Sketched Multiview Subspace Learning for Hyperspectral Anomalous Change Detection

Shizhen Chang, Member, IEEE, Michael Kopp, Member, IEEE, and Pedram Ghamisi, Senior Member, IEEE

Abstract—In recent years, multiview subspace learning has been garnering increasing attention. It aims to capture the inner relationships of the data that are collected from multiple sources by learning a unified representation. In this way, comprehensive information from multiple views is shared and preserved for the generalization processes. As a special branch of temporal series hyperspectral image (HSI) processing, the anomalous change detection (ACD) task focuses on detecting very small changes among different temporal images. However, when the volume of datasets is very large or the classes are relatively comprehensive, the existing methods may fail to find those changes between the scenes, and end up with terrible detection results. In this article, inspired by the sketched representation and multiview subspace learning, a sketched multiview subspace learning (SMSL) model is proposed for HSI ACD. The proposed model preserves major information from the image pairs and improves the computational complexity using a sketched representation matrix. Furthermore, the differences between scenes are extracted using the specific regularizer of the self-representation matrices. To evaluate the detection effectiveness of the proposed SMSL model, experiments are conducted on a benchmark hyperspectral remote sensing dataset and a natural hyperspectral dataset and compared with other state-of-the-art approaches. The code of the proposed method will be available at https://github.com/ShizhenChang/SMSL.

Index Terms—Anomalous change detection (ACD), hyperspectral image (HSI) processing, multiview subspace learning, remote sensing, sketched self-representation, temporal analysis.

I. INTRODUCTION

As one of the typical research fields of image processing, change detection focuses on measuring the progression of one or more patterns that perform substantially differently among multitemporal images of the same scene [1]. In response to the demands of diverse disciplines, change detection has been widely applied for remote sensing [2], [3], [4], video surveillance [5], [6], medical diagnosis and treatment [7], [8], civil infrastructure [9], [10], and underwater sensing [11], [12]. With the development of hyperspectral imaging technology [13], [14], using the wealth of spectral information together with spatial correlations between the objects to capture changes in the multitemporal hyperspectral images (HSIs) has attracted increasing attention [15].

Generally speaking, the change detection task needs to squeeze the input image pairs into one output image that can reflect the potentially changed object with higher values. Our challenge is to distinguish real changes from the interference caused by sensor noise, camera motion, illumination variation, shadows, or atmospheric absorption [16]. Based on the interests of particular application, change detection can be categorized into two main topics, general change detection [17], [18] and anomalous change detection (ACD) [19], [20]. In this article, we focus on exploring the potential of anomalous changes in HSIs. What we are interested in are those relatively small and unusual changes within the broad changed or unchanged backgrounds, such as small objects that suddenly appear, disappear, or move.

To detect those small objects, the basic idea of the ACD algorithms is to define a predictor to minimize the spectral difference in the backgrounds while highlighting the changes in very high-dimensional multitemporal images [21]. A simplified diagram of ACD is shown in Fig. 1. In recent years, a number of ACD algorithms have been proposed in the literature to cope with this problem. One effective predictor is the chronochrome (CC) [22], [23] method, which uses the joint second-order statistic between two images to capture small changes. The spectral differences in backgrounds are modeled by the least-square linear regression. Another widely used method, covariance equalization (CE) [23], [24], assumes that the images have the same statistical distribution after the whitening/dewhiten transform. Based on the nonlinear Gaussian distribution, the cluster kernel Reed–Xiaoli (CKRX) algorithm [19] was proposed and applied for change detection. The CKRX method groups background pixels into clusters and

Manuscript received 13 July 2022; revised 9 October 2022; accepted 4 November 2022. Date of publication 9 November 2022; date of current version 21 November 2022. (Corresponding author: Shizhen Chang.)

Shizhen Chang and Michael Kopp are with the Institute of Advanced Research in Artificial Intelligence, Landstraßer Hauptstraße 5, 1030 Vienna, Austria (e-mail: shizhen.chang@iarai.ac.at; michael.kopp@iarai.ac.at).

Pedram Ghamisi is with the Institute of Advanced Research in Artificial Intelligence, Landstraßer Hauptstraße 5, 1030 Vienna, Austria, and also with the Machine Learning Group, Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute for Resource Technology (HIF), 09599 Dresden, Germany (e-mail: pedram.ghamisi@iarai.ac.at).

Digital Object Identifier 10.1109/TGRS.2022.3220814
then applies a fast eigendecomposition algorithm to generate the anomaly index. Focusing on the Gaussian and elliptically contoured (EC) distribution, the kernel ACD algorithm was proposed [25], which extends the detection process to the nonlinear counterparts based on the theory of reproducing kernels’ Hilbert space. The traditional models assume the distribution of land covers obeys a statistical model, and then deduce the detection output through statistical assumptions.

With the development of the signal processing and machine learning theories, more detectors have been proposed. Wu et al. proposed a slow feature analysis (SFA)-based method to explore potential small changes by assuming that the background signals are invariant slow varying features [26]. The top bands of the residual images that are highly related to changes are finally selected as the input for Reed–Xiaoli (RX) anomaly detection [27]. Based on the sparse representation theory, a joint sparse representation (JSR) ACD method [20] was proposed. For every test pixel, the surrounding pixels in its dual window are sparsely represented by a randomly selected stacked background dictionary matrix, and the detection output is determined by the active bases of the dictionary. To better capture the nonlinear features from bitemporal images, an autoencoder-based two-Siamese network was proposed [28], which uses bidirectional predictors to minimize the reconstruction errors between the image pairs. To fully explore the feature correlations of the images, a self-supervised hyperspectral spatial–spectral feature understanding network (HyperNet) was proposed [29], which achieves pixel-level features for change detection. Unlike the traditional predictors, the machine/deep-learning-based models focus on capturing the feature differences between backgrounds and anomalous changes, to extract the changed objects from multitemporal images. However, those slight differences among the temporal images are still very difficult to be perfectly represented.

Nowadays, multiview data analysis has gained increasing attention in many real-world applications, since data are usually collected from diverse domains or obtained from various feature subsets [30], [31]. To better combine the consensus and complementary information among multiple views, the model is designed to give a comprehensive understanding and improve the generalization performance [32]. Among all the topics related to multiview learning, subspace learning is one of the most typical, which aims to obtain a latent subspace to align features for the inputs [33], [34]. The representative methods are successfully used for classification [35] and clustering [36], [37]. However, for large-volume data, the huge complexity and memory consumption of the multiview learning methods will cause a serious computational burden. On the other hand, the traditional multiview learning methods pay more attention to homogeneous information of each view and, thus, will ignore the correlation among the views.

For the proposed hyperspectral ACD task, instead of capturing the most common information among the views, we are interested in extracting small abnormal objects and excluding them from the background instances. To address this issue, a sketched multiview subspace learning (SMSL) model with a consistent constraint and a specific constraint is proposed in this article. The flowchart of the proposed SMSL model is shown in Fig. 2, which illustrates its three main steps. First, considering the large volume of the HSIs, a sketched dictionary is calculated from the union matrix of all the images. Then the residual fractions between the neighboring views corresponding to the specific matrix and the noisy matrix are obtained. Finally, the ACD map is derived by calculating the norm of the residuals. Our main contributions are summarized as follows.

1) It is the first attempt to combine multiview subspace learning with change detection to distinguish small anomalous changes from temporal images. Considering the large volume of the hyperspectral datasets and high computational consumption of the optimization process, a sketched dictionary is used to preserve most of the information from the original data with much smaller sample size.
2) With a low-rank regularizer that constrains the consistent part of the coefficient matrix and two regularizers that constrain the specific part, the SMSL model guarantees that the common information of the data are the lowest rank represented, while the differences are maximally separated.
3) Experiments conducted on sufficient HSI datasets demonstrate the effectiveness of the proposed method.
Detailed analysis concludes the convergence of the model and the performance related to the parameters.

II. RELATED WORKS

A. Subspace Learning

Subspace learning mainly focuses on recovering the subspace structure of the data. Recently, the self-representative methods, which assume that data instances can be approximately formed by a combination of other instances of the data, have been used for the clustering and classification tasks. The representation models are generated via structured convex regularizers. For example, the sparse subspace clustering (SSC) [38] method calculates data clusters in a low-dimensional subspace using the $\ell_1$-norm. For a given data $X \in \mathbb{R}^{d \times N}$, where $d$ represents the dimension and $N$ represents the total number of samples of the data, the objective function of SSC is written as

$$\min_{Z} ||Z||_1, \quad \text{s.t.} \ X = XZ, \ \text{diag}(Z) = 0$$

(1)

where $Z \in \mathbb{R}^{N \times N}$ is the coefficient matrix.

Based on the low-rank representation (LRR) model, the objective was proposed to solve the following problem:

$$\min_{Z,E} ||Z||_*, + \lambda ||E||_{2,1}, \quad \text{s.t.} \ X = XZ + E$$

(2)

where the coefficient matrix $Z$ is low-ranked in this case, and $E \in \mathbb{R}^{d \times N}$ is the error matrix corresponds to the sample-specific corruptions.

B. Multiview Subspace Learning

Given that real-world data are usually collected from multiple sources or represent different feature types, the multiview subspace learning methods are proposed to learn a common subspace of different views. For instance, Guo [40] proposed a convex subspace learning model which jointly solves the optimization problem and learns the common subspace using a sparsity inducing norm. Ding and Fu [41] proposed a robust multiview subspace learning algorithm that uses dual low-rank decomposition and two supervised graph regularizers to obtain the view-invariant subspace.

Specifically, by jointly exploring the consistency and specificity for subspace representation learning, Luo et al. [42] design a novel multiview self-representation model for clustering.

Let $X^i \in \mathbb{R}^{d_i \times N}$ be the $i$th view of all the data, where $d_i$ denotes the dimension of $X$; the multiview self-representation model can be formulated as

$$X^i = X^iZ^i + E^i$$

(3)

where $Z^i$ and $E^i$ correspond to the coefficient matrix and the error matrix of $X^i$, respectively.

By arguing that the coefficient matrix of different views contains a consistency term $C$ and a view-specific term $D^i$, (3) can be rewritten as

$$X^i = X^i(C + D^i) + E^i.$$

(4)

To better excavate the common information and ensure the unification property among all the views, the nuclear norm is imposed to constrain the consistent matrix and the $\ell_2$-norm is chosen for the specific matrix. Thus, the model is optimized via the augmented Lagrange multiplier (ALM) and the finally clustering result is derived by spectral clustering.

III. METHODOLOGY

The idea of multiview subspace learning leads to new perspectives for multitemporal data analysis. However, the optimization process of the multiview subspace learning models includes inverse derivations of the data itself, so the good performance is usually compromised by a very high price of computational complexity. More importantly, directly applying the multiview subspace learning model to detect anomalous changes cannot fully consider the abnormal information among the views. In this article, to better extract specific fractions and effectively solve the subspace learning model, an SMSL model is proposed for hyperspectral ACD.

A. Sketched Dictionary

As has been introduced in Section II, the existing subspace learning models use data itself as the dictionary to find the subspaces, but the large consumption of both storage and processing time makes it hard to handle large-scale datasets.

For the proposed hyperspectral ACD problem, which has abundant spectral information and relatively large data volume, the existing multiview-based subspace learning models are infeasible for deriving a detection result in a computationally efficient manner. Thus, to compress images into a scalable size and preserve main information, the sketched dictionary of the inputs is designed.

Assume the hyperspectral dataset $\{X^s \in \mathbb{R}^{L \times N}\}$ has $s$ phases in total, where $s \in \{1, 2, \ldots, S\}$, and $L$ and $N$ are the dimension and the number of pixels of the $s$th phase, respectively. To preserve most of the common information and enlarge the differences in the coefficients among the views, we define a sketching matrix $R \in \mathbb{R}^{N \times N_H}$ to obtain the sketched dictionary $H \in \mathbb{R}^{L \times N_H}$

$$H = [X^1, X^2, \ldots, X^S]R.$$

(5)

The sketching matrix is intended to compress data while retaining as much information as possible. It has been proven that the sketched representation can largely reduce the computational complexity, meanwhile preserving the data information [43], [44]. In particular, defining a random matrix using Johnson–Lindenstrauss transforms (JLT) can hold this property. In this article, the random JLT matrix is independent and identically distributed (i.i.d.) with the values of the entries drawn from $\mathcal{N}(0, 1)$ normal distribution scaled by the factor $1/\sqrt{N_H}$.

B. Sketched Multiview Subspace Learning

After obtaining the sketched dictionary $H$, the SMSL model is proposed for ACD. For each phase of the data $X^i$, assuming it can be approximated formulated by the matrix multiplication of the sketched dictionary and the corresponding coefficient
matrix and a noisy matrix, the expressive function can be derived as follows:

\[ X' = HZ' + E' \]  

(6)

where \( Z' \) and \( E' \) denote the coefficient matrix and noisy matrix corresponding to \( X' \), respectively. Then, combining this with the consistency and specificity property of the coefficient matrix, (6) can be modeled as

\[ X' = H(C + D') + E'. \]  

(7)

Fig. 3 displays the SMSL model with consistency and specificity where the common information shared through the dictionary, and the differences in the specific parts among the views are large. Generally, the objective function of the SMSL model is expressed as:

\[
\min_{E', C, D'} \sum_s L'(X', H(C + D')) + \lambda_1 \Omega_C
\]

\[
+ \lambda_2 \sum_s \Omega_{D'} + \lambda_3 \sum_s F_{s,t \neq s} (D', D')
\]

where \( \lambda_1, \lambda_2, \lambda_3 > 0 \) are the three tradeoff parameters to balance the four terms, \( \sum_s L'(X', H(C + D')) \) is the total loss of subspace representation, and the reconstructed coefficient matrices \( C \) and \( D' \) are regularized with \( \Omega_C \) and \( \Omega_{D'} \). \( F_{s,t \neq s} (D', D') \) measures the difference between \( D' \) and \( D' \). In our work, we want the difference between \( D' \) and \( D' \) as large as possible, so the relaxed exclusivity is used to verify the similarity of the matrices.

Definition 1 (Relaxed Exclusivity [45], [46]): The definition of relaxed exclusivity between \( U \in \mathbb{R}^{m \times n} \) and \( V \in \mathbb{R}^{m \times n} \) is \( F(U, V) = ||U \odot V||_1 = \sum_{i,j} |u_{ij} \cdot v_{ij}| \), where \( || \cdot ||_1 \) is the \( \ell_1 \) norm, \( \cdot \odot \) represents the absolute value, and \( \odot \) designates the Hadamard product which operates an elementwise multiplication of the two matrices.

It is noted that minimizing the relaxed exclusivity term can guarantee that the two matrices are as orthogonal as possible. More specifically, the performance of detecting anomalies is strongly related to the comprehensive understanding of the common part and the identification of the differences. Thus, we design the proposed SMSL model with multiple regularization terms as

\[
\begin{align*}
\min_{E', C, D'} & \sum_{s=1}^S ||E'||_{2,1} + \lambda_1 ||C||_s + \frac{\lambda_2}{2} \sum_{s=1}^S ||D'||_F^2 \\
& + \lambda_3 \sum_{s=1}^S ||D' \odot D'||_1
\end{align*}
\]

s.t. \( X' = H(C + D') + E' \), \( (C + D')^T 1 = 1 \)  

(8)

where \( || \cdot ||_{2,1} \) is a \( \ell_{2,1} \)-norm that ensures that the columns of the matrix are sparse, \( || \cdot ||_s \) is the nuclear norm that ensures that the matrix is low-rank, and \( || \cdot ||_F \) is the Frobenius norm. By adding a sum-to-one constraint to the columns of the coefficient matrix, the images are assumed to be absolutely represented by the representative model and the consistent part is more robust to anomalies.

C. Optimization

According to the objective function of our SMSL model in (8), we can simultaneously obtain the subspace representation of multiviews and optimize the consistent matrix. To pursue the optimal solutions of all the variables, the proposed model is divided into several subproblems, and the ALM algorithm is used.

By introducing two auxiliary variables \( W' \) and \( J \) to replace \( E' \) and \( C \), respectively, our model can be equivalently rewritten as

\[
\begin{align*}
\min_{E', W', C, D'} & \sum_{s=1}^S ||W'||_{2,1} + \lambda_1 ||J||_s + \frac{\lambda_2}{2} \sum_{s=1}^S ||D'||_F^2 \\
& + \lambda_3 \sum_{s=1}^S \sum_{t \neq s} ||D' \odot D'||_1
\end{align*}
\]

s.t. \( X' = H(C + D') + E' \), \( E' = W' \)

(9)

\( (C + D')^T 1 = 1 \, \, C = J \).

Then, the augmented Lagrange function is formulated as

\[
\begin{align*}
L(C, J, D', E', W', Y_1, Y_2, Y_3, Y_4) & = \sum_{s=1}^S ||W'||_{2,1} \\
& + \lambda_1 ||J||_s + \frac{\lambda_2}{2} \sum_{s=1}^S ||D'||_F^2 + \lambda_3 \sum_{s=1}^S \sum_{t \neq s} ||D' \odot D'||_1 \\
& + \frac{\mu}{2} \sum_{s=1}^S \left( ||X' - H(C + D') - E' + \frac{Y_1}{\mu}||_F^2 \\
& + ||(C + D')^T - 1 + \frac{Y_2}{\mu}||_F^2 \\
& + ||E' - W' + \frac{Y_3}{\mu}||_F^2 + \frac{\mu}{2} ||C - J + \frac{Y_4}{\mu}||_F^2 \right)
\end{align*}
\]

(10)

where \( \{Y_1, Y_2, Y_3\}_{k \in [3]} \) and \( Y_4 \) are the Lagrange multipliers, and \( \mu > 0 \) is a penalty parameter. To optimize the above

\(^1\)For a matrix \( A \in \mathbb{R}^{m \times n} \), the definition of its \( \ell_{2,1} \)-norm is: \( ||A||_{2,1} = \sum_{j=1}^n (\sum_{i=1}^m a_{ij}^2)^{1/2} \).
unconstrained function, we divide it into six subproblems and optimize each of them with other variables fixed, alternatively. The optimization process can be organized as follows:

1) **C-Subproblem**: Fixing the other variables, the variable $C$ is optimized by the following subproblem:

$$C^* = \text{arg min}_C \frac{\mu}{2} \sum_{s=1}^S \left\| X^s - H(C + D^s) - E^s + \frac{Y_1^s}{\mu} \right\|^2_F$$

$$+ \frac{\mu}{2} \sum_{s=1}^S \left( \frac{1}{\mu} \sum_{t=1}^{m} S_{1t} \right)$$

By differentiating the objective function with respect to $C$ and setting the derivation to zero, the update rule of $C^*$ is obtained:

$$C^* = A^{-1} B$$

where $A = S H^T H + S I_1^T + I$

and

$$B = \sum_{s=1}^S H^T \left( X^s - HD^s - E^s + \frac{Y_1^s}{\mu} \right)$$

$$- \sum_{s=1}^S I^T \left( \frac{1}{\mu} \sum_{t=1}^{m} S_{1t} \right) + J - \frac{Y_4}{\mu}.$$ 

2) **J-Subproblem**: With fixed variables, $J$ can be optimized by the following problem:

$$J^* = \text{arg min}_J \lambda_1 \|J\|_1 + \frac{\mu}{2} \left\| C - J + \frac{Y_4}{\mu} \right\|^2_F.$$ 

Note that the above function is equivalent to

$$J^* = \text{arg min}_J \lambda_1 \|J\|_1 + \frac{1}{2} \left\| C - J + \frac{Y_4}{\mu} \right\|^2_F.$$ 

According to [47], the approximation of the nuclear norm can be solved with the singular value thresholding (SVT) algorithm; the update rule is thus written as:

$$J^* = UD_{\frac{\lambda_1}{\mu}}(\Sigma)V^T$$

where $\Sigma = C + (Y_4/\mu)$, and the soft-thresholding operator $D_\tau(\cdot)$ is defined as:

$$D_{\tau}(\cdot) = \max(\cdot - \tau, 0) + \min(\cdot + \tau, 0).$$

3) **D^s-Subproblem**: We can find that the variables $\{D^s\}_{s \in [S]}$ are independent of each other, so for a fixed phase $s$, we can solve the following function with other variables fixed:

$$D^s = \text{arg min}_{D^s} \frac{\lambda_2}{2} \left\| D^s \right\|_F^2 + \lambda_3 \sum_{t \neq s} \left\| D^t \odot D^s \right\|_1$$

$$+ \frac{\mu}{2} \left\| X^s - H(C + D^s) - E^s + \frac{Y_1^s}{\mu} \right\|_F^2$$

$$+ \frac{\mu}{2} \left\| 1^T(C + D^s) - 1^T + \frac{Y_2^s}{\mu} \right\|_F^2.$$ 

To calculate the optimal solution of $D^s$, the key point is to obtain the partial derivation with respect to $D^s$ for the $\ell_1$ norm of the Hadamard product function.

**Lemma 1**: For finite dimensional matrices $U, V \in \mathbb{R}^{m \times n}$, the partial derivative of the function $F(U, V) = ||U \odot V||_1$ with respect to $V$ is

$$\frac{\partial F}{\partial V} = |U| \odot \text{sign}(V)$$

where $\text{sign}(\cdot)$ is the componentwise sign function.

**Proof**: The partial derivation with respect to each element $v_{ij}$ of $V$ can be written as:

$$\frac{\partial}{\partial v_{ij}} \left( \sum_{i,j} |U \odot V|_{i,j} \right)$$

$$= \frac{\partial}{\partial v_{ij}} \left( \sum_{i,j} |u_{ij}| |v_{ij}| \right)$$

$$= \sum_{i,j} |u_{ij}| \frac{\partial |v_{ij}|}{\partial v_{ij}} = \sum_{i,j} |u_{ij}| \delta_{ik} \delta_{lj} \text{sign}(v_{kl})$$

$$= |u_{ij}| \text{sign}(v_{ij})$$

where the operator $\delta_{ik} = 1$ when $k = i$; else, $\delta_{ik} = 0$. So Lemma 1 holds.

For the current view $s$, assuming that the elements of $D^s$ are all positive and following Lemma 1, the closed-form solution of $D^s$ can be written as:

$$D^s = (\lambda_2 I + \mu H^T H + \mu I_1^T)^{-1} \times$$

$$\left( -\lambda_3 \sum_{t \neq s} |D^t| + \mu H^T \left( X^s - HC - E^s + \frac{Y_1^s}{\mu} \right) \right)$$

$$- \mu I \left( 1^T C - 1^T + \frac{Y_2^s}{\mu} \right).$$

where $D^s$ can be conceptualized as the constant matrices of the current view. The update rule of $D^{s*}$ is

$$D^{s*} = \max\{D^s, 0\}.$$ 

4) **E^s-Subproblem**: By fixing other variables, the update rule of $E^s$ is:

$$E^{s*} = \text{arg min}_{E^s} \frac{\mu}{2} \left\| X^s - H(C + D^s) - E^s + \frac{Y_1^s}{\mu} \right\|_F^2$$

$$+ \frac{\mu}{2} \left\| E^s - W^s + \frac{Y_3^s}{\mu} \right\|_F^2.$$ 

Then the optimal value can be obtained by taking a partial derivation with respect to $E^s$ and setting it to zero:

$$E^{s*} = \frac{1}{2} \left( X^s - H(C + D^s) + \frac{Y_1^s}{\mu} + W^s - \frac{Y_3^s}{\mu} \right).$$
D1F12H1_D1F12H2. (d) Changemask of "D1F12H1_D2F22H2".

12: 11: end while
10: break;

Algorithm 1 SMSL-ACD: SMSL for Hyperspectral ACD

Input: The multitemporal dataset \( \{X_t\}_{t \in [S]} \), the sketched dictionary \( H \), and the parameters \( \lambda_1, \lambda_2, \text{and} \lambda_3. \)

1: Initialize coefficient matrices \( C = J = D = Y_4 = 0 \), \( E^t = W^t = Y_1^t = Y_2^t = 0 \), set parameters \( \mu = 10^{-5} \), \( \mu_{\text{max}} = 10^5 \), \( \rho = 1.1 \), maximum iteration times \( \text{iter} = 60 \), and the stopping threshold \( \epsilon = 10^{-5}. \)

2: while iterations < \( \text{iter} \) do
3:   Update \( C \) and \( J \) according to subproblem 1)–2);
4:   for \( s \in [1, 2, \ldots, S] \) do
5:       Update \( D^s, E^s, \text{and} W^s \) according to subproblem 3)–5);
6:       Update the multipliers \( Y_1^s, Y_2^s, \text{and} Y_3^s \) according to subproblem 6);
7:   end for
8:   Update the multipliers \( Y_4 \) and \( \mu \) according to subproblem 6);
9:   if converges then
10:      break;
11: end while
12: return \( D^t \) and \( E^t. \)

Output: The specific matrices \( H \cdot D^t, \text{noisy matrices} E^t \)

5) \( W^t \)-Subproblem: With other variables being fixed, the subproblem of updating \( W^t \) is:

\[
W^{t*} = \arg \min_{W^t} ||W^t||_{2,1} + \frac{\mu}{2} \left\| E^t - W^t + \frac{Y_1^t}{\mu} \right\|_F^2.
\]

Following Lemma 4.1 in [39], the closed-form solution of the above function is:

\[
[W^{t*}]_{i,i} = \begin{cases} 
\frac{||Q_{:,i}||_2 - \frac{1}{\mu} ||Q^*_{:,i}||_2}{||Q^*_{:,i}||_2}, & \text{if} \ ||Q_{:,i}||_2 > \frac{1}{\mu} \\
0, & \text{otherwise}
\end{cases}
\]

where \( Q = E^t + (Y_1^t/\mu) \), and \( Q_{:,i} \) denotes its \( i \)-th column.

6) Updating the Multipliers and \( \mu \):

\[
Y_1^t = Y_1^t + \mu(X^t - H(C + D^t) - E^t)
\]

\[
Y_2^t = Y_2^t + \mu(I^\top(C + D^t) - 1)\bigg)\]

\[
Y_3^t = Y_3^t + \mu(E^t - W^t)
\]

\[
Y_4 = Y_4 + \mu(C - J)
\]

\[
\mu = \min(\rho \mu, \mu_{\text{max}}).
\]

The complete steps of the proposed SMSL model are shown in Algorithm 1, where the convergence conditions are:

\[
||X^t - H(C + D^t) - E^t||_\infty < \epsilon
\]

\[
||E^t - W^t||_\infty < \epsilon
\]

\[
||C + D^t||^\top I - 1||_\infty < \epsilon, \text{and} ||C - J||_\infty < \epsilon.
\]

D. Decision Rule

Through the aforementioned optimization process, the specific part of each pixel is derived and can be written as

\[ S_{xi}^s = H d_i^s \]  \hspace{1cm} (11)

where \( x_i^s \) is the \( i \)-th pixel of \( X_s^t \), and \( d_i \) denotes the coefficient vector of the specific part. For bitemporal images, the output of abnormal pixels is also influenced by the differences and the noise. Thus, we define the residual corresponding to \( x_i \) as the ACD result

\[ r_i^s = ||S_{xi}^s - S_x^t||_2 + ||e_i^t - e_i^s||_2 \]  \hspace{1cm} (12)

where \( e_i^t \) denotes the \( i \)-th column of the noisy matrix \( E^t \).

More generally, the decision rule can be written as the sum of the residuals

\[ d(x_i) = \sum_{s} r_i^s. \]  \hspace{1cm} (13)
E. Complexity and Convergence

1) Complexity Analysis: The proposed model totally contains \(3S + 2\) optimization process, and the complexity is analyzed as follows. Considering that the size of the sketched dictionary is much smaller than the original inputs, the complexity of updating the coefficient matrix \(C\), \(D_s\), and the auxiliary variable \(J\) is simplified as \(O(N_{hi}^2N)\). The complexity of updating variable \(E^1\) and multiplier \(Y^1_1\) is \(O(LN_{hi}N)\) due to the matrix multiplication. The complexity of the multiplier \(Y^1_2\) is \(O(N_{hi}N)\). For the subproblem \(W^1\), the complexity is \(O(LN)\). Then, the overall complexity of the SMSL model is \(O(SN_{hi}^2N + SLN_{hi}N + SN_{hi}N + SLN)\), which is basically \(O(SN_{hi}^2N + SLN_{hi}N)\). Since the size of the sketched dictionary and the image bands are much smaller than the image size, i.e., \(N_{hi} \ll N\) and \(L \ll N\), in contrast to the traditional subspace learning methods [48], [49], the computational consumption of our model is obviously decreased and the limitations of multiview learning are greatly addressed.

2) Convergence Analysis: Unfortunately, the convergence of the ALM method with three or more pending matrices is very difficult to prove theoretically. Inspired by previous research [48], [49], [50], the gap generated in each iteration is calculated and shown in the experimental results. We find that the SMSL model can be expected to have good convergence properties.

IV. EXPERIMENTS

In this section, experiments are conducted on a natural HSI dataset and two hyperspectral remote sensing image pairs for bitemporal ACD. Accordingly, the results are shown with the
Fig. 9. ROC curves of all the methods in (a) Object_1550_1558, (b) D1F12H1_D1F12H2, and (c) D1F12H1_D2F22H2.

Fig. 10. Spectral curves of the randomly selected pixels in three datasets. (a) and (b) Object_1550_1558. (c) and (d) D1F12H1_D1F12H2. (e) and (f) D1F12H1_D2F22H2.
corresponding analysis, and the proposed method is compared with five state-of-the-art anomalous change detectors: the difference RX [19], CC [22], CE [24], SFA [26], and JSR [20]. The convergence and the parametric sensitivity are tested as well. All the experiments are conducted in MATLAB on an Intel Core i7-8550U CPU with 16 GB of RAM.

A. HSI Datasets

1) Object_1550_1558 Dataset: This dataset was originally derived from the BGU ICVL HSI dataset.² A Specim PS Kappa D × 4 hyperspectral camera and a rotary stage for spatial scanning are used to acquire the data [51]. We selected one image pair from the database and cropped a subimage for ACD. Each subimage contains 600 × 900 spatial resolution over 31 spectral bands from 400 to 700 nm at 10-nm increments. The change mask was created in ENVI image analysis and processing software by combining the human observation and the predetection result of the differential image. Fig. 4(a) and (b) are the RGB-colored images of the Object_1550 and the Object_1558, respectively, and Fig. 4(c) is the change mask of the changed objects.

2) Viareggio Datasets: The Viareggio dataset [52] provided by the Viareggio 2013 Trial contains three hyperspectral remote sensing images of the same study area. Both D1F12H1 and D1F12H2 were acquired on May 8, 2013, where the illumination conditions are very similar. D2F22H2 was acquired the following day, on May 9, 2013, when the illumination conditions were quite different. For ACD, this dataset is usually divided into two pairs: “D1F12H1_D1F12H2” and “D1F12H1_D2F22H2.” Each image contains 375 × 450 pixels with 128 spectral bands. The ground-truth mask for pairwise change detection is provided by the trial. Fig. 5 illustrates the RGB-colored images and the change masks of two pairs.

B. Experimental Results

1) Performance Comparison: Since the sketched dictionary is constructed by the JLT transformation, the result could be influenced by the random matrix. Taking into consideration the computation consumption of the multiview subspace learning model, we repeat our construction process ten times, and the result is derived using the average sketched dictionary. To quantitatively evaluate the performances of different methods, the receiver operating characteristic curve (ROC) and the area under the ROC curve (AUC) values are used for illustration. For a fair comparison, we tuned the parameters of the compared methods and reported the best results.

The visualized results of three datasets are shown in Fig. 6–8, and the ROC curves and the corresponding AUC values of all the algorithms are shown in Fig. 9. It can be seen that for the natural hyperspectral dataset, the mirror and the palette located on the table are quite different from the surroundings, but they are not changed in different scenes. The compared methods failed to suppress the mirror and the palette and have relatively small outputs. For the “D1F12H1_D1F12H2” and “D1F12H1_D2F22H2” image pairs, the background materials are quite complicated, while the anomalous pixels are rare and small, which means highlighting the anomalous changes from the background should be more difficult. It can be seen from Figs. 7 and 8 that the difference RX, CC, CE, and SFA methods failed to suppress the background pixels and result in bad visualization detection maps. Compared with other methods, the JSR and SMSL methods are more robust to the backgrounds. Combining the quantitative evaluations shown in Fig. 9, we can clearly see that the ROC curves of the proposed SMSL are more close to the top left corner and reach the largest AUC values in all the three datasets.

To better illustrate the effectiveness of the proposed model, we randomly select five pixels that belong to the anomalous changes and the backgrounds from the images and make a comparison of their spectral characteristics for the three datasets. The spectral difference between the original inputs \(X^1\) and \(X^2\), the specific matrix \(H \cdot D^1\) and \(H \cdot D^2\), and the noisy matrix \(E^1\) and \(E^2\) are given in Fig. 10. We can see that overall the anomalous changes have relatively larger spectral differences between the original inputs than the backgrounds. For the “Object_1550_1558” and the “D1F12H1_D2F22H2” datasets, the differences between the two times are well-preserved on anomalous pixels; meanwhile, the differences between background pixels are also well-suppressed. For the “D1F12H1_D1F12H2” dataset, the SMSL model can enlarge the differences between anomalous pixels and decrease the differences between background pixels.

2) Parametric Analysis: As has been discussed before, the sketched dictionary we designed aims to preserve the most important information while effectively saving computational consumption and storage. However, it is clear that with the reduction in the size of the sketching matrix, the consumption of the optimization process is saved; meanwhile, we may lose more information that is originally contained in the inputs. Therefore, we adjust the size of the sketch dictionary from 100 to 500 and test the performances of SMSL in three datasets. As shown in Fig. 11, the AUC values of the proposed SMSL model are generally improved as the size of the sketched dictionary increases. To find a compromise between preserving the majority information and effectively computation, the size of \(H\) is set to 500.

²http://icvl.cs.bgu.ac.il/hyperspectral/
It is known from (8) that three tradeoff parameters are used to balance the constraints of coefficient matrix $C$ and $D^s$ for the optimization procedure, where $\lambda_1$ reflects the influence of the consistency term $||C||_*$, $\lambda_2$ relates to the quantity of the specific matrices, and $\lambda_3$ reflects the orthogonality between the specific matrices. Considering that $\lambda_2$ and $\lambda_3$ are two penalty scalars on $D^s$, we jointly analyze the influence of $\lambda_2$ and $\lambda_3$ by setting their values as $\{0.1, 1, 10, 100, 1000\}$. The AUC values of the proposed SMSL method are illustrated by 3-D colored surfaces, as shown in Fig. 12, where $\lambda_1$ is set to specific values.

It can be seen that $\lambda_2$ and $\lambda_3$ are negatively correlated: when the value of $\lambda_2$ is smaller than 10 and $\lambda_3$ is larger than 10, the proposed model has stable and better performances. Our intuitive understanding is that paying more attention to the difference between the specific matrices and setting smaller weights on the sum of the specific matrices is helpful to detect anomalous changes. We can also observe in three datasets that the relationships between these two parameters are not much influenced by $\lambda_1$, which indicates that the parametric settings of the two specificity terms are independent of the consistency term. The promising result can be obtained when $\lambda_2 = 10$ and $\lambda_3 = 10$.

Then, we fix the values of $\lambda_2$ and $\lambda_3$ and observe the influence of $\lambda_1$ in three datasets. The results are shown in Fig. 13. It can be found that as the value of $\lambda_1$ increases, the performance of the SMSL model varies slightly. When the value of $\lambda_1$ equals 1, the SMSL model performs the best in the Object_1550_1558 dataset. For the Viareggio dataset, we can observe that the influence of $\lambda_1$ on the performance of the proposed model is not consistent, which may be caused by the different illumination conditions between two image pairs. For the “D1F12H1_D1F12H2” image pair, we have the best result when $\lambda_1 = 10$. The detection accuracy reaches the best when $\lambda_1 = 1$ in the “D1F12H1_D1F22H2” image pair. Therefore, the recommended range of $\lambda_1$ is $[1, 10]$ by experience.

3) Convergence Study: Finally, the convergence property of the proposed SMSL is explored on three datasets. As shown in Fig. 14, the stopping criteria related to the residuals at every iteration are given. It can be observed that the tendency of the
residuals keep decreasing in the subsequent iterations.

Fig. 14. Convergence curves about the overall stopping criteria of reconstruction errors versus the iterations on three datasets.

SMSL is quickly converged at the first ten iterations, and the residuals keep decreasing in the subsequent iterations.

V. CONCLUSION

In this article, which focuses on the hyperspectral ACD problem, we propose an SMSL model for large-scale multitemporal images. Unlike the existing multiview learning models that require redundant computational consumption and storage, the proposed SMSL model extracts the main information from the inputs and constructs a sketched dictionary for self-representation learning. In addition, the common information and the differences are simultaneously learned by defining a consistent part and specific parts for the coefficients. Furthermore, this article discussed a multiple suboptimization process using the ALM algorithm in detail and conducted sufficient experiments on large-scale hyperspectral datasets. The experimental results demonstrate the superiority of our approach compared to other classical methods of background suppression and anomalous changes extraction. In addition, the parametric analysis with respect to the tradeoff parameters and the convergence study of the model are also discussed.

ACKNOWLEDGMENT

The authors would like to thank the handling editor and the anonymous reviewers for their careful reading and helpful remarks. The authors would also like to thank the Institute of Advanced Research in Artificial Intelligence (IARAI) for its support.

REFERENCES

[1] N. Acito, M. Diani, G. Corsini, and S. Resta, “Introductory view of anomalous change detection in hyperspectral images within a theoretical Gaussian framework,” IEEE Aerosp. Electron. Syst. Mag., vol. 32, no. 7, pp. 2–27, Jul. 2017.
[2] F. Bovolo and L. Bruzzone, “A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain,” IEEE Trans. Geosci. Remote Sens., vol. 45, no. 1, pp. 218–236, Jan. 2006.
[3] Q. Wang, Z. Yuan, Q. Du, and X. Li, “GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 1, pp. 3–13, Jan. 2018.
[4] 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.
[5] R. T. Collins, A. J. Lipton, and T. Kanade, “Introduction to the special section on video surveillance,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 745–746, Aug. 2000.
[6] C. Stauffer and W. E. L. Grimson, “Learning patterns of activity using real-time tracking,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 747–757, Aug. 2000.
[7] L. Lemieux, U. C. Wiesemann, N. F. Moran, D. R. Fish, and S. D. Borovon, “The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization,” Med. Image Anal., vol. 2, no. 3, pp. 227–242, Sep. 1998.
[8] D. Grattarola, D. Zambon, L. Livi, and C. Alippi, “Change detection in graph streams by learning graph embeddings on constant-curvature manifolds,” IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 6, pp. 1856–1869, Jun. 2020.
[9] A. Tanega, L. Ballan, and M. Pollefeys, “Geometric change detection in urban environments using images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 11, pp. 2193–2206, Nov. 2015.
[10] G. Nagy, T. Zhang, W. R. Franklin, E. Landis, E. Nagy, and D. T. Keane, “Volume and surface area distributions of cracks in concrete,” in Proc. Int. Workshop Vis. Form., Berlin, Germany: Springer, 2001, pp. 759–768.
[11] Y. Wang et al., “Target tracking control of a biomimetic underwater vehicle through deep reinforcement learning,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 8, pp. 3741–3752, Aug. 2022.
[12] G. Vivekananda, R. Swathi, and A. Sujith, “Multi-temporal image analysis for ULC classification and change detection,” Eur. J. Remote Sens., vol. 54, no. 2, pp. 189–199, Mar. 2021.
[13] P. Ghamisi et al., “Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art,” IEEE Geosci. Remote Sens. Mag., vol. 7, no. 4, pp. 37–78, Dec. 2017.
[14] S. Chang, B. Du, and L. Zhang, “A subspace selection-based discriminative forest method for hyperspectral anomaly detection,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 6, pp. 4033–4046, Jun. 2020.
[15] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, “A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges,” IEEE Trans. Remote Sens. Mag., vol. 7, no. 2, pp. 140–158, Jun. 2019.
[16] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, “Image change detection algorithms: A systematic survey,” IEEE Trans. Image Process., vol. 14, no. 3, pp. 294–307, Mar. 2005.
[17] 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, pp. 1–12, 2022.
[18] M. T. Eismann and J. Meola, “Hyperspectral change detection: Methodology and challenges,” Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Jul. 2008, pp. II–605.
[19] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, “A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 11, pp. 6497–6504, Nov. 2016.
[20] C. Wu, B. Du, and L. Zhang, “Hyperspectral anomalous change detection based on joint sparse representation,” ISPRS J. Photogramm. Remote Sens., vol. 146, pp. 137–150, Dec. 2018.
[21] M. T. Eismann, J. Meola, and R. C. Hardie, “Hyperspectral change detection in the presence of diurnal and seasonal variations,” IEEE Trans. Geosci. Remote Sens., vol. 46, no. 1, pp. 237–249, Jan. 2008.
[22] A. Schaum, “Theoretical foundations of NRL spectral target detection algorithms,” Appl. Opt., vol. 54, no. 31, p. F286, 2015.
[23] J. Meola and M. T. Eismann, “Image misregistration effects on hyperspectral change detection,” Proc. SPIE, vol. 6966, May 2008, Art. no. 69660Y.
[24] A. P. Schaum and A. Stocker, “Hyperspectral change detection and supervised matched filtering based on covariance equalization,” Proc. SPIE, vol. 5425, pp. 77–90, Aug. 2004.
[25] J. A. Padron-Hidalgo, V. Laparra, N. Longbotham, and G. Camps-Valls, “Kernel anomalous change detection for remote sensing imagery,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 10, pp. 7743–7755, Oct. 2019.
[26] C. Wu, L. Zhang, and B. Du, “Hyperspectral anomaly change detection with slow feature analysis,” Neurocomputing, vol. 151, pp. 175–187, Mar. 2015.
[27] I. S. Reed and X. Yu, “Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution,” IEEE Trans. Acoust., Speech Signal Process., vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
[28] M. Ha, 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, 2021.
[29] M. Hu, C. Wu, and L. Zhang, “HyperNet: Self-supervised hyperspectral spatial–spectral feature understanding network for hyperspectral change detection,” 2022, arXiv:2207.09634.

[30] C. Zhang, Y. Cui, Z. Han, J. T. Zhou, H. Fu, and Q. Hu, “Deep partial multi-view learning,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 5, pp. 2402–2415, May 2022.

[31] P. S. Dhillon, D. P. Foster, and L. H. Ungar, “Multi-view learning of word embeddings via CCA,” in Proc. NIPS, 2011, pp. 1–9.

[32] J. Guo, Y. Sun, J. Gao, Y. Hu, and B. Yin, “Rank consistency induced multiview subspace clustering via low-rank matrix factorization,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 7, pp. 3157–3170, Jul. 2022.

[33] J. Zhao, X. Xie, X. Xu, and S. Sun, “Multi-view learning overview: Recent progress and new challenges,” Inf. Fusion, vol. 38, pp. 43–54, Nov. 2017.

[34] Y. Chen, X. Xiao, Z. Hua, and Y. Zhou, “Adaptive transition probability matrix learning for multiview spectral clustering,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 9, pp. 4712–4726, Sep. 2022.

[35] X. You, J. Xu, W. Yuan, X.-Y. Jing, D. Tao, and T. Zhang, “Multi-view common component discriminant analysis for cross-view classification,” Pattern Recognit., vol. 92, pp. 37–51, Aug. 2019.

[36] M. Yin, J. Gao, S. Xie, and Y. Guo, “Multiview subspace clustering via tensorial t-product representation,” IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 3, pp. 851–864, Mar. 2019.

[37] C. Zhang, H. Fu, J. Wang, W. Li, X. Cao, and Q. Hu, “Tensorized multi-view subspace representation learning,” Int. J. Comput. Vis., vol. 128, no. 8, pp. 2344–2361, 2020.

[38] E. Elhamifar and R. Vidal, “Sparse subspace clustering: Algorithm, theory, and applications,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[39] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, “Robust recovery of subspace structures by low-rank representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 1, pp. 171–184, Jan. 2013.

[40] Y. Guo, “Convex subspace representation learning from multi-view data,” in Proc. AAAI Conf. Artif. Intell., 2013, vol. 27, no. 1, pp. 1–7.

[41] Z. Ding and Y. Fu, “Robust multi-view subspace learning through low-rank decompositions,” in Proc. AAAI Conf. Artif. Intell., 2016, vol. 30, no. 1, pp. 1–7.

[42] S. Luo, C. Zhang, W. Zhang, and X. Cao, “Consistent and specific multi-view subspace clustering,” in Proc. 32nd AAAI Conf. Artif. Intell., 2018, pp. 1–8.

[43] P. A. Traganitis and G. B. Giannakis, “Sketched subspace clustering,” IEEE Trans. Signal Process., vol. 66, no. 7, pp. 1663–1675, Apr. 2018.

[44] H. Zhao, H. Zhang, L. Zhang, and P. Li, “Nonlocal means regularized sketched reweighted sparse and low-rank subspace clustering for large hyperspectral images,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 5, pp. 4164–4178, May 2021.

[45] X. Guo, “Exclusivity regularized machine,” 2016, arXiv:1603.08318.

[46] X. Wang, X. Gao, Z. Lei, C. Zhang, and S. Z. Li, “Exclusivity-consistency regularized multi-view subspace clustering,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 923–931.

[47] J. F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” SIAM J. Opt., vol. 20, no. 4, pp. 1956–1982, Jun. 2010.

[48] C. Zhang et al., “Generalized latent multi-view subspace clustering,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 1, pp. 86–99, Jan. 2020.

[49] J. Chen, S. Yang, H. Mao, and C. Fahy, “Multiview subspace clustering using low-rank representation,” IEEE Trans. Cybern., vol. 52, no. 11, pp. 12364–12378, Nov. 2022.

[50] Q. Zheng, J. Zhu, Z. Li, S. Pang, J. Wang, and Y. Li, “Feature concatenation multi-view subspace clustering,” Neurocomputing, vol. 379, pp. 89–102, Feb. 2020.

[51] B. Arad and O. Ben-Shahar, “Sparse recovery of hyperspectral signal from natural RGB images,” in Proc. Eur. Conf. Comput. Vis. Berlin, Germany: Springer, 2016, pp. 19–34.

[52] N. Acito, S. Matteoli, A. Rossi, M. Diani, and G. Corsini, “Hyperspectral airborne ‘Viareggio 2013 trial’ data collection for detection algorithm assessment,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 6, pp. 2365–2376, Jun. 2016.