different number  $p$  of unlabeled samples. In order to illustrate the performance of the proposed MCS<sup>4</sup>CD method, we deliberately set a very small initial training samples, such as 20, 50, or 100 in each class. Correspondingly, the step size of increasing training samples is set to 20, 50, and 50. For more intuitive observation, the classification performance of different algorithms of the four HSI datasets is expressed in the form of histogram, as shown in Fig. 10. The histogram gives the OA accuracy obtained by different CD algorithms (including supervised and the proposed methods) with different  $p$ . Since

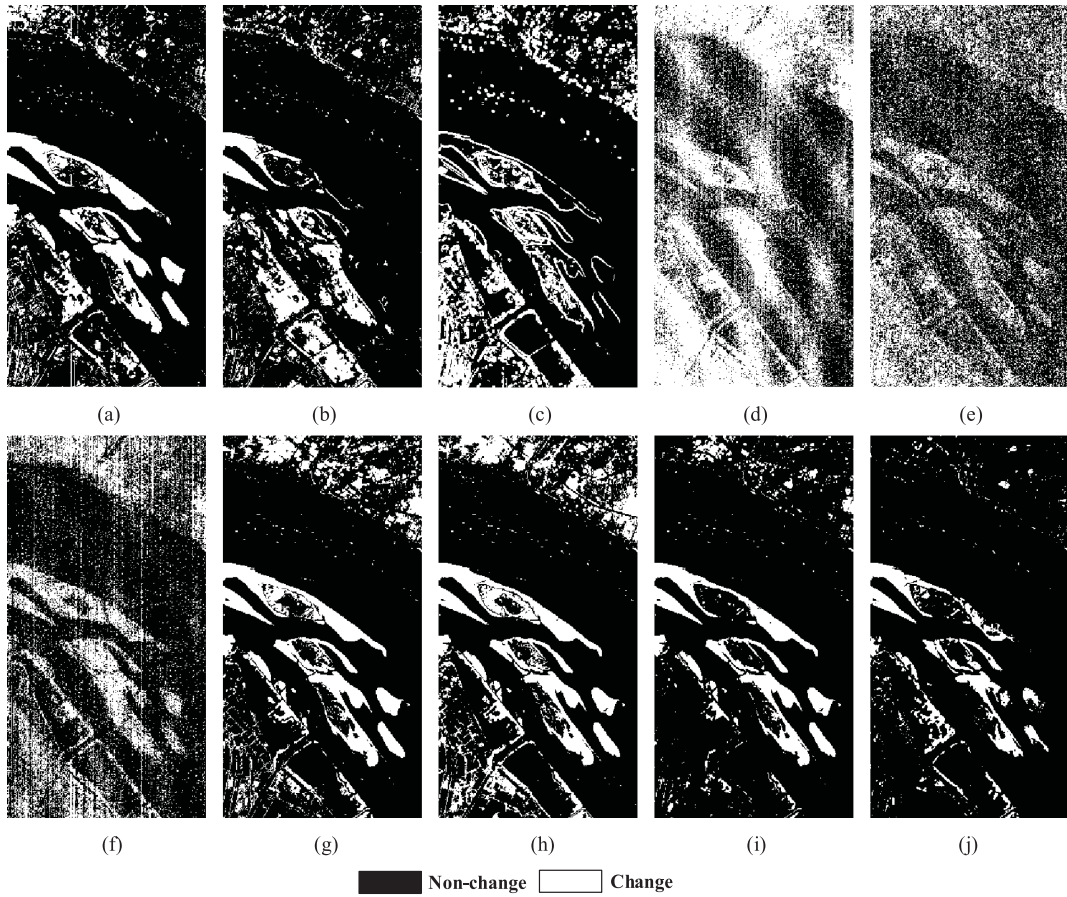


Fig. 7. Classification results of different change detection methods for the River dataset. (a) CVA. (b) AAD. (c) TDI-GC. (d) MAD. (e) IR-MAD. (f) ISFA. (g) DPCA. (h) SISFA. (i) MCS<sup>4</sup>CD. (j) Ground truth.

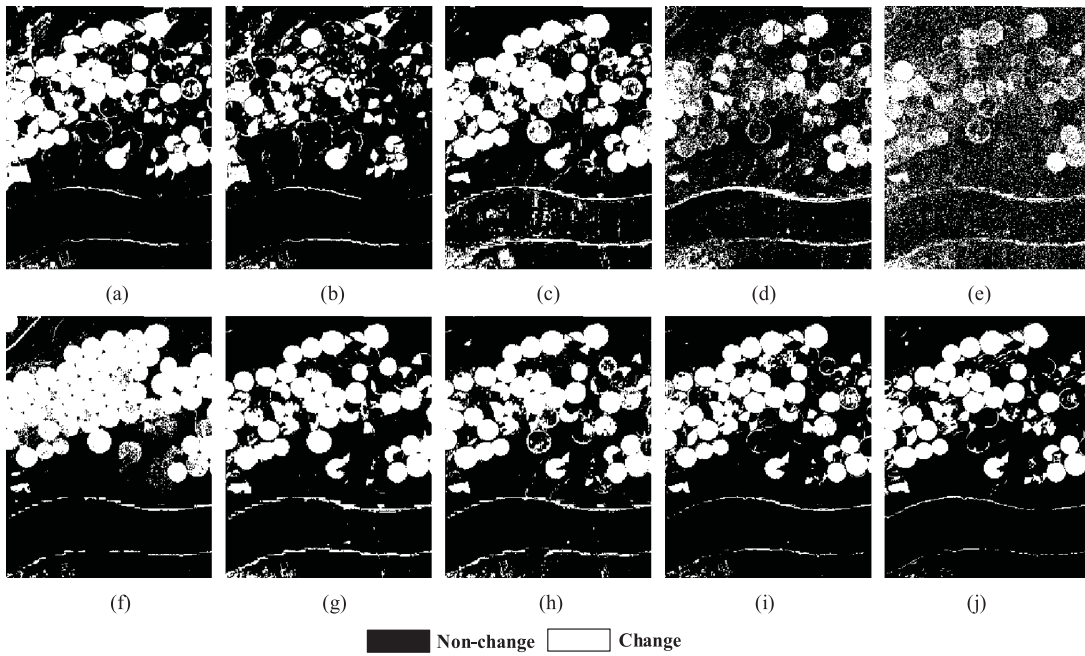


Fig. 8. Classification results of different change detection methods for the USA dataset. (a) CVA. (b) AAD. (c) TDI-GC. (d) MAD. (e) IR-MAD. (f) ISFA. (g) DPCA. (h) SISFA. (i) MCS<sup>4</sup>CD. (j) Ground truth.



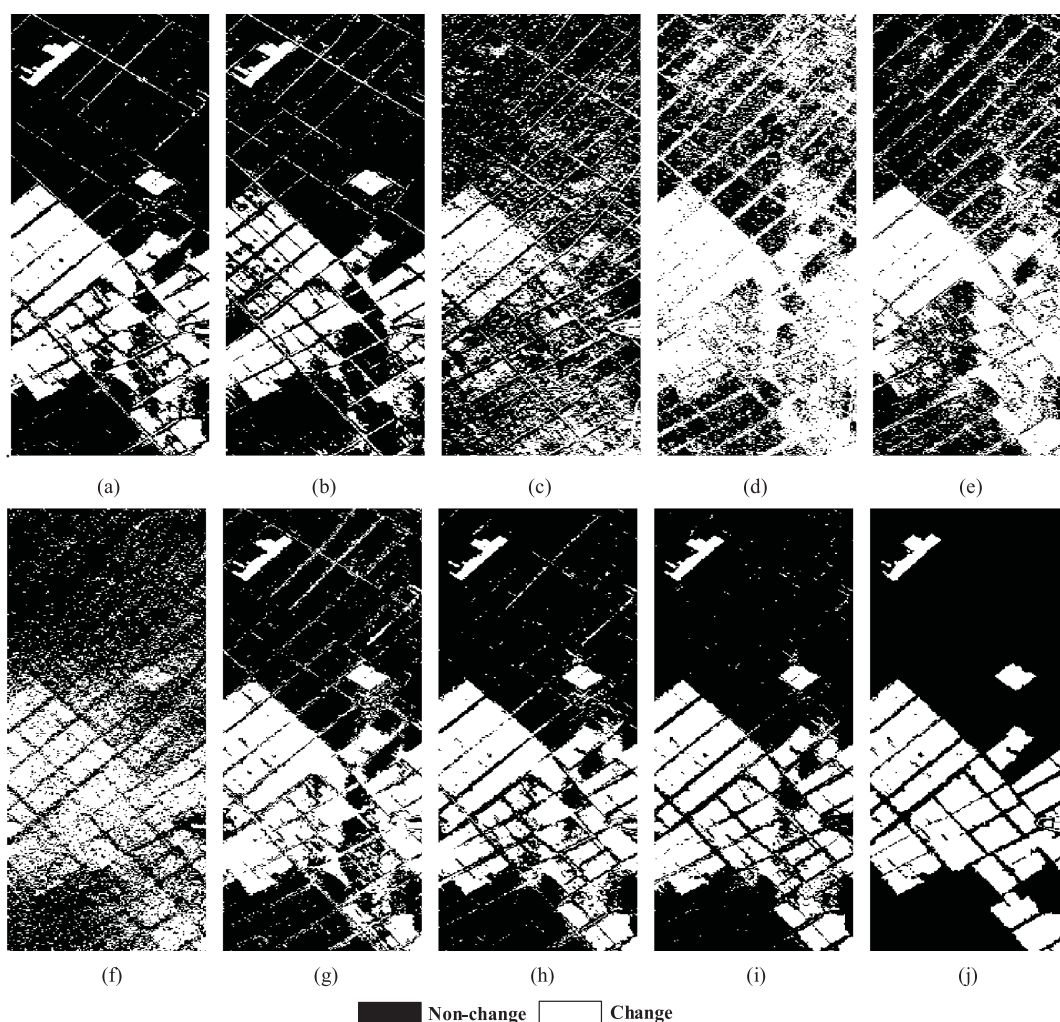


Fig. 9. Classification results of different change detection methods for the China dataset. (a) CVA. (b) AAD. (c) TDI-GC. (d) MAD. (e) IR-MAD. (f) ISFA. (g) DPCA. (h) SISFA. (i) MCS<sup>4</sup>CD. (j) Ground truth.

TABLE III  
CLASSIFICATION ACCURACY IN RELATION TO DIFFERENT CHANGE DETECTION METHODS FOR THE USA DATASET

Method	CVA	AAD	TDI-GC	MAD	IR-MAD	ISFA	DPCA	SISFA	MCS <sup>4</sup> CD
OA (%)	92.11	83.57	87.98	83.13	74.83	81.05	89.42	91.58	<b>94.46</b>
Kappa	0.7911	0.4938	0.6996	0.5226	0.3495	0.5489	0.7176	0.7743	<b>0.8458</b>
MA	0.0845	0.1276	0.1375	0.1139	0.2068	0.2057	0.0983	0.0822	<b>0.0476</b>
FA	<b>0.0186</b>	0.0772	0.0201	0.1050	0.1296	0.0467	0.0407	0.0280	0.0245

TABLE IV  
CLASSIFICATION ACCURACY IN RELATION TO DIFFERENT CHANGE DETECTION METHODS FOR THE CHINA DATASET

Method	CVA	AAD	TDI-GC	MAD	IR-MAD	ISFA	DPCA	SISFA	MCS <sup>4</sup> CD
OA (%)	87.88	81.20	69.40	64.04	74.08	78.58	84.20	88.94	<b>91.59</b>
Kappa	0.7199	0.5463	0.3031	0.3334	0.4637	0.5341	0.6472	0.7468	<b>0.8013</b>
MA	0.0939	0.1573	0.2406	0.4755	0.2925	0.2138	0.1498	0.0940	<b>0.0593</b>
FA	0.0834	0.1067	0.2122	0.0840	0.1068	0.1106	0.0860	0.0688	<b>0.0642</b>

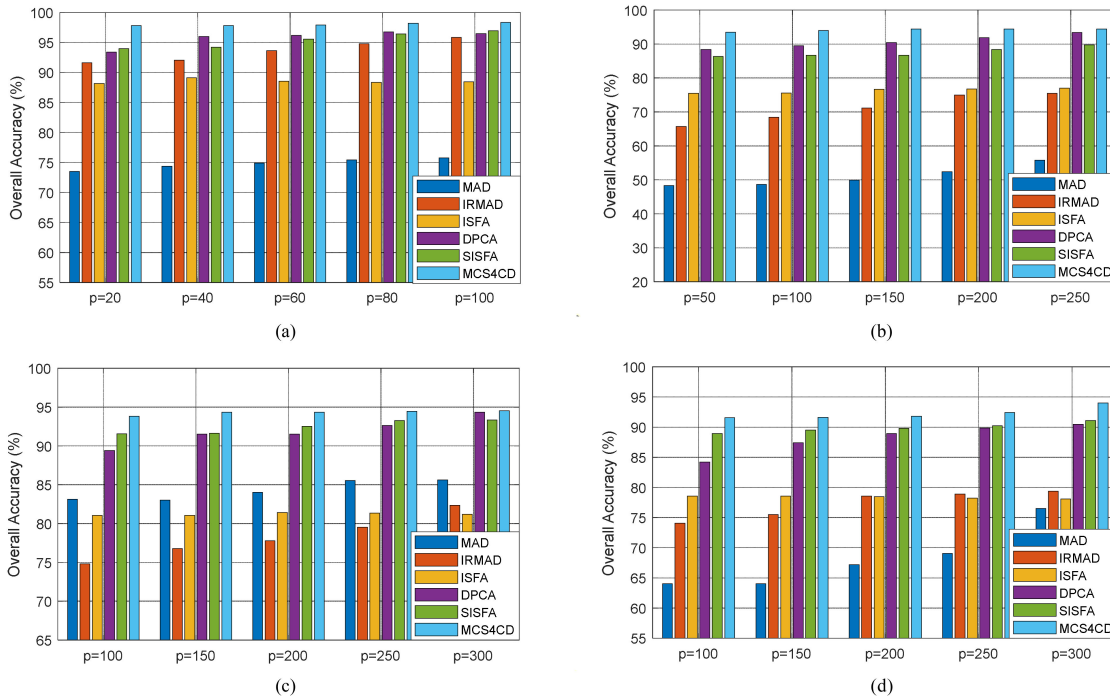


Fig. 10. Classification accuracy in relation to different number of labeled training samples. (a) Hermiston Dataset. (b) River dataset. (c) USA dataset. (d) China dataset.

unsupervised CD algorithm does not need training samples, CVA, AAD, and TDI-GC algorithms are not involved in this experiment.

As shown in Fig. 10, the OA values are increased when using more labeled samples. It indicates that the number of labeled samples affects the OA accuracy that the algorithm can achieve. The more labeled samples used initially, the better the final result. In addition, it can also be observed that the proposed MCS<sup>4</sup>CD algorithm produces consistently higher OA than the other supervised CD algorithm. For example, at  $p = 150$ , the OA of the proposed MCS<sup>4</sup>CD algorithm is 13.28% greater than that of the ISFA for the USA dataset and 4.21% greater than that of the DPCA for the China dataset. The proposed MCS<sup>4</sup>CD method can achieve more competitive results compared to the other five CD methods with limited labeled samples.

### C. Effect of the Spectral Subspace Dimension

Detailed spectral information provided by HSI data present both challenges and opportunities. The unique characteristics of high-dimensional data can usually be represented in a lower dimensional space (i.e., spectral subspace) without losing considerable information about class separability. To analyze the influence of the number of input dimension on the final accuracy and determine the parameter  $\mathcal{H}$ , we set the different dimension and test their accuracy in this section. Fig. 11 presents the OA curves of MCS<sup>4</sup>CD in relation to the parameter  $\mathcal{H}$  for the four datasets.

Several observations can be made from Fig. 11.

1) The OA curves rise with the increase of  $\mathcal{H}$ .

In Table I, the OA of ISFA algorithm can reach 88.17%. For example, at  $\mathcal{H} = 10$ , the OA of the proposed SISFA algorithm

is 5.81% greater than that of the ISFA. It reveals that full dimensionality is worse than subspace in terms of OA accuracies. Feature extraction-based dimensionality reduction helps in data compression, thereby eliminating redundant features and improving classification accuracy.

2) When dimension  $\mathcal{H}$  increases, with a constant and limited number of training samples, a higher OA can be obtained.

After the feature dimension increases to a certain critical point, continuing to increase will cause the performance of the classifier to deteriorate. This problem can be explained by the Hughes phenomenon. In order to achieve better classification effect, Hughes phenomenon can be effectively avoided in the experiment. The method we adopt is to select the SVM classifier with good classification performance for small samples. We can also increase the features for classification within a certain range and select a sufficient number of training samples, so that the number of training samples has a greater ratio to the dimension of features.

3) According to Fig. 11(b), the OA curve shows an oscillating trend.

This is because of the introduction of feature extraction algorithms. The main reason is that PCA is to eliminate the correlation between variables, and assumes that this correlation is linear, and it cannot get good results for nonlinear dependence. Therefore, the solution space will be unstable, which may lead to the incoherent and fluctuation of OA results.

### D. Effect of the Suitable Search Radius $D$

For the MCS<sup>4</sup>CD, one of the key questions is how to confirm the suitable radius  $d$ , which influences the CD accuracy and

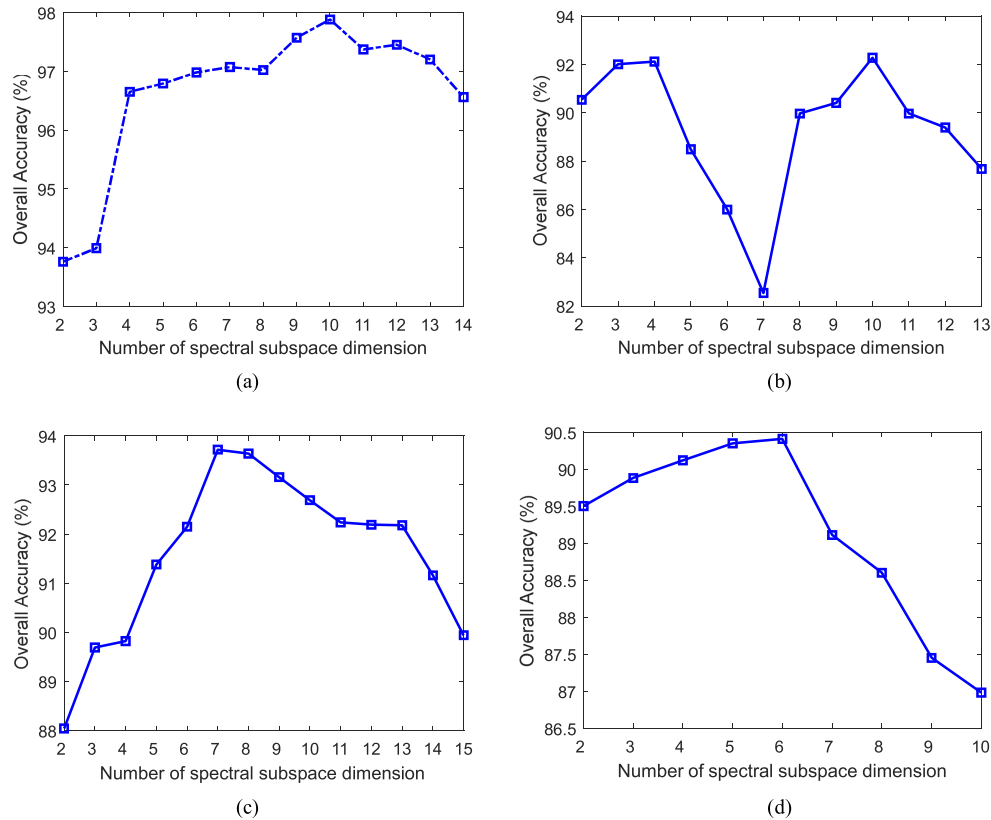


Fig. 11. Classification accuracy in relation to different spectral subspace dimension. (a) Hermiston Dataset. (b) River Dataset. (c) USA Dataset. (d) China Dataset.

the numbers of selected unlabeled samples (i.e., pseudo labeled samples). To analyze the influence of SNI extraction strategy on the final accuracy and determine the optimal parameter  $d$ , we set the three different radii and test their accuracy in this section. Fig. 12 shows the OA curves of MCS<sup>4</sup>CD with respect to the unlabeled samples and the radius of CN ( $d$ ) for the four datasets.

Fig. 12(a) shows the result of the Hermiston dataset. The OA becomes higher with the increase of the number of pseudo labeled sample, when the initial training samples and search radius  $d$  are fixed. When the number  $s$  of pseudo labeled samples is 0, 20, 40, 60, 80, 100, 120 and the search radius  $d$  ranges from 1 to 3, the results corresponding  $d = 2$  or  $d = 3$  are acceptable. Fig. 12(b)–(d) show the results of the other three datasets, respectively. The OA are first increased and then decreased with the increase of the pseudo labeled samples. When the number of pseudo labeled samples is 0, 20, 40, 60, 80, 100, 120 and the search radius  $d$  ranges from 1 to 3, the results corresponding  $d = 2$  is the best for the USA dataset. The comparison results of River and China datasets show that the search radius is the same, and  $d$  is 3.

Examining the four pictures, all the considered radius can significantly improve the initial supervised classification performances. The classification accuracy can be improved by the distribution information of unlabeled samples. The greatest OA accuracy can be obtained at around  $s = 40$ . As shown in Fig. 12(b) and (d), the curves rise as  $s$  increases to  $s = 40$  or  $s = 60$ , and afterward, they decrease gradually as  $s$  increases. This is due to the accumulation of errors caused by too many

unlabeled samples in the labeling process. Pseudolabeling is the process of using labeled data to predict unlabeled data and label it. First, the model has been trained on a dataset containing labels, and the model is used to generate pseudo labels for unlabeled samples. Finally, the original labels and pseudo labels are combined together for final model training. Therefore, the training model is easy to be affected by wrong labels, which makes the misclassification spread. In order to avoid the negative effect of incorrect labels on classification, an appropriate neighborhood radius and number of unlabeled should be selected.

### E. Change Detection in Complex Background

In order to achieve better experimental accuracy, most of the currently published CD datasets are developed under the conditions of good visual conditions, clear pictures, and simple environment backgrounds. However, in real life scenarios, the changes of ground objects are often diverse, and the environment background is more complex. In our experiments, for further testing the performance of the algorithm in a complex background, Xiong County dataset is selected as the experimental object to verify the effectiveness. Compared with the four published datasets, the flying height of Xiong County is relatively low, so the captured scene background is complex and includes more ground objects, such as buildings, trees, grassland, cars, and foodstuff in residential areas.

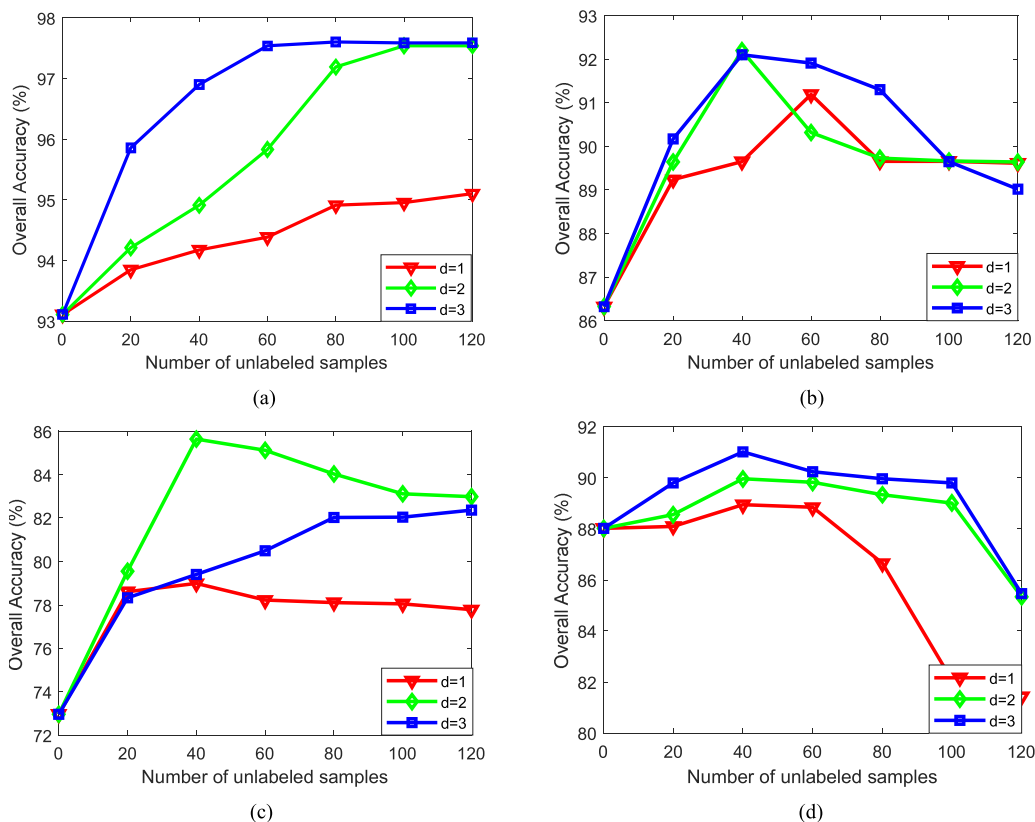


Fig. 12. Classification accuracy in relation to different radii. (a) Hermiston Dataset. (b) River Dataset. (c) USA Dataset. (d) China Dataset.

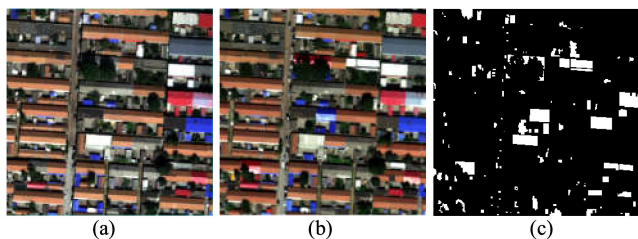


Fig. 13. Xiong County dataset and land-cover ground truth map. (a) Image acquired on October 3, 2017. (b) Image acquired on September 22, 2018. (c) Ground Truth.

This dataset shows the residential areas of Xiong County, Xiong'an New Area, China, which were collected on October 3, 2017, and September 22, 2018, respectively. An aerial hyperspectral remote sensing image of Xiong County was acquired using the visible and near infrared imaging spectrometer designed by Shanghai Institute of Technical Physics, Chinese Academy of Sciences. The flight height is 2000 m. The spectral range is 400–1000 nm, with 250 bands and a spatial resolution of 0.5 m. This image size is  $300 \times 300$  pixels. Between the two acquisition dates, some changes related to residents' lives between the areas occurred on the ground. The true color composite image of the Xiong County dataset and its corresponding land-cover CD map are shown in Fig. 13. In particular, three different kinds of change can be observed:

- 1) Changes in building roofing materials (i.e., yellow tile roof to blue colored steel). Peeling, aging, and fading of roof paint coating. Changes in roofs related to saturation problems of the sensor.
- 2) Seasonal changes in tree and vegetation areas. (i.e., with the growth of the year, the shape and volume of the tree crown change, and the branches and leaves gradually become denser). And the changes that the dense tree crown visually obscures the buildings below.
- 3) Changes along daily life of local residents (i.e., residents dried foodstuff under their houses and cars parked by the road). To quantitatively analyze this dataset, we obtained reference change maps based on visual interpretation and manual delineation.

The CD maps obtained by different methods on the Xiong County dataset are shown in Fig. 14. Table V lists the classification accuracy obtained by nine algorithms. These traditional unsupervised methods, except for the TDI-GC algorithm, obtain more obvious structural outline of the building area in the CD result map. It can be observed that many unchanged regions are detected as the changed ones in Fig. 14(a)–(c). For the binary CD tasks, CVA achieves an OA of 79.13%. On the contrary, the CD maps obtained by the methods under the supervised and semi-supervised framework have a relatively clear background in Fig. 14(d)–(i). The OA values of the five supervised methods are greater than 82.61%. Among them, it is clear that DPCA obtains the better change detection result, where the values of OA

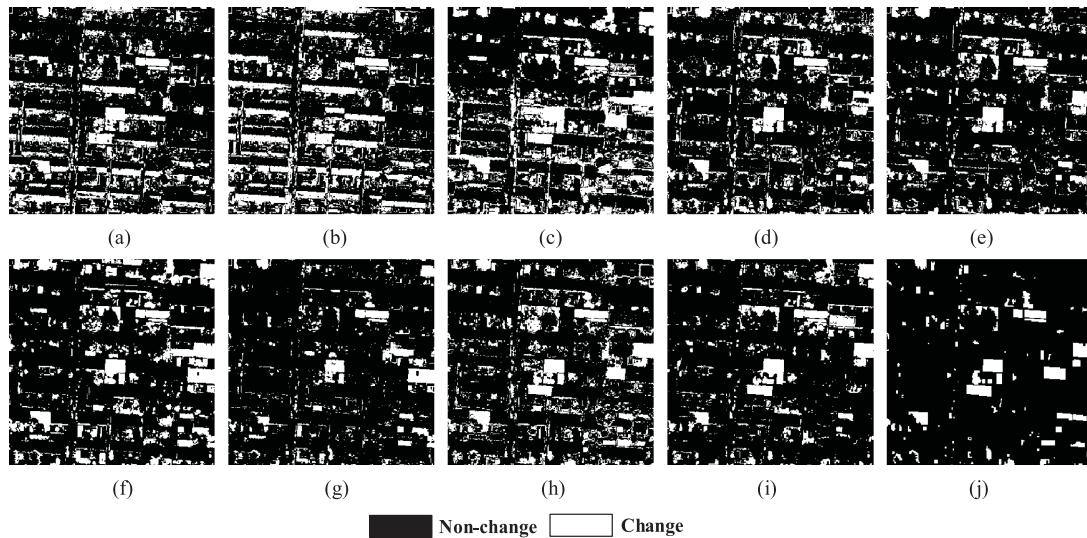


Fig. 14. Classification results of different change detection methods for the Xiong County dataset. (a) CVA. (b) AAD. (c) TDI-GC. (d) MAD. (e) IR-MAD. (f) ISFA. (g) DPCA. (h) SISFA. (i) MCS<sup>4</sup>CD. (j) Ground truth.

TABLE V  
CLASSIFICATION ACCURACY IN RELATION TO DIFFERENT CHANGE DETECTION METHODS FOR THE XIONG COUNTY DATASET

Method	CVA	AAD	TDI-GC	MAD	IR-MAD	ISFA	DPCA	SISFA	MCS <sup>4</sup> CD
OA (%)	79.13	72.88	78.89	84.68	87.70	82.61	88.73	86.70	<b>89.19</b>
Kappa	0.2480	0.1355	0.3070	0.3639	0.4212	0.3197	0.3528	0.4560	<b>0.4562</b>
MA	<b>0.0366</b>	0.0521	0.2157	0.1426	0.1078	0.1636	0.0470	0.1340	0.0907
FA	0.1967	0.2543	<b>0.0177</b>	0.0274	0.0284	0.0299	0.0772	0.0123	0.0288

are up to 88.73%. Semi-supervised MCS<sup>4</sup>CD algorithm achieves the best result (OA = 89.19%) visually, and unsupervised AAD algorithm obtains the worst result (OA = 72.88%) with many undetected regions.

Among these competing methods, the CD results show that there are many small and fragmented areas in the changed area, indicating that when the supervised and MCS<sup>4</sup>CD algorithm extracts the change information, some changed pixels will be misclassified as unchanged pixels, and then a large number of spots appear in the image. SSL algorithms, which consider both labeled (the number of initial training sample labels is 200) and unlabeled (unlabeled samples select 500 as the pseudo labeled set) data, can improve the experimental accuracy. However, in complex scenes, the improvement of accuracy is limited. The complexity of surface scene, the diversity of surface feature types, and imaging conditions significantly affect the accuracy of CD results. In the process of CD, there are a lot of false detection and missed detection. The results of traditional algorithms under complex backgrounds are easily affected by environmental background, resulting in a large amount of spots. Therefore, using traditional algorithms to detect changes in a complex background is very challenging. The machine learning algorithm represented by deep learning essentially adopts the way of supervised learning, which uses a large number of sample data to learn the essential characteristics of the target, and then predicts and discriminates the unknown data accordingly. It can extract high-level feature changes and suppress speckle noise.

However, the experimental effect of deep learning network depends on sufficient training samples. If the number of sample data can meet the requirements of deep learning, we can continue to study in the field of deep algorithms in the future.

## V. CONCLUSION

This study proposed a semi-supervised CD method with slow feature analysis, including multilayer cascade screening strategy and data transformation strategy in spectral subspace domain, for HSI classification with a small number of labeled samples. In the semi-supervised process, the slow feature extraction in a high-dimensional space, the use of BT algorithm, and the decision strategy for the label of unlabeled samples are all key points. On one hand, we use feature extraction method to reduce too many spectral bands, which has a negative impact on the expected CD performance. On the other hand, BT, circular neighbor, and SVM are combined together to improve the judgment accuracy of unlabeled samples. Experimental results with four classical HSI datasets indicate that the proposed MCS<sup>4</sup>CD approach can show well performance.

In the future, we will explore the CD method based on nonlinear transformation and spatial constraint method for HSI datasets with limited labeled samples.

## REFERENCES

- [1] Z. Zhu, "Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 370–384, 2017.
- [2] S. Liu *et al.*, "Unsupervised change detection in multispectral remote sensing images via spectral-spatial band expansion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3578–3587, Sep. 2019.
- [3] X. Zhang *et al.*, "Change detection of built-up land: A framework of combining pixel-based detection and object-based recognition," *ISPRS J. Photogrammetry Remote Sens.*, vol. 119, pp. 402–414, 2016.
- [4] D. 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. 4913–4928, Jul. 2019.
- [5] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.
- [6] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.
- [7] C. Wu, L. Zhang, and B. Du, "Hyperspectral anomaly change detection with slow feature analysis," *Neurocomputing*, vol. 151, pp. 175–187, 2015.
- [8] M. Hu, C. Wu, L. Zhang, and B. Du, "Hyperspectral anomaly change detection based on autoencoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3750–3762, Mar. 2021.
- [9] W. Shi *et al.*, "Change detection based on artificial intelligence: State-of-the-art and challenges," *Remote Sens.*, vol. 12, no. 10, 2020, Art. no. 1688.
- [10] C. E. Woodcock *et al.*, "Transitioning from change detection to monitoring with remote sensing: A paradigm shift," *Remote Sens. Environ.*, vol. 238, no. 18, 2019, Art. no. 111558.
- [11] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.
- [12] C. Zhang *et al.*, "A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 183–200, 2020.
- [13] H. C. Li *et al.*, "Deep nonsmooth nonnegative matrix factorization network with semi-supervised learning for SAR image change detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 160, pp. 167–179, 2020.
- [14] D. Hong *et al.*, "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 52–87, Jun. 2021.
- [15] L. T. Luppino, F. M. Bianchi, G. Moser, and S. N. Anfinsen, "Unsupervised image regression for heterogeneous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9960–9975, Dec. 2019.
- [16] M. Gong, Y. Yang, T. Zhan, X. Niu, and S. Li, "A generative discriminatory classified network for change detection in multispectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 1, pp. 321–333, Jan. 2019.
- [17] D. Hong *et al.*, "Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model," *ISPRS J. Photogramm. Remote Sens.*, vol. 178, pp. 68–80, 2021.
- [18] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Joint and progressive subspace analysis (JPSA) with spatial-spectral manifold alignment for semi-supervised hyperspectral dimensionality reduction," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3602–3615, Jul. 2021.
- [19] K. Tan *et al.*, "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.
- [20] C. Liu, J. Li, and L. He., "Superpixel-based semisupervised active learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 1, pp. 357–370, Jan. 2019.
- [21] R. Toutati, M. Mignotte, and M. Dahmane, "Anomaly feature learning for unsupervised change detection in heterogeneous images: A deep sparse residual model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 588–600, Jan. 2020.
- [22] R. D. Johnson and E. S. Kasischke, "Change vector analysis: A technique for the multispectral monitoring of land cover and condition," *Int. J. Remote Sens.*, vol. 19, no. 3, pp. 411–426, 1998.
- [23] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, 1998.
- [24] 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, Mar. 2007.
- [25] J. S. Deng *et al.*, "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.
- [26] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014.
- [27] H. Huang, G. Shi, H. He, Y. Duan, and F. Luo, "Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2604–2616, Jun. 2020.
- [28] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.
- [29] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data," *Inf. Fusion*, vol. 59, pp. 44–58, Jul. 2020.
- [30] X. Deng *et al.*, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Inf. Sci.*, vol. 340, pp. 250–261, 2016.
- [31] H. G. Lewis and M. Brown, "A generalized confusion matrix for assessing area estimates from remotely sensed data," *Int. J. Remote Sens.*, vol. 22, no. 16, pp. 3223–3235, Jul. 2001.
- [32] A. Luque *et al.*, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit.*, vol. 91, pp. 216–231, 2019.
- [33] B. P. Salmon, W. Kleynhans, C. P. Schwegmann, and J. C. Olivier, "Proper comparison among methods using a confusion matrix," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 3057–3060.
- [34] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.
- [35] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.
- [36] C. Kwan, "Methods and challenges using multispectral and hyperspectral images for practical change detection applications," *Information*, vol. 10, no. 11, 2019, Art. no. 353.
- [37] J. Wang *et al.*, "Unsupervised change detection between SAR images based on hypergraphs," *ISPRS J. Photogramm. Remote Sens.*, vol. 164, pp. 61–72, 2020.
- [38] Z. Hou, W. Li, L. Li, R. Tao, and Q. Du, "Hyperspectral change detection based on multiple morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5507312.
- [39] D. Hong *et al.*, "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615, doi: [10.1109/TGRS.2021.3130716](https://doi.org/10.1109/TGRS.2021.3130716).
- [40] Z. Hou, W. Li, R. Tao, and Q. Du, "Three-order Tucker decomposition and reconstruction detector for unsupervised hyperspectral change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6194–6205, Jun. 2021.



**Lian Liu** received the Ph.D. degree in information and communication engineering from the College of Information and Communications Engineering, Harbin Engineering University, Harbin, China, in 2019.

She is currently a Postdoctoral Associate with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. Her research interests include change detection, image processing, and machine learning.



**Danfeng Hong** (Senior Member, IEEE) received the M.Sc. degree (*summa cum laude*) in computer vision from the College of Information Engineering, Qingdao University, Qingdao, China, in 2015, the Dr.-Ing degree (*summa cum laude*) in signal processing in earth observation (SiPEO) from the Technical University of Munich, Munich, Germany, in 2019.

He is currently a Professor with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences (CAS). Before joining CAS,

he has been a Research Scientist and led a Spectral Vision Working Group at the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. He was also an Adjunct Scientist at GIPSA-lab, Grenoble INP, CNRS, Univ. Grenoble Alpes, Grenoble, France, from 2020 to 2022. His research interests include signal / image processing and analysis, hyperspectral remote sensing, machine / deep learning, artificial intelligence, and their applications in Earth Vision.

Dr. Hong was a recipient of the Best Reviewer Award of the IEEE TGRS in 2021 and the Jose Bioucas Dias award for recognizing the outstanding paper at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) in 2021. He is an Editorial Board Member of Remote Sensing and a Topical Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is also a Leading Guest Editor of the *International Journal of Applied Earth Observation and Geoinformation*, *IEEE Journal of Selected Topics in Applied Earth Observations*, and *Remote Sensing*.



**Li Ni** received the Ph.D. degree in cartography and geographical information system from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2015.

She is currently an Associate Researcher with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences. Her research interests include the land surface temperature retrieval and application of hyperspectral remote sensing.



**Lianru Gao** (Senior Member, IEEE) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2002, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, China, in 2007.

He is currently a Professor with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, CAS. He also has been a Visiting Scholar with the University of Extremadura, Cáceres, Spain, in 2014, and at the Mississippi State University (MSU), Starkville, USA, in 2016. His research focuses on hyperspectral image processing and information extraction. In the last ten years, he was the PI of ten scientific research projects at national and ministerial levels, including projects by the National Natural Science Foundation of China (2016–2019, 2018–2020, 2022–2025), and by the Key Research Program of the CAS (2013–2015). He has published more than 180 peer-reviewed papers, and there are more than 100 journal papers included by Science Citation Index (SCI). He has coauthored three academic books including “*Hyperspectral Image Information Extraction*.”

Dr. Gao was the recipient of the Outstanding Science and Technology Achievement Prize of the CAS in 2016, and was supported by the China National Science Fund for Excellent Young Scholars in 2017, and won the Second Prize of The State Scientific and Technological Progress Award in 2018, he received the recognition of the best reviewers of the IEEE JSTARS in 2015, and the best reviewers of the IEEE TGRS in 2017. He obtained 29 National Invention Patents in China.