Learning Structurally Incoherent Background and Target Dictionaries for Hyperspectral Target Detection

Tan Guo, Fulin Luo, *Member, IEEE*, Lei Zhang, *Senior Member, IEEE*, Bob Zhang, *Senior Member, IEEE*, Xiaoheng Tan, and Xiaocheng Zhou

***Abstract*—Existing sparsity-based hyperspectral image (HSI) target detection methods have two key problems. 1) The background dictionary is locally constructed by the pixels between the inner and outer windows, surrounding and enclosing the central test pixel. The dual-window strategy is intricate and might result in impure background dictionary deteriorating the detection performance. 2) For an unbalanced binary classification problem, the target dictionary atoms are generally inadequate compared with the background dictionary, which might yield unstable performance. For the issues, this article proposes a novel structurally incoherent background and target dictionaries (SIBTD) learning model for HSI target detection. Specifically, with the concept that the observed HSI data is composed of low-rank background, sparsely distributed targets, and dense Gaussian noise, the background and target dictionaries can be jointly derived from the observed HSI data. Additionally, the introduction of structural incoherence can enhances the discrimination between the target and background dictionaries. Thus, the developed model can not only lead to a pure and unified background dictionary but also augment the target dictionary for improved detection performance. Besides, an efficient optimization algorithm is devised to solve SIBTD model, and the performance of SIBTD is verified on three benchmark HSI datasets in comparison with several state-of-the-art detectors.**

Manuscript received May 2, 2020; revised May 26, 2020 and June 8, 2020; accepted June 9, 2020. Date of publication June 15, 2020; date of current version July 2, 2020. This work was supported in part by the National Key R&D Program of China under Grant 2019YFB2102001, in part by the National Natural Science Foundation of China under Grant 61801336 and Grant 61771079, in part by the Science and Technology Research Program of the Chongqing Municipal Education Commission under Grant KJQN201800632, in part by the China Postdoctoral Science Foundation under Grant 2019M662717, in part by the Open Research Project of the Hubei Key Laboratory of Intelligent Geo-Information Processing under Grant KLIGIP-2018A06, in part by the Fundamental Research Funds for the Central Universities under Grant 2042020kf0013, and in part by the Open Research Fund of Key Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University under Grant 2019LSDMIS06. *(Corresponding author: Fulin Luo.)*

Tan Guo is with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: guot@cqupt.edu.cn).

Fulin Luo is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: luoflyn@163.com).

Lei Zhang and Xiaoheng Tan are with the School of Microelectronics and Communications Engineering, Chongqing University, Chongqing 400044, China (e-mail: leizhang@cqu.edu.cn; txh@cqu.edu.cn).

Bob Zhang is with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: bobzhang@um.edu.mo).

Xiaocheng Zhou is with the Key Laboratory of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fuzhou 350116, China (e-mail: zhouxc@fzu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3002549

***Index Terms*—Dictionary decomposition, hyperspectral image (HSI), low-rank constraint, sparse model, target detection.**

I. Introduction

UNLIKE the conventional RGB image, which is composed of only three spectral bands, hyperspectral imaging sensors can sense reflected light in hundreds or even more than a thousand bands, and thus a 3-D data structure called hyperspectral image (HSI) can be obtained with two spatial dimensions and one spectral dimension [1], [2]. Thus, HSI can provide almost continuous spectral curves of the materials on the ground surface. Since different surface material can be characterized by different deterministic spectrums, the abundant spectral information in HSI makes it have great potential for target detection [3], [5]. As the name implies, target detection techniques aim to separate pixels with specific characteristic from its surrounding background pixels, which has been widely applied in both civil and military applications, such as detecting rare minerals, land mines, and man-made objects. When the target prior knowledge is unavailable, target detection becomes anomaly detection aiming at discriminating abnormal materials from a normal background [6], [7].

Obviously, target detection can be considered as a binary classification problem wherein the test pixels are required to be labeled as target or background. In fact, a high-quality target detector can calculate and output the response values corresponding to the input test pixels and encourage the response differences between the target and background pixels as much as possible [8]–[10]. Over past decades, a lot of target detection approaches have been developed, which can be roughly classified into two types: Probabilistic model and subspace model. Probabilistic models propose to model the background and target components in HSI with certain distribution, e.g., multivariate normal distribution [11]. Classical probabilistic models include spectral matched filter (SMF) [12] and adaptive coherence estimator (ACE) [13]. For example, the SMF model first estimates the background covariance matrix and then adopts the generalized likelihood ratio test to perform detection with a single target spectrum. Alternatively, subspace models assume that the background or target pixels lie in a low-dimensional subspace with different distribution. Typical subspace-based models are orthogonal subspace projection (OSP) [14] and sparsity-based

detectors [15], [16]. The OSP assumes that the target has some components orthogonal to the background subspace and can be detected by maximizing the signal-to-noise (SNR) ratio in the subspace orthogonal to the background subspace. In recent years, sparse representation (SR) has been emerging as an efficient methodology and has been successfully applied for many computer vision applications [38]–[50]. With SR theory, the sparse representation-based target detector (SRD) has been developed, which first represents a test pixel using the union of the background and target dictionaries via sparsity-inducing algorithms, such as the orthogonal matching pursuit [17], and the test pixel is classified by comparing the residuals between the input test pixel with the two test pixels respectively reconstructed by the target and background dictionaries. In the sparsity-based detectors, no explicit assumption on the statistical distribution characteristics, such as Gaussian distribution, about the background and target distribution is needed. Besides, the independence between the training samples is unnecessary, and thus better model generalization ability can be expected [15], [16], [24], [25]. However, owing to imaging technology, the high spectral resolution is always obtained at the expense of spatial resolution for compromise. Limited by low spatial resolution, the HSI usually contains both pure and mixed pixels. A pure pixel contains only one single material, whereas a mixed pixel or subpixel might cover multiple materials, and its spectral signature represents the aggregate of different materials in the corresponding spatial location. Therefore, a pixel of HSI is possibly a combination of different materials' spectra, and target detection is converted into the problem of determining whether the target is present or absent in test pixel. Hence, the problem can be considered from the perspective of binary hypothesis model, i.e., target absent hypothesis and target present hypothesis. Based on the binary hypothesis, the sparse representation-based binary hypothesis detector (SRBBHD) has been suggested wherein the test pixel is modeled by the background dictionary under the null hypothesis or by the background and target union dictionary under the alternative hypothesis [18]. Thus, the difference prior knowledge between the target and background dictionaries can be fully utilized.

Nevertheless, an important question for SRD and its variants is the constructions of the background and target dictionaries. Typically, the background dictionary can be constructed via some background samples, e.g., trees, road, buildings, and vegetation. However, a universal background dictionary constructed by all the background pixels requires high computational costs and makes the problem even more unbalanced. Therefore, the background dictionary is usually generated locally for each test pixel via a dual concentric window, which can separate the local area around each test pixel into two regions, a small inner window region (IWR) centered within a larger outer window region (OWR) [19]. On the other hand, target dictionary is generally constructed by some priori spectral libraries that contain the targets of interest. Target pixels are, however, usually insufficient, which results in the problem that the target training samples and background training samples are not equivalent in data volume.

Consequently, a more elaborately designed dictionary would be desirable instead of using the raw samples as the dictionary due to the large distribution difference between background and target components as well as the sparsity of the targets [20], [21]. Thus, directly learning compact and discriminative target and background dictionaries from data is more flexible and data-driven [22], [23]. By considering the data characteristics and prior knowledge of HSI, the background dictionary should have the property of low rankness, while the target components are sparsely distributed in spatial. Based on the above considerations, the HSI data is regarded as being made up of a low-rank background HSI, a sparse target HSI, and a Gaussian noise component HSI. As shown in Fig. 1, this article proposes to directly learn background and target dictionaries from the observed HSI data via sparsity and low-rank constraints, and a novel structurally incoherent background and target dictionaries (SIBTD) learning model is developed for HSI target detection. In summary, the main contributions and novelties of this article can be summarized as follows.

1) SIBTD can learn target and background dictionaries from observed HSI data via low-rank and sparsity constraints. Specifically, the background component is modeled using a union of multiple subspaces, leading to a pure and unified background dictionary. Equally important is that a target compensation dictionary can be simultaneously obtained and further concatenated with the known target spectrum for an augmented target dictionary.
2) The obtained background and target compensation dictionaries are encouraged to be as independent as possible for enhanced discriminative ability. Besides, an efficient iteration-based optimization algorithm is carefully devised to solve the SIBTD model such that a promising solution can be guaranteed.
3) HSI target detection is performed by combining the developed SIBTD model with the classwise collaborative representation model, and the experimental results on three HSI datasets show that the proposed target detection method can yield promising performance.

The reminder of this article is structured as follows. The related works are briefly reviewed in Section II. Section III presents the proposed SIBTD model and its optimization procedure. The experimental results and discussions are given in Section IV. Section V concludes this article.

II. Related Work

A. SRD and SRBBH

Suppose an HSI dataset is with the size of $h \times w \times b$, where h and w are the height and width of the image scene, respectively, and b is the number of spectral bands. The observed 3-D HSI can be rearranged into a 2-D matrix by sequentially ordering the pixels. Thus, the HSI dataset can be obtained in matrix form as $\mathbf{X} \in \Re^{b \times e}$ with $= h \times w$. By assuming that pixels belonging to the same category approximately lie in a low-dimensional subspace spanned by the training samples, SRD employs training samples from both the target and background classes to linearly

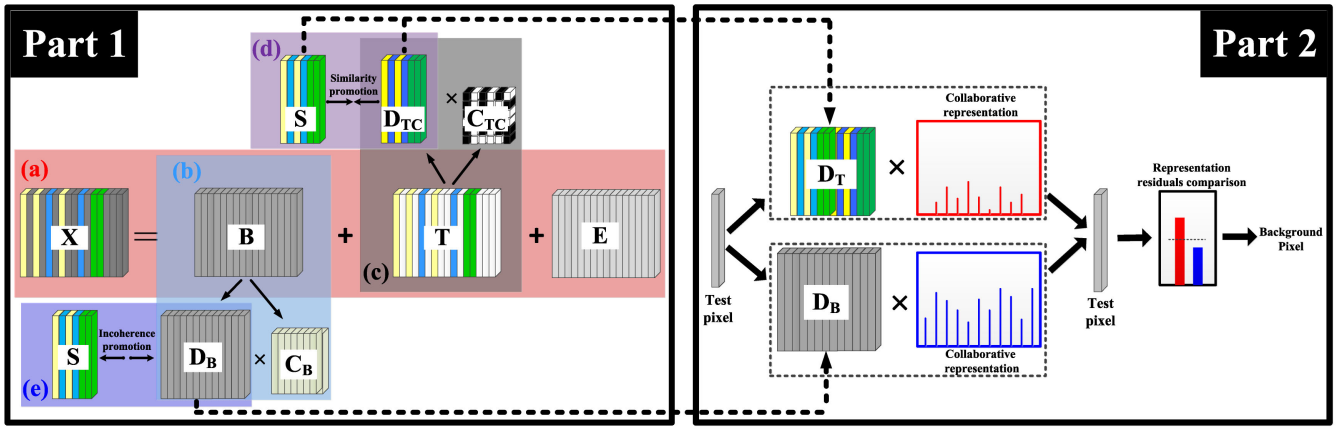

Fig. 1. Illustration for the proposed structurally incoherent background and target dictionaries learning model for HSI target detection. **Part 1**: SIBTD model. (a) Observed HSI dataset $\mathbf{X}$ is decomposed to get the low-rank background component $\mathbf{B}$, the sparsely distributed target component $\mathbf{T}$, and the noise component $\mathbf{E}$. $\mathbf{B}$ and $\mathbf{T}$ are further simultaneously decomposed to pursue a pure and unified background dictionary $\mathbf{D}_{\mathrm{B}}$ and a target compensation dictionary $\mathbf{D}_{\mathrm{TC}}$ as shown in (b) and (c). As (d) and (e) show, $\mathbf{D}_{\mathrm{TC}}$ and $\mathbf{D}_{\mathrm{B}}$ are learnt with the supervision of known target spectrum $\mathbf{S}$, and $\mathbf{D}_{\mathrm{TC}}$ is concatenated with $\mathbf{S}$ for final augmented and discrimination enhanced target dictionary $\mathbf{D}_{\mathrm{T}}$, i.e., $\mathbf{D}_{\mathrm{T}} = [\mathbf{S}, \mathbf{D}_{\mathrm{TC}}]$. **Part 2**: Classwise representation learning for target detection. The test pixel is respectively represented by the target dictionary and background dictionary derived from the developed SIBTD model via collaborative representation technology. The final target detection decision is made by comparing the representation residuals calculated on the target dictionary and background dictionary.

model the class-unknown test pixels. In SRD, a test pixel in an HSI is first represented as a sparse linear combination of a number of samples from the target and background union dictionary, and then the detection decision is made by checking which dictionary can yield the smallest representation residual [15]. To be specific, a test pixel $\boldsymbol{x}$ can be represented in terms of the target and background dictionary pixels as follows:

$$\begin{aligned}\boldsymbol{x} \approx\ & \left(c_{B,1}\boldsymbol{d}_{B,1} + c_{B,2}\boldsymbol{d}_{B,2} + \cdots + c_{B,N_B}\boldsymbol{d}_{B,N_B}\right) \\ & + \left(c_{T,1}\boldsymbol{d}_{T,1} + c_{T,2}\boldsymbol{d}_{T,2} + \cdots + c_{T,N_T}\boldsymbol{d}_{T,N_T}\right) \\ \approx\ & \mathbf{D}_{\mathrm{B}}\boldsymbol{c}_{\mathrm{B}} + \mathbf{D}_{\mathrm{T}}\boldsymbol{c}_{\mathrm{T}} = \mathbf{D}\boldsymbol{c}. \end{aligned} \tag{1}$$

The union dictionary $\mathbf{D} \in \Re^{b\times(N_B+N_T)}$ is composed of the background dictionary $\mathbf{D}_{\mathrm{B}} \in \Re^{b\times N_B}$ with N_B atoms and the target dictionary $\mathbf{D}_{\mathrm{T}} \in \Re^{b\times N_T}$ with N_T atoms. $\boldsymbol{c}_{\mathrm{B}}$ and $\boldsymbol{c}_{\mathrm{T}}$ are SR vectors whose entries correspond to the weights of the samples in $\mathbf{D}_{\mathrm{B}}$ and $\mathbf{D}_{\mathrm{T}}$ for representing test pixel $\boldsymbol{x}$. An SR with only a few nonzero entries is proved to be efficient in revealing the membership of the test pixel. Thus, the following l_0-norm minimization problem has been formulated for a discriminative sparse solution

$$\min_{\boldsymbol{c}} \|\boldsymbol{x} - \mathbf{D}\boldsymbol{c}\|_2 \ \text{s.t.}\ \|\boldsymbol{c}\|_0 \le K_0. \tag{2}$$

During the optimization process, the pixels from the background and target dictionaries compete against the others to gain their share in representing the test pixel. The residuals recovered from the two subspaces spanned by the target and background dictionaries are calculated as follows:

$$R_B(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{D}_{\mathrm{B}}\boldsymbol{c}_{\mathrm{B}}\|_2 \tag{3}$$

$$R_T(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{D}_{\mathrm{T}}\boldsymbol{c}_{\mathrm{T}}\|_2. \tag{4}$$

The discriminative information derived from the representation residuals on the two subdictionaries can be utilized for detecting targets. Specifically, the detection decision can be made by comparing the above two residuals

$$\mathbf{D}(\boldsymbol{x}) = R_B(\boldsymbol{x}) - R_T(\boldsymbol{x}). \tag{5}$$

If the above difference value is larger than a certain threshold, the pixel can be identified as target, or background otherwise. However, SRD calculates a uni-structure representation on the union dictionary and ignores the difference prior knowledge between the target and background dictionaries. SRBBHD is proposed to alleviate this problem based on the binary hypotheses [18]. Specifically, a given test pixel can be modeled via the following binary hypothesis:

$$\mathrm{H}_0 :\ \boldsymbol{x} = \mathbf{D}_{\mathrm{B}}\boldsymbol{c}_{\mathrm{B}} + \boldsymbol{e}\ (\text{Targetabsent}) \tag{6}$$

$$\mathrm{H}_1 :\ \boldsymbol{x} = \mathbf{D}_{\mathrm{B}}\boldsymbol{c}'_{\mathrm{B}} + \mathbf{D}_{\mathrm{T}}\boldsymbol{c}_{\mathrm{T}} + \boldsymbol{e}\ = \mathbf{D}\boldsymbol{c} + \boldsymbol{e}\ (\text{Targetpresent}). \tag{7}$$

The above representation vectors on the background and target dictionaries can be respectively calculated by solving the following optimization problems:

$$\min_{\boldsymbol{c}_{\mathrm{B}}} \|\boldsymbol{x} - \mathbf{D}_{\mathrm{B}}\boldsymbol{c}_{\mathrm{B}}\|_2 \ \text{s.t.}\ \|\boldsymbol{c}_{\mathrm{B}}\|_0 \le K_1 \tag{8}$$

$$\min_{\boldsymbol{c}} \|\boldsymbol{x} - \mathbf{D}\boldsymbol{c}\|_2 \ \text{s.t.}\ \|\boldsymbol{c}\|_0 \le K_2. \tag{9}$$

The competition between two hypotheses is revealed by the competition between the above two representation residuals

$$D(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{D}_{\mathrm{B}}\widehat{\boldsymbol{c}_{\mathrm{B}}}\|_2 - \|\boldsymbol{x} - \mathbf{D}\hat{\boldsymbol{c}}\|_2. \tag{10}$$

Similarly, if the difference value is larger than certain threshold, the test pixel can be claimed as a target, or background otherwise. Nevertheless, a key problem for many sparsity-based target detection methods is the constructions of target and background dictionaries [2], [26], [27]. Generally, the target dictionary can be formed via the target training pixels that are selected from the

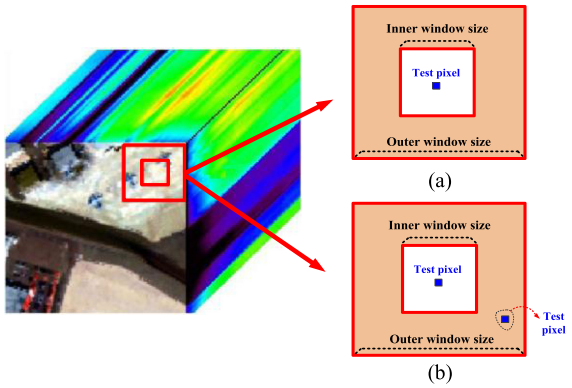

Fig. 2. Dual-window strategy for background dictionary construction in sparsity-based detectors. (a) Expected pure background dictionary. However, during the sliding process of dual window, some target pixels might inevitably fall into the dual window, especially when targets are densely and evenly distributed in the scene, as shown in (b), which will lead to an impure background dictionary for the test pixel, and result in deteriorated performance.

global image scene, and the priori information is usually given by the target spectrum obtained from a target spectral library. As for background dictionary, a universal dictionary constructed by all the training samples needs high computational costs and makes the problem even more unbalanced. Therefore, a locally adaptive method is employed for the background pixel selection. In detail, the background dictionary is generated locally for each test pixel through a dual concentric window that separates the local area around each test pixel into two regions: A small inner window centered within a larger outer window. The IWR is used to exclude the pixels of interest to be tested, while the OWR is employed to model the local background around the test pixel. Since there is no specific method for setting the sizes of the detection windows, the window sizes are set manually [19]. The size of IWR should be set larger than or equal to the size of all the possible targets of interests in HSI to avoid the target pixels from appearing in background dictionary as shown in Fig. 2(a). However, as shown in Fig. 2(b), the target pixel might appear in the background dictionary, which will lead to an impure background dictionary. Thus, a universal and compact background dictionary would be preferred and urgently needed.

B. Robust Principal Component Analysis and Low-Rank Representation Models

Suppose a dataset $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n] \in \Re^{b\times n}$ contains n training data points with dimensionality b, conventional principal component analysis (PCA) learns a low-rank projection $\mathbf{V} \in \Re^{b\times r}$ by solving the following problem:

$$\min_{\boldsymbol{V}} \left\|\mathbf{X} - \boldsymbol{V}\boldsymbol{V}^T\mathbf{X}\right\|_F^2 \text{ s.t. } \boldsymbol{V}^T\boldsymbol{V} = \mathbf{I}_r. \tag{11}$$

The solution to the above optimization problem can be efficiently obtained by singular value decomposition, which is computationally stable. Due to the advantages of computational stability and efficiency, PCA has been one of the most widely applied tools for error correction [28]. However, in real scenario, the application and performance of PCA are limited due to lack of robustness to gross corruptions. To overcome the drawback, recent years have witnessed a surge of robust principal component analysis (RPCA) methods. Candes *et al*. [29] established a RPCA model which has been emerging as a powerful tool for many applications. The model is mathematically formulated as follows:

$$\min_{\boldsymbol{D},\boldsymbol{E}} \|\boldsymbol{D}\|_* + \lambda\|\boldsymbol{E}\|_1 \text{ s.t. } \mathbf{X} = \boldsymbol{D} + \boldsymbol{E} \tag{12}$$

where $\lambda > 0$ is a parameter and $\|\cdot\|_*$ denotes the nuclear norm, i.e., a convex approximation for rank function. RPCA can well handle the gross corruptions with large magnitude and sparsely distributed noise $\mathbf{E}$ using sparsity-induced norm $\|\cdot\|_1$. However, RPCA implicitly assumes that the underlying data structure is a single low-rank subspace, which is inapplicable when the data is drawn from a union of multiple subspaces. To cope with the data with complex structure, Liu *et al.* [30] suggest a more general rank minimization model defined as follows:

$$\min_{\boldsymbol{C},\boldsymbol{E}} \|\boldsymbol{C}\|_* + \lambda\|\boldsymbol{E}\|_{2,1} \text{ s.t. } \mathbf{X} = \mathbf{D}\boldsymbol{C} + \boldsymbol{E} \tag{13}$$

where $\mathbf{D}$ is a dictionary that can linearly span the data space with multiple subspaces and $\mathbf{C}$ is the low-rank representation matrix of $\mathbf{X}$ on $\mathbf{D}$. RPCA can be seen as a special case of LRR by setting $\mathbf{D} = \mathbf{I}$. Benefiting from the well modeling for data structure and noise, LRR can well handle the data drawn from a union of multiple subspaces.

III. Proposed Structurally Incoherent Background and Target Dictionaries Learning-Based Hyperspectral Target Detection

This section will present the proposed SIBTD learning model for hyperspectral target detection. The model formulation and the optimization procedure will be introduced in detail.

A. Model Formulation

The two formulations in SRBBH can be rewritten in a more general additive model, i.e.,

$$\mathbf{x} = \mathbf{D}_\mathrm{B}\mathbf{c}_\mathrm{B} + \mathbf{D}_\mathrm{T}\mathbf{c}_\mathrm{T} + \boldsymbol{e}. \tag{14}$$

If $\mathbf{c}_\mathrm{T} \approx 0$, the target is absent, and the target is probably present when $\mathbf{c}_\mathrm{T} \neq 0$. The matrix form corresponding to the above formula is

$$\mathbf{X} = \mathbf{D}_\mathrm{B}\mathbf{C}_\mathrm{B} + \mathbf{D}_\mathrm{TC}\mathbf{C}_\mathrm{TC} + \mathbf{E}. \tag{15}$$

The above formula shows that the observed HSI dataset $\mathbf{X} \in \Re^{b\times n}$ can be represented by the background dictionary $\mathbf{D}_\mathrm{B} \in \Re^{b\times n_B}$ and the target dictionary $\mathbf{D}_\mathrm{T} \in \Re^{b\times n_T}$. Conversely, $\mathbf{D}_\mathrm{B}$ and $\mathbf{D}_\mathrm{T}$ can be decomposed and learned from $\mathbf{X}$. To this end, some regularizations and constraints are required. In HSI, a key observation is that the background component contains some different materials with multiple spectra such that the background samples should be drawn from several low-dimensional subspaces. Thus, a low-rank-based background learning formula is suggested with a low-rank background dictionary $\mathbf{D}_\mathrm{B}$ and a

background representation matrix $\mathbf{C}_{\mathrm{B}}$ as

$$\min_{\mathbf{D}_{\mathrm{B}},\mathbf{D}_{\mathrm{T}},\mathbf{C}_{\mathrm{T}},\mathbf{C}_{\mathrm{B}},E} \ \operatorname{rank}(\mathbf{D}_{\mathrm{B}})+\psi(\mathbf{D}_{\mathrm{TC}},\mathbf{C}_{\mathrm{TC}})+\gamma\|\mathbf{E}\|_F^2$$
$$\text{s.t.} \quad X=\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}+\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}+\mathbf{E}. \tag{16}$$

Instead of learning target dictionary $\mathbf{D}_{\mathrm{T}}$, the above model aims to learn a target compensation dictionary $\mathbf{D}_{\mathrm{TC}}\in\Re^{b\times n_{TC}}$ with n_{TC} atoms, and $\mathbf{D}_{\mathrm{TC}}\in\Re^{b\times n_{TC}}$ has the same size with the known target spectrum $\mathbf{S}$. The number of atoms in background dictionary can be set as $n_B=\tau\times n_{\mathrm{TC}}$ with $\tau\geq 1$ for simplicity. Consider that $\operatorname{rank}(\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}})\leq \min\{\operatorname{rank}(\mathbf{D}_{\mathrm{B}}),\operatorname{rank}(\mathbf{C}_{\mathrm{B}})\}$, and a slack version can be expressed as $\|\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}\|_*\leq\min\{\|\mathbf{D}_{\mathrm{B}}\|_*,\|\mathbf{C}_{\mathrm{B}}\|_*\}$. Thus, a low-rank background dictionary $\mathbf{D}_{\mathrm{B}}$ will promote the background component $\mathbf{B}=\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}$ to be of low rank. Dense Gaussian noise in data is modeled by $\|\mathbf{E}\|_F^2$. The term $\psi(\mathbf{D}_{\mathrm{TC}},\mathbf{C}_{\mathrm{TC}})$ refers to some regularizations for the target compensation dictionary $\mathbf{D}_{\mathrm{TC}}$ and the corresponding representation matrix $\mathbf{C}_{\mathrm{TC}}$. Besides, the total image area of all the targets should be spatially small relative to the whole image. Due to the properties of matrix multiplication, right multiplication with a column sparse matrix will lead to a sparse matrix by column. Therefore, a regularization term for the column sparsity is added to model $\mathbf{C}_{\mathrm{TC}}$ as $\sum_{j=1}^{n}\sqrt{\sum_{i=1}^{n_{\mathrm{TC}}}(\mathbf{C}_{\mathrm{TC}})_{ij}^2}=\|\mathbf{C}_{\mathrm{TC}}\|_{2,1}$, which is utilized as a surrogate for $\|\mathbf{C}_{\mathrm{TC}}\|_{2,0}$ to obtain a column-sparse solution. Thus, the proposed SIBTD learning model is finally mathematically formulated as follows:

$$\min_{\mathbf{D}_{\mathrm{B}},\mathbf{D}_{\mathrm{TC}},\mathbf{C}_{\mathrm{TC}},\mathbf{C}_{\mathrm{B}},\mathbf{E}} \|\mathbf{D}_{\mathrm{B}}\|_*+\alpha\|\mathbf{C}_{\mathrm{B}}\|_F^2+\beta\|\mathbf{C}_{\mathrm{TC}}\|_{2,1}+\gamma\|\mathbf{E}\|_F^2$$
$$+\eta\|\mathbf{S}^T\mathbf{D}_{\mathrm{B}}\|_F^2+\frac{\varphi}{2}\|\mathbf{D}_{\mathrm{TC}}-\mathbf{S}\|_F^2$$
$$\text{s.t.} \quad \mathbf{X}=\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}+\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}+\mathbf{E} \tag{17}$$

where $\|\mathbf{D}_{\mathrm{TC}}-\mathbf{S}\|_F^2$ can encourage the similarity between target compensation dictionary $\mathbf{D}_{\mathrm{TC}}$ and the known target spectrum $\mathbf{S}$ measured by Frobenius norm. $\|\mathbf{C}_{\mathrm{B}}\|_F^2$ can make the solution for $\mathbf{C}_{\mathrm{B}}$ more stable and avoid overfitting. The term $\|\mathbf{S}^T\mathbf{D}_{\mathrm{B}}\|_F^2$ promotes the independence between background dictionary $\mathbf{D}_{\mathrm{B}}$ with $\mathbf{S}$. Note the property that $\|\mathbf{S}^T\mathbf{D}_{\mathrm{B}}\|_F^2\leq\|\mathbf{S}\|_F^2\|\mathbf{D}_{\mathrm{B}}\|_F^2$, and minimizing $\|\mathbf{S}^T\mathbf{D}_{\mathrm{B}}\|_F^2$ can be achieved by minimizing $\|\mathbf{S}_F^2\mathbf{D}_{\mathrm{B}}\|_F^2$. Model (17) can be relaxed into the following one:

$$\min_{\mathbf{D}_{\mathrm{B}},\mathbf{D}_{\mathrm{TC}},\mathbf{C}_{\mathrm{TC}},\mathbf{C}_{\mathrm{B}},\mathbf{E}} \|\mathbf{D}_{\mathrm{B}}\|_*+\alpha\|\mathbf{C}_{\mathrm{B}}\|_F^2+\beta\|\mathbf{C}_{\mathrm{TC}}\|_{2,1}$$
$$+\gamma\|\mathbf{E}\|_F^2+\eta'\|\mathbf{D}_{\mathrm{B}}\|_F^2+\frac{\varphi}{2}\|\mathbf{D}_{\mathrm{TC}}-\mathbf{S}\|_F^2$$
$$\text{s.t.} \quad \mathbf{X}=\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}+\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}+\mathbf{E} \tag{18}$$

where $\eta'=\eta\|\mathbf{S}\|_F^2$ is a constant during model optimization. In summary, the above model can simultaneously separate the low-rank background, sparse target, and dense noise components from the observed HSI data. The obtained background component is further decomposed to pursue a pure and unified background dictionary $\mathbf{D}_{\mathrm{B}}$. Besides, the obtained target component is separated to get the target compensation dictionary $\mathbf{D}_{\mathrm{TC}}$ with the supervision of known target spectrum $\mathbf{S}$. $\mathbf{D}_{\mathrm{TC}}$ can be concatenated with $\mathbf{S}$ for an augmented and discrimination enhanced target dictionary $\mathbf{D}_{\mathrm{T}}$ as $\mathbf{D}_{\mathrm{T}}=[\mathbf{D}_{\mathrm{TC}},\mathbf{S}]$. Thus, the problem of insufficient target samples can be alleviated. Optionally, one can first augment the known target spectrum, obtaining an augmented known target spectrum $\mathbf{S}_{\mathrm{Aug}}$ several times than the original known target spectrum in data volume. Then, the obtained augmented known target spectrum $\mathbf{S}_{\mathrm{Aug}}$ can be input into (18) instead of $\mathbf{S}$ for dictionaries learning.

B. Model Optimization

To solve the proposed model (18), an auxiliary variable $\mathbf{J}$ is first introduced, and the optimization problem can be converted into the following equivalent one:

$$\min_{\mathbf{D}_{\mathrm{B}},\mathbf{C}_{\mathrm{B}},\mathbf{J},\mathbf{D}_{\mathrm{TC}},\mathbf{C}_{\mathrm{TC}},\mathbf{E}} \|\mathbf{J}\|_*+\alpha\|\mathbf{C}_{\mathrm{B}}\|_F^2+\beta\|\mathbf{C}_{\mathrm{TC}}\|_{2,1}+\gamma\|\mathbf{E}\|_F^2$$
$$+\eta'\|\mathbf{D}_{\mathrm{B}}\|_F^2+\frac{\varphi}{2}\|\mathbf{D}_{\mathrm{TC}}-\mathbf{S}\|_F^2$$
$$\text{s.t.} \quad \mathbf{X}=\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}+\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}+\mathbf{E},\ \mathbf{D}_{\mathrm{B}}=\mathbf{J} \tag{19}$$

which can be solved based on the Augmented Lagrange Multiplier scheme [31]. The augmented Lagrangian function for the above problem is formulated as follows:

$$\begin{aligned}\mathcal{L}(\mathbf{D}_{\mathrm{B}},\mathbf{J},\mathbf{D}_{\mathrm{TC}},\mathbf{C}_{\mathrm{TC}},\mathbf{C}_{\mathrm{B}},\mathbf{E},\boldsymbol{Y}_1,\boldsymbol{Y}_2)&=\|\mathbf{J}\|_*+\alpha\|\mathbf{C}_{\mathrm{B}}\|_F^2\\&+\beta\|\mathbf{C}_{\mathrm{TC}}\|_{2,1}+\gamma\|\mathbf{E}\|_F^2+\eta'\|\mathbf{D}_{\mathrm{B}}\|_F^2+\frac{\varphi}{2}\|\mathbf{D}_{\mathrm{TC}}-\mathbf{S}\|_F^2\\&+tr\left(\mathbf{Y}_1^T(\mathbf{X}-\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}-\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{E})\right)\\&+tr\left(\mathbf{Y}_2^T(\mathbf{D}_{\mathrm{B}}-\mathbf{J})\right)+\frac{\mu}{2}(\|\mathbf{X}-\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}\\&-\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{E}\|_F^2+\|\mathbf{D}_{\mathrm{B}}-\mathbf{J}\|_F^2)\end{aligned} \tag{20}$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are the Lagrange multipliers and $\mu>0$ is a penalty parameter. The variables $\mathbf{D}_{\mathrm{B}},\mathbf{J},\mathbf{D}_{\mathrm{TC}},\mathbf{C}_{\mathrm{TC}},\mathbf{C}_{\mathrm{B}},\mathbf{E}$ are updated alternately by minimizing the loss function with the other variables fixed. The updating procedure is as follows.

Step 1 (Updating *J*): Fix the other variables and solve the following problem:

$$\min_{\boldsymbol{J}}\frac{1}{\mu}\|\boldsymbol{J}\|_*+\frac{1}{2}\left(\|\boldsymbol{J}-\left(\mathbf{D}_{\mathrm{B}}+\frac{\mathbf{Y}_2}{\mu}\right)\|_F^2\right)=\mathbf{U}\boldsymbol{\mathcal{S}}_{1/\mu}[\Sigma]\boldsymbol{V}^T. \tag{21}$$

where $(\mathbf{U},\boldsymbol{\Sigma},\mathbf{V}^T)=\mathrm{SVD}(\mathbf{D}_{\mathrm{B}}+\frac{\mathbf{Y}_2}{\mu})$ and $\boldsymbol{\mathcal{S}}_{\varepsilon}[\cdot]$ is the soft thresholding (shrinkage) operator defined as follows [31]:

$$\boldsymbol{\mathcal{S}}_{\varepsilon}[Z]=\begin{cases}z-\varepsilon, & \text{if } z>\varepsilon\\ z+\varepsilon, & \text{if } z<-\varepsilon\\ 0, & \text{otherwise}\end{cases}. \tag{22}$$

Step 2 (Updating $\boldsymbol{C}_{\mathrm{B}}$): Fix the other variables and solve the following problem:

$$\operatorname*{argmin}_{\mathbf{C}_{\mathrm{B}}}\alpha\mathbf{C}_{\mathrm{B}}{}_F^2+\frac{\mu}{2}\left(\|\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}-\left(\mathbf{X}-\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{E}+\frac{\mathbf{Y}_1}{\mu}\right)\|_F^2\right). \tag{23}$$

Letting the derivation of the objective function w.r.t $\mathbf{C}_{\mathrm{B}}$ to be $\mathbf{0}$, the following equation can be obtained:

$$2\alpha\mathbf{C}_{\mathrm{B}}+\mu\mathbf{D}_{\mathrm{B}}^T\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}=\mu\mathbf{D}_{\mathrm{B}}^T\left(\mathbf{X}-\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{E}+\frac{\mathbf{Y}_1}{\mu}\right). \tag{24}$$

The solution for $\mathbf{C}_{\mathrm{B}}$ is obtained as follows:

$$\mathbf{C}_{\mathrm{B}} = \mu\big(2\alpha\mathbf{I}+\mu\mathbf{D}_{\mathrm{B}}^{T}\mathbf{D}_{\mathrm{B}}\big)^{-1}\mathbf{D}_{\mathrm{B}}^{T}\left(\mathbf{X}-\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{E}+\frac{\mathbf{Y}_1}{\mu}\right). \tag{25}$$

Step 3 (Updating $\boldsymbol{C}_{\mathrm{TC}}$): Fix the other variables and solve the following problem:

$$\min_{\mathbf{C}_{\mathrm{TC}}} \beta\mathbf{C}_{\mathrm{TC}2,1}+\frac{\mu}{2}\left(\|\mathbf{X}-\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}-\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{E}+\frac{\mathbf{Y}_1}{\mu}\|_F^2\right). \tag{26}$$

Let $\mathbf{R} = \mathbf{X}-\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}-\mathbf{E}+\frac{\mathbf{Y}_1}{\mu}$, and the problem is transformed to the following form:

$$\min_{\mathbf{C}_{\mathrm{T}}} \beta\mathbf{C}_{\mathrm{TC}2,1}+\frac{\mu}{2}\left(\|\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{R}\|_F^2\right). \tag{27}$$

The above $l_{2,1}$-norm plus Frobenius norm combined minimization problem can also be solved using ALM scheme. First, an auxiliary variable $\boldsymbol{L}$ is introduced to the make the problem easy to be optimized

$$\min_{\mathbf{C}_{\mathrm{TC}},\boldsymbol{L}} \beta\|\boldsymbol{L}\|_{2,1}+\frac{\mu}{2}\left(\|\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{R}\|_F^2\right) \text{s.t. } \mathbf{C}_{\mathrm{TC}} = \mathbf{L}. \tag{28}$$

The augmented Lagrange function is

$$\begin{aligned}\mathcal{L}\left(\mathbf{C}_{\mathrm{TC}},\boldsymbol{L},\boldsymbol{Z}\right) = \beta\|\boldsymbol{L}\|_{2,1}+\frac{\mu}{2}\left(\|\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{R}\|_F^2\right)\\ +\frac{\mu_{\mathbf{C}_{\mathrm{TC}}}}{2}\left(\|\mathbf{C}_{\mathrm{TC}}-\boldsymbol{L}+\frac{\mathbf{Z}}{\mu_{\mathbf{C}_{\mathrm{TC}}}}\|_F^2\right)\end{aligned} \tag{29}$$

where $\mathbf{Z}$ is the Lagrange multiplier, and $\mu_{\boldsymbol{C}_{\mathrm{TC}}} > 0$ is the penalty parameter. The subproblems for $\mathbf{L}$ and $\mathbf{C}_{\mathrm{TC}}$ are solved alternatively in the following way.

Updating $\mathbf{L}$

$$\min_{L} \frac{\beta}{\mu_{\mathbf{C}_{\mathrm{TC}}}}\|\boldsymbol{L}\|_{2,1}+\frac{1}{2}\left(\|\boldsymbol{L}-\left(\mathbf{C}_{\mathrm{TC}}+\frac{\mathbf{Z}}{\mu_{\mathbf{C}_{\mathrm{TC}}}}\right)\|_F^2\right) \tag{30}$$

which can be solved using the following lemma [32].

Lemma 1. Given a matrix $\mathcal{Q}$, if $\boldsymbol{\mathcal{W}}^*$ is the optimal solution for the following problem:

$$\min_{\boldsymbol{\mathcal{W}}} \xi\|\boldsymbol{\mathcal{W}}\|_{2,1}+\frac{1}{2}\|\boldsymbol{\mathcal{W}}-\mathcal{Q}\|_F^2 \tag{31}$$

then the i-column of $\boldsymbol{\mathcal{W}}^*$ is

$$[\boldsymbol{\mathcal{W}}^*]_{:,i} = \begin{cases}\frac{\|[\mathcal{Q}]_{:,i}\|_2-\xi}{\|[\mathcal{Q}]_{:,i}\|_2}\mathcal{Q}_{:,i}, & \text{if } \|[\mathcal{Q}]_{:,i}\|_2 > \xi\\ 0, & \text{otherwise}\end{cases}. \tag{32}$$

Updating $\boldsymbol{C}_{\mathrm{TC}}$

$$\min_{\mathbf{C}_{\mathrm{TC}}} \frac{\mu}{2}\left(\|\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{R}\|_F^2\right)+\frac{\mu_{\mathbf{C}_{\mathrm{TC}}}}{2}\left(\|\mathbf{C}_{\mathrm{TC}}-L+\frac{\mathbf{Z}}{\mu_{\mathbf{C}_{\mathrm{TC}}}}\|_F^2\right). \tag{33}$$

By setting the deviation of the above objective function w.r.t $\mathbf{C}_{\mathrm{TC}}$ to be 0, the following equation can be obtained:

$$\mu\mathbf{D}_{\mathrm{TC}}^{T}\left(\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{R}\right)+\mu_{\mathbf{C}_{\mathrm{TC}}}\left(\mathbf{C}_{\mathrm{TC}}-\mathbf{L}+\frac{\mathbf{Z}}{\mu_{\mathbf{C}_{\mathrm{TC}}}}\right) = 0. \tag{34}$$

Algorithm 1: Solving the Subproblem (28) for $\mathbf{C}_{\mathrm{TC}}$.

Input: $\mathbf{X}$, $\mathbf{D}_{\mathrm{B}}$, $\mathbf{D}_{\mathrm{TC}}$, $\mathbf{C}_{\mathrm{B}}$, $\mu_{\max-\boldsymbol{C}_{\mathrm{TC}}} = 10^8$, $\rho_{\boldsymbol{C}_{\mathrm{TC}}} = 1.15$, $\mu_{\boldsymbol{C}_{\mathrm{TC}}} = 10^{-3}$, $\varepsilon_{\boldsymbol{C}_{\mathrm{TC}}} = 10^{-6}$.
Initialize: $\mathbf{Z} = \mathbf{L} = 0$.
While not converged **do**
1. Update $\mathbf{L}$ and $\mathbf{C}_{\mathrm{TC}}$ as in (30) and (35).
2. Update the Lagrange multipliers as $\mathbf{Z} := \mathbf{Z}+\mu_{\boldsymbol{C}_{\mathrm{TC}}}(\mathbf{C}_{\mathrm{TC}}-\mathbf{L})$.
3. Update the penalty parameter as $\mu_{\boldsymbol{C}_{\mathrm{TC}}} := \min(\rho_{\boldsymbol{C}_{\mathrm{TC}}}\mu_{\boldsymbol{C}_{\mathrm{TC}}}, \mu_{\max-\boldsymbol{C}_{\mathrm{TC}}})$.
4. Check the convergence condition $\mathbf{C}_{\mathrm{TC}}-\mathbf{L}_{\infty} < \varepsilon_{\boldsymbol{C}_{\mathrm{TC}}}$.
End while
Output: $\mathbf{C}_{\mathrm{TC}}$.

The solution for the above problem is

$$\mathbf{C}_{\mathrm{TC}} = \left(\mathbf{D}_{\mathrm{TC}}^{T}\mathbf{D}_{\mathrm{TC}}+\mu_{\mathbf{C}_{\mathrm{T}}}\mathbf{I}\right)^{-1}\left(\mathbf{D}_{\mathrm{TC}}^{T}\mathbf{R}+\frac{\mu_{\mathbf{C}_{\mathrm{T}}}}{\mu}\mathbf{L}-\frac{\mathbf{z}}{\mu}\right). \tag{35}$$

The detailed procedure for solving $\mathbf{C}_{\mathrm{TC}}$ is summarized in Algorithm 1.

Step 4 (Updating $\boldsymbol{D}_{\mathrm{B}}$): Fix the other variables and solve the following problem:

$$\min_{\mathbf{D}_{\mathrm{B}}} \frac{\mu}{2}\left(\begin{matrix}\|\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}+\left(\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{X}+\mathbf{E}-\frac{\mathbf{Y}_1}{\mu}\right)\|_F^2\\ +\|\mathbf{D}_{\mathrm{B}}+\left(\frac{\mathbf{Y}_2}{\mu}-\mathbf{J}\right)\|_F^2\end{matrix}\right)+\eta'\|\mathbf{D}_{\mathrm{B}}\|_F^2. \tag{36}$$

Let $\mathbf{M} = \mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-\mathbf{X}+\mathbf{E}-\frac{\mathbf{Y}_1}{\mu}$, $\mathbf{N} = \frac{\mathbf{Y}_2}{\mu}-\mathbf{J}$, and the simplified formula is

$$\min_{\mathbf{D}_{\mathrm{B}}} \frac{\mu}{2}\left(\|\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}+\mathbf{M}\|_F^2+\|\mathbf{D}_{\mathrm{B}}+\mathbf{N}\|_F^2\right)+\eta'\|\mathbf{D}_{\mathrm{B}}\|_F^2. \tag{37}$$

Let the deviation of the above objective function w.r.t $\mathbf{D}_{\mathrm{B}}$ be 0, and the solution can be gotten as follows:

$$\mathbf{D}_{\mathrm{B}} = -\left(\mu\mathbf{M}\mathbf{C}_{\mathrm{B}}^{T}+\mu\mathbf{N}\right)\left(\mu\mathbf{C}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}^{T}+\mu\mathbf{I}+2\eta'\mathbf{I}\right)^{-1}. \tag{38}$$

Step 6 (Updating $\mathbf{D}_{\mathrm{TC}}$): Fix the other variables and solve the following problem:

$$\min_{\mathbf{D}_{\mathrm{TC}}} \frac{\varphi}{2}\|\mathbf{D}_{\mathrm{TC}}-\mathbf{S}\|_F^2+\frac{\mu}{2}\left(\left\|\begin{matrix}\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}-(\mathbf{X}-\\ \mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}-\mathbf{E}+\frac{\mathbf{Y}_1}{\mu})\end{matrix}\right\|_F^2\right). \tag{39}$$

Letting $\mathbf{q} = \mathbf{X}-\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}-\mathbf{E}+\frac{\mathbf{Y}_1}{\mu}$, we have

$$\varphi\left(\mathbf{D}_{\mathrm{TC}}-\mathbf{S}\right)+\mu\left(\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}^{T}-\mathbf{q}\mathbf{C}_{\mathrm{TC}}^{T}\right) = \mathbf{0}. \tag{40}$$

The solution can be accordingly obtained as in the following equation:

$$\mathbf{D}_{\mathrm{TC}} = \left(\mu\mathbf{q}\mathbf{C}_{\mathrm{TC}}^{T}+\varphi\mathbf{S}\right)\left(\varphi\mathbf{I}+\mu\mathbf{C}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}^{T}\right)^{-1}. \tag{41}$$

Step 6 (Updating $\boldsymbol{E}$): Solve following model with others fixed:

$$\min_{E} \gamma\|\mathbf{E}\|_F^2+\frac{\mu}{2}\left(\left\|\mathbf{E}-\left(\mathbf{X}-\mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}}-\mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}}+\frac{\mathbf{Y}_1}{\mu}\right)\right\|_F^2\right). \tag{42}$$

Algorithm 2: Solving the SIBTD Model (19).

Input: $\mathbf{X}$, $\mathbf{S}$, $\mu_{\max} = 10^8$, $\rho = 1.2$, $\varepsilon = 10^{-6}$, $\mu = 0.9$, α, β, γ, η, φ, and τ.
Initialize: Initialize $\mathbf{D}_{\mathrm{B}}$ and $\mathbf{D}_{\mathrm{TC}}$ using RPCA and *K*-means, $\mathbf{C}_{\mathrm{B}} = \mathbf{C}_{\mathrm{TC}} = \mathbf{E} = \mathbf{Y}_1 = \mathbf{Y}_2 = 0$.
While not converged **do**
1. Update $\mathbf{J}$, $\mathbf{C}_{\mathrm{B}}$, $\mathbf{C}_{\mathrm{TC}}$, $\mathbf{D}_{\mathrm{B}}$, $\mathbf{D}_{\mathrm{TC}}$ and $\mathbf{E}$ via (21), (25), **Algorithm 1**, (38), (41), and (44), respectively.
2. Update the Lagrange multipliers and penalty parameter as in (45).
3. Check the convergence conditions 1 and 2 as follows. $X - \mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}} - \mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}} - E_\infty < \varepsilon$, $\|\mathbf{D}_{\mathrm{B}} - \mathbf{J}\|_\infty < \varepsilon$
End while
Output: $\mathbf{D}_{\mathrm{B}}$, $\mathbf{C}_{\mathrm{B}}$, $\mathbf{D}_{\mathrm{TC}}$, $\mathbf{C}_{\mathrm{TC}}$.

Set the derivation w.r.t $\mathbf{E}$ to be 0

$$2\gamma\mathbf{E} + \mu\left(\mathbf{E} - \left(\mathbf{X} - \mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}} - \mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}} + \frac{\mathbf{Y}_1}{\mu}\right)\right) = 0. \tag{43}$$

The solution for $\mathbf{E}$ can be gotten as follows:

$$\mathbf{E} = \frac{\mu}{2\gamma + \mu}\left(\mathbf{X} - \mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}} - \mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}} + \frac{\mathbf{Y}_1}{\mu}\right). \tag{44}$$

Update the multipliers and penalty parameter as follows:

$$\begin{cases} \mathbf{Y}_1 := \mathbf{Y}_1 + \mu\left(\mathbf{X} - \mathbf{D}_{\mathrm{B}}\mathbf{C}_{\mathrm{B}} - \mathbf{D}_{\mathrm{TC}}\mathbf{C}_{\mathrm{TC}} - \mathbf{E}\right) \\ \mathbf{Y}_2 := \mathbf{Y}_2 + \mu\left(\mathbf{D}_{\mathrm{B}} - \mathbf{J}\right) \\ \mu := \min\left(\mu_{\max}, \rho\mu\right) \end{cases}. \tag{45}$$

The SIBTD model can be efficiently solved by iteratively running the above steps until the convergence conditions are satisfied. Generally, the complete optimization process is outlined in Algorithm 2. One can choose to separate the HSI data using RPCA, and the obtained results are clustered via *K*-means algorithm to initialize $\mathbf{D}_{\mathrm{B}}$ and $\mathbf{D}_{\mathrm{TC}}$.

C. Target Detection Based on SIBTD

Once the optimal solutions are obtained, the final augmented and discrimination enhanced target dictionary is gotten as $\mathbf{D}_{\mathrm{T}} = [\mathbf{D}_{\mathrm{TC}}, \mathbf{S}]$, which can be utilized as the universal target dictionary for target detection by combining with background dictionary and different representation strategies. Recent works have shown that better detection performance can be achieved by fully exploring the difference prior knowledge between target and background dictionaries [18], [19], [27], [33]. To this end, this article adopts a classwise strategy for discriminative representation learning and the final target detection. Specifically, the input test pixel is locally represented by the background dictionary, and collaboratively represented by the target dictionary

$$\min_{\boldsymbol{c}_{\mathrm{B}}} \|\mathbf{x} - \mathbf{D}_{\mathrm{B}}\boldsymbol{c}_{\mathrm{B}}\|_2^2 + \lambda_1\|\mathbf{W}\boldsymbol{c}_{\mathrm{B}}\|_2^2 \tag{46}$$

where the weight matrix $\mathbf{W}$ is a diagonal matrix, whose elements are acquired by the following way:

$$\mathbf{W} = \begin{bmatrix} \|\mathbf{x} - \mathbf{d}_{\mathrm{B}_1}\|_2 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \|\mathbf{x} - \mathbf{d}_{\mathrm{B}_n}\|_2 \end{bmatrix} \tag{47}$$

where $\mathbf{d}_{\mathrm{B}_n}$ means the *n*th atom in the background dictionary $\mathbf{D}_{\mathrm{B}}$. If $\mathbf{x}$ is far away from the *n*th atom in background dictionary, i.e., $\|\mathbf{x} - \mathbf{d}_{\mathrm{B}n}\|_2$ is big, a larger penalty will be imposed on the corresponding background dictionary atom $\mathbf{d}_{Bn}$, or a smaller penalty will be assigned. This representation learning strategy can promote a local representation which is also sparse and discriminative. With the above definition, a closed form solution for (46) can be gotten as follows:

$$\boldsymbol{c}_{\mathrm{B}} = \left(\mathbf{D}_{\mathrm{B}}^T\mathbf{D}_{\mathrm{B}} + \lambda_1\mathbf{W}^T\mathbf{W}\right)^{-1}. \tag{48}$$

Besides, a collaborative representation $\boldsymbol{c}_{\mathrm{T}}$ for test pixel $\mathbf{x}$ is calculated on the target dictionary

$$\min_{\boldsymbol{c}_{\mathrm{T}}} \|\mathbf{x} - \mathbf{D}_{\mathrm{T}}\boldsymbol{c}_{\mathrm{T}}\|_2^2 + \lambda_2\|\boldsymbol{c}_{\mathrm{T}}\|_2^2. \tag{49}$$

The solution for (49) can be similarly gotten as follows:

$$\boldsymbol{c}_{\mathrm{T}} = \left(\mathbf{D}_{\mathrm{T}}^T\mathbf{D}_{\mathrm{T}} + \lambda_2\mathbf{I}\right)^{-1}. \tag{50}$$

The detection evidence is then acquired as

$$D(x) = \|\boldsymbol{x} - \mathbf{D}_{\mathrm{B}}\boldsymbol{c}_{\mathrm{B}}\|_2 - \|\boldsymbol{x} - \mathbf{D}_{\mathrm{T}}\boldsymbol{c}_{\mathrm{T}}\|_2. \tag{51}$$

If $D(\boldsymbol{x})$ is larger than a certain threshold, then the test pixel $\boldsymbol{x}$ can be claimed as target and, otherwise, as background.

IV. Experimental Verifications

A. Hyperspectral Datasets

Three real HSIs collected by two different sensors were utilized in the experiments to evaluate the performance of the proposed target detector. The first dataset was collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor [34] with a spatial resolution of 2 m and 210 spectral bands. After removing the low SNR, water absorption, and bad bands (1–4, 76, 87, 101–111, 136–153, and 198–210), 162 bands remained. The HYDICE dataset and its ground-truth information are shown in Fig. 3(c), which has 150 × 150 pixels, and the vehicles were selected as the targets to detect, which has 21 pixels.

The second and third datasets are collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) from San Diego [35] with a spatial resolution of 3.5 m. After removing the low SNR, water absorption, and bad bands (1–6, 33–35, 97, 107–113, 153–166, and 221–224), 189 bands remained. The AVIRIS I dataset and AVIRIS II dataset and their ground-truth information are shown in Fig. 3(b) and (c), which have 60 × 60 pixels and 100 × 100 pixels, respectively. For the AVIRIS I dataset, there are 14 airplanes covering 119 pixels for detection. For the AVIRIS II dataset, the airplanes with 58 pixels were selected as the targets to be detected.

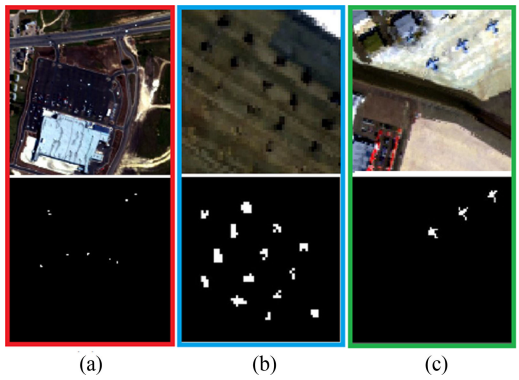

Fig. 3. HSI image scene and the ground truth. (a) HYDICE dataset. (b) AVIRIS I dataset. (c) AVIRIS II dataset.

B. Comparing Methods and Experimental Settings

To evaluate the performance of the proposed target detector, several state-of-the-art models are selected for comparison, including 1) the local ACE, 2) the local SMF, 3) the SRD, 4) SRBBHD, and 5) combined sparse and collaborative representation (CSCR) for hyperspectral target detection [19]. The source codes for the comparing methods were provided by the original author or implemented according to the detector ideas. For all the detectors, the same given target spectrum was used as the input priori target spectrum. For the detectors local ACE and local SMF, the pixels falling into dual window are used to construct the background covariance matrix. For SRD and SRBBH, the pixels falling into dual window are used to construct background dictionary and the level of sparsity is set at 4 in two AVIRIS datasets and 10 in HYDICE dataset for STD. For the detector CSCR, the pixels falling into dual window were used to estimate each center pixel.

C. Evaluation Criteria

First, receiver operating characteristic (ROC) curves were employed to evaluate the detection performance. ROC curves have been widely used as a performance evaluation tool in target detection applications [37], which demonstrates the relationship between the target detection probability P_d and the false alarm probability P_f. The definitions for P_d and P_f are formulated as follows:

$$P_d = \frac{N_{\text{detected}}}{N_t} \tag{52}$$

$$P_f = \frac{N_{\text{miss}}}{N_{\text{all}}} \tag{53}$$

where N_{detected} refers to the number of detected target pixels under a certain threshold, and N_{miss} means the number of background pixels misjudged as targets. N_t represents the number of real target pixels in the image scene, and N_{all} means all pixels in the image scene.

Limited by the dual concentric window, the marginal pixels at the edge of the image are not processed during the detection. As a result, these pixels are excluded when using ROC curves to evaluate the detection performance. The detector with the higher detection rate under the same level of false alarm rate or the same detection rate under the lower level false alarm rate has the better detection performance. However, ROC curves are often used to evaluate the detection performances of more than one method, which usually makes it difficult to judge which method performs better as they perform almost the same. Therefore, the area under ROC curves is employed, and the larger AUC value the detector achieves, the better detection performance it achieves [36], [37].

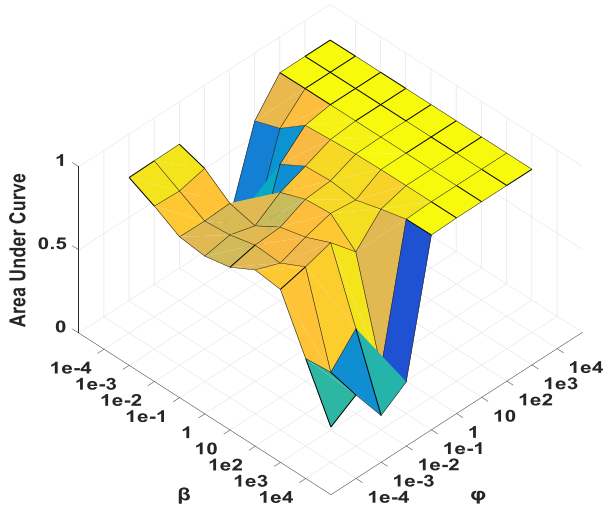

Fig. 4. AUC performance of SIBTD with different β and φ on AVIRIS I dataset.

D. Experimental Results and Analysis

1) Parameter and Convergence Analysis: This section will discuss several key parameters in the proposed HSI target detection method, including the α, β, γ, η, and φ in target and background dictionaries learning model as well as parameters λ_1 and λ_2 in learning target detection-oriented representation. The AVIRIS I dataset is utilized for example, and the parameter analysis scheme is as follows. All the parameters are tuned in three groups. Specifically, β and φ are first tuned with all the other parameters fixed, and then α and η are tuned with the rest parameters fixed. Parameter γ is separately tuned by fixing α, β, η, and φ. For all the experiments, AUC is utilized to evaluate the detection performance, and the results are shown in Figs. 4–7. As introduced in Section III, the number of background dictionary atoms is tunable, and the ratio between the number of background dictionary atoms and the number of target dictionary atoms is denoted as τ. The detection performance of SIBTD under different τ is studied and shown in Fig. 8.

From the above experimental results, a parameter setting range with $\alpha \in [10^{-4}, 1]$, $\beta \in [10^{-1}, 10^{2}]$, $\gamma \in [1, 10^{3}]$, $\eta \in [10^{-4}, 10]$, $\varphi \in [10^{2}, 10^{4}]$, $\tau \in [4, 8]$, $\lambda_1 \in [10^{-4}, 10^{-2}]$, and $\lambda_2 \in [10^{-4}, 10^{-2}]$ were suggested. In this parameter space, a good and stable performance can be expected. Besides, an optimization algorithm has been carefully devised to efficiently solve the developed SIBTD model. To show this, the convergence curves for the two convergence conditions in Algorithm 2 on AVIRIS II dataset were plotted in Fig. 9, which show that

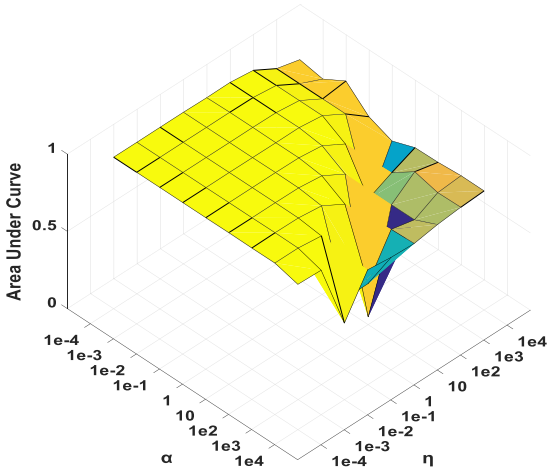

Fig. 5. AUC performance of SIBTD with different α and η on AVIRIS I dataset.

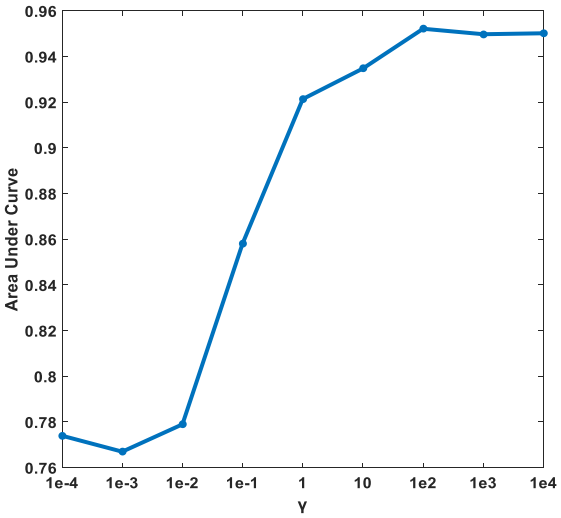

Fig. 6. AUC of SIBTD on AVIRIS I dataset with different γ.

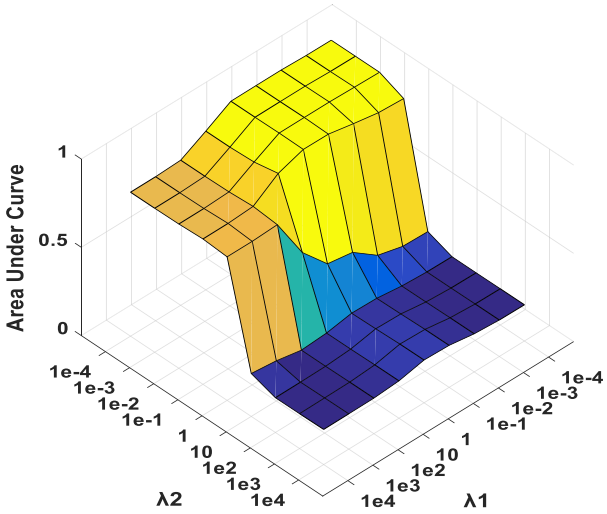

Fig. 7. AUC performance of SIBTD with varying λ_1 and λ_2 on AVIRIS I dataset.

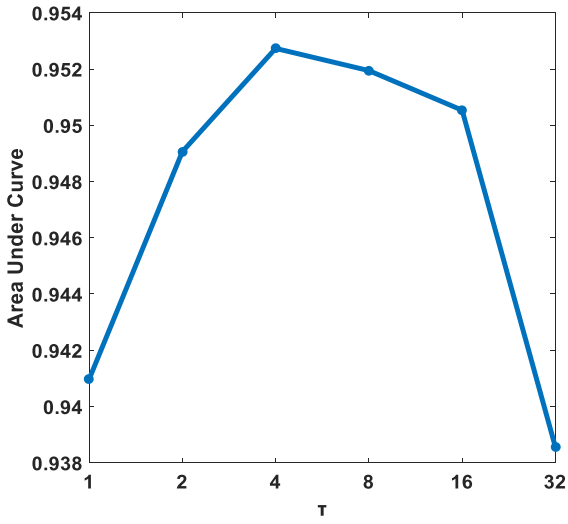

Fig. 8. AUC of SIBTD on AVIRIS I dataset with target and background dictionaries under different τ.

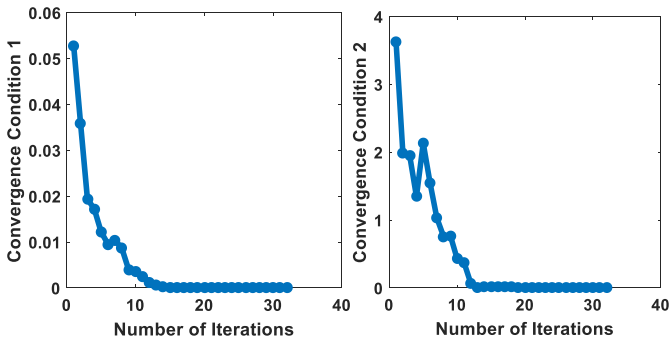

Fig. 9. Convergence curves of Algorithm 2 on AVIRIS II dataset, and an AUC of 0.9972 was achieved in this example.

Algorithm 2 performs well in practice and can quickly converge about 30 iterations. The experimental results also show that the devised optimization algorithm has good converge property, leading to promising model performance.

2) Detection Performance: This section will evaluate the detection performance of SIBTD model in comparison with several state-of-the-art detectors, i.e., ACE, SMF, STD, SRBBH, and CSCR. For the benchmark HSI datasets, according to previous experiments on examining the effect of dual-window size on the detection performance, the most commonly adopted dual-window $(w_{\text{in}}, w_{\text{out}})$ combinations were searched within {(5, 13), (5, 17), (7, 13), (7, 17)} for the comparing detectors to get the best detection performance. The AUC performances of different detectors on different datasets are listed in Tables I–IV. From the results, one can see that the performances of the detectors adopting dual-window strategy for local background dictionary construction are sensitive to the sizes of the outer window and the inner window.

The detailed experimental results on different datasets of the proposed SIBTD in comparison with other detectors are

TABLE I
AUC PERFORMANCE OF DIFFERENT DETECTORS ON THE HYDICE DATASET WITH DIFFERENT WINDOW SIZES

Detectors	w_{in} \ w_{out}	13	17
ACE	5	0.9140	**0.9531**
	7	0.9476	0.9380
SMF	5	0.9087	0.8995
	7	**0.9382**	0.9139
STD	5	**0.9844**	0.9250
	7	0.9843	0.9625
SRBBH	5	0.9138	0.7797
	7	**0.9376**	0.8137
CSCR	5	0.9535	0.9405
	7	**0.9556**	0.9408

TABLE II
AUC PERFORMANCE OF DIFFERENT DETECTORS ON THE AVIRIS I DATASET WITH DIFFERENT WINDOW SIZES

Detectors	w_{in} \ w_{out}	13	17
ACE	5	**0.7821**	0.7346
	7	0.7704	0.6688
SMF	5	0.7768	0.7063
	7	**0.7846**	0.6305
STD	5	0.9303	0.9141
	7	**0.9429**	0.9231
SRBBH	5	0.8663	0.6722
	7	**0.8775**	0.6838
CSCR	5	0.9631	0.9536
	7	**0.9679**	0.9548

TABLE III
AUC PERFORMANCE OF DIFFERENT DETECTORS ON THE AVIRIS II DATASET WITH DIFFERENT WINDOW SIZES

Detectors	w_{in} \ w_{out}	13	17
ACE	5	0.4841	0.5595
	7	**0.6426**	0.6219
SMF	5	0.5684	0.6669
	7	0.7404	**0.7439**
STD	5	0.9491	0.9380
	7	0.9609	**0.9747**
SRBBH	5	0.7676	0.7625
	7	**0.7805**	0.7767
CSCR	5	**0.9951**	0.9940
	7	0.9849	0.9166

provided by ROC curves and AUC values, as shown in Fig. 10. As mentioned before, the x- and y-axes in ROC curves represent the false alarm rate and the detection probability. The closer a curve gets to the upper-left side, the better performance the corresponding method achieves. The results show that the proposed SIBTD can achieve better AUC performance.

3) Analysis for the Structural Incoherence Constraint: One key idea for the developed SIBTD model is to directly learn a pure and unified background dictionary and a target compensation dictionary with the supervision of known target spectrum. The background dictionary is encouraged to be as incoherent as possible from the known target spectrum, and a target compensation dictionary is simultaneously promoted to be similar with the known target spectrum. To intuitively show the learning performance, a 2-D principal component subspace is derived, and the target dictionary and known target spectrum are projected into this subspace for visualization using t-SNE [51]. As shown in Fig. 11, the background and target dictionary atoms derived from the classical dual-window strategy are mixed together without clear boundary, which might reduce the discrimination ability of the dictionary atoms, and thus the detection performance is restricted. In contrast, the boundary between the background and target dictionary atoms learned by our SIBTD model is clear, which can intuitively verify the good discrimination ability of the proposed dictionary construction method. From the above extensive experimental results, several discussions can be made as follows.

TABLE IV
AUC VALUES OF DIFFERENT DETECTORS ON DIFFERENT DATASETS

Detectors	Datasets		
	HYDICE	AVIRIS I	AVIRIS II
ACE	0.9531	0.7799	0.6426
SMF	0.9384	0.7881	0.7439
STD	0.9848	0.9398	0.9747
SRBBH	0.9372	0.8775	0.7805
CSCR	0.9557	0.9655	0.9951
SIBTD	0.9928	0.9819	0.9981

1) Classic detectors, such as ACE and SMF, assume the target and background components following certain distribution. However, good performances were achieved on the HYDICE dataset. However, much worse performances were gotten when meeting AVIRIS I and AVIRIS II datasets. This is because the data distributions assumed in ACE and SMF cannot always hold in reality. The HSI data can be acquired using sensors with different characteristics under different imaging environments, which might bring difficulties for ACE and SMF to achieve promising detection performance in application.
2) As a universal methodology, spare representation has a more relaxed assumption that the within-class samples should reside in an identical subspace, and the samples from the identical class tends to represent each other better. From the perspective of theory and experiment, the SRD methods can generally achieve better performance than ACE and SMF. As a preceding work, STD applies SR theory to learn a uni-structure SR on the union of target and background dictionaries. The SR is discriminative and can reflect the membership of the test pixel. However, the different roles of background and target dictionaries in representing the test pixel are not well utilized, which might be unfavorable to the detection performance. This drawback can be alleviated by fully exploring the difference prior knowledge as in SRBBH and CSCR. Both of them consider the characteristic differences between the background and target dictionaries and adopt a re-representation strategy for test pixel representation using target dictionary, background dictionary, or a union of them via sparse or collaborative representation methods. The key strategy can guide SRBBH and CSCR to achieve

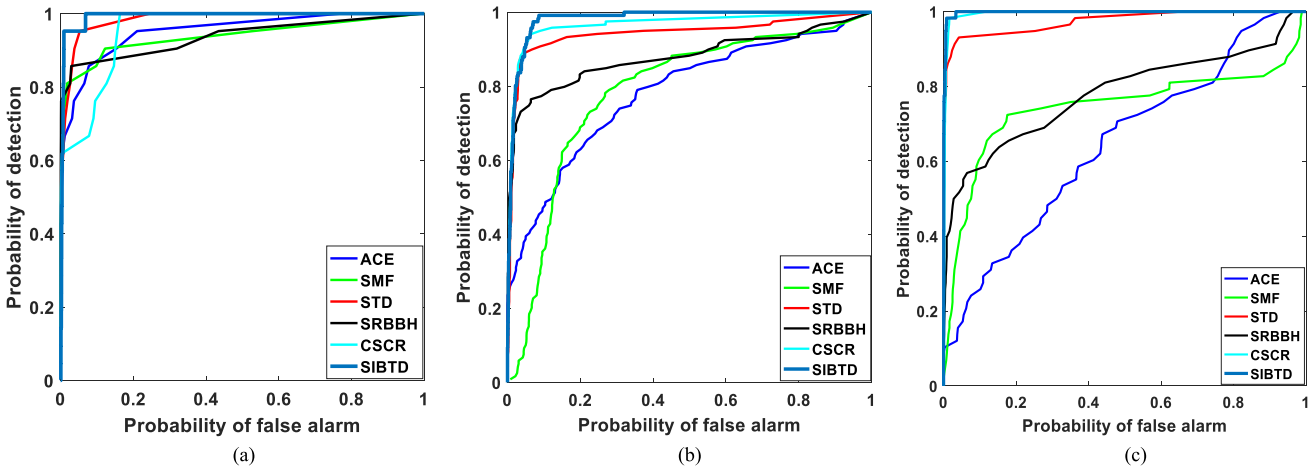

Fig. 10. ROC performances of all detectors for three datasets. (a) HYDICE. (b) AVIRIS I. (c) AVIRIS II.

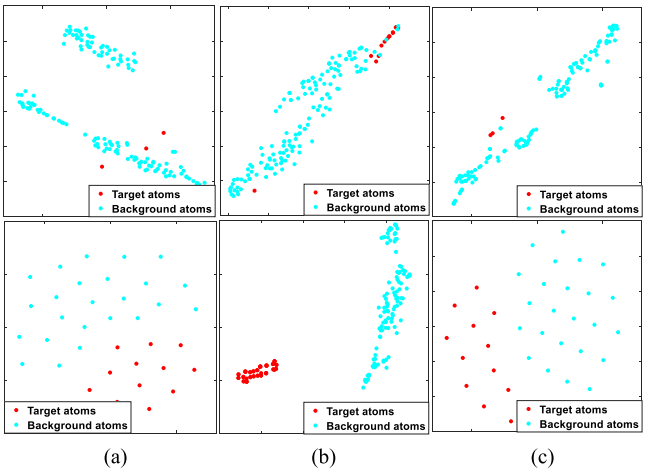

Fig. 11. Visualization of the background and target dictionaries of different datasets via t-SNE [51]. The first row is obtained by dual-window strategy, and the second row is learned by the proposed SIBTD model. (a) HYDICE dataset. (b) AVIRIS I dataset. (c) AVIRIS II dataset.

better performance. However, as the experimental results show, the dual-window method used for background and target dictionaries will lead to unstable detection performance.

3) For the limitations of existing HSI target detection methods, this article proposes to learn and construct the background and target dictionaries from the observed raw HSI data using sparse and low-rank theory. Experimental results show that the developed detector can yield promising detection performance. The benefits can be attributed to the three features of the proposed SIBTD model. First, a low-rank background component and sparely distributed target component can be decomposed from the raw HSI data supervised by the known target spectrum. The background dictionary derived from SIBTD is compact with the atom number tunable. Second, the well modeling of noise can lead to a pure background and target dictionary pair. Third, as shown by the numerical results and the visualization experiments, the introduction of incoherent constraint can help to enhance the discrimination and separability of the background and target dictionaries.

V. Conclusion

One key problem for HSI target detection is how to accurately characterize the background and target components. This article focuses on the target and background dictionaries construction problem in sparsity-based HSI target detection methods. As a supervised method, SIBTD can construct a compact and discriminative background and target dictionary pair with the supervision of known target spectrum. Besides, the devised optimization algorithm shows good convergence property and can efficiently solve the SIBTD model. The benefits and advantageous detection performance are verified on three benchmark HSI datasets in comparing with several state-of-the-art methods. Further work will consider improving the transfer ability of SIBTD, i.e., the background and target dictionaries learned in one application scenario can be used in other different but related application scenarios.

Acknowledgment

The authors would like to thank the Editor-in-Chief, Associate Editor, and anonymous reviewers for their insightful comments.

References

[1] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, 2014, Art. no. 010901.

[2] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, Nov. 2016.

[3] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 24–33, Jan. 2014.

[4] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.

[5] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Is there a best hyperspectral detection algorithm?" *Proc. SPIE*, vol. 7334, no. 1, Apr. 2009, Art. no. 733402.

[6] Q. Guo *et al.*, "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.

[7] S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "Models and methods for automated background density estimation in hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2837–2852, May 2013.

[8] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.

[9] M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 374–387, Feb. 2005.

[10] S. H. Peterson, D. A. Roberts, M. Beland, R. F. Kokaly, and S. L. Ustin, "Oil detection in the coastal marshes of Louisiana using MESMA applied to band subsets of AVIRIS data," *Remote Sens. Environ.*, vol. 159, pp. 222–231, Mar. 2015.

[11] D. Manolakis and G. S. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[12] N. M. Nasrabadi, "Regularized spectral matched filter for target recognition in hyperspectral imagery," *IEEE Signal Process. Lett.*, vol. 15, pp. 317–320, 2008.

[13] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.

[14] C.-I. Chang, "Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.

[15] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Jun. 2011.

[16] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Simultaneous joint sparsity model for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 676–680, Jul. 2011.

[17] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[18] Y. Zhang, B. Du, and L. Zhang, "A sparse representation-based binary hypothesis model for target detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1346–1354, Mar. 2015.

[19] W. Li, Q. Du, and B. Zhang, "Combined sparse and collaborative representation for hyperspectral target detection," *Pattern Recognit.*, vol. 48, no. 12, pp. 3904–3916, Dec. 2015.

[20] M. Ahron, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[21] D.-S. Phan and S. Venkatesh, "Joint learning and dictionary construction for pattern recognition," *in Proc. IEEE CVP R*, Jun. 2008, pp. 1–8.

[22] A. W. Bitar, L. Cheong, and J. Ovaelez, "Sparse and low rank matrix composition for automatic target detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5239–5251, Aug. 2019.

[23] L. Ma, C. Wang, B. Xiao, and W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," *in Proc. IEEE CVPR*, Jun. 2012, pp. 2586–2593.

[24] X. Lu, W. Zhang, and X. Li, "A hybrid sparsity and distance-based discrimination detector for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1704–1717, Dec. 2018.

[25] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 116–126, Jan. 2014.

[26] D. Zhu, B. Du, and L. Zhang, "Target dictionary construction-based sparse representation hyperspectral target detection methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1254–1264, Apr. 2019.

[27] D. Zhu, B. Du, and L. Zhang, "Binary-class collaborative representation for target detection in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1100–1104, Jul. 2019.

[28] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[29] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, 2009, Art. no. 11. [Online]. Available: https://dl.acm.org/doi/10.1145/1970392.1970395

[30] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," *in Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.

[31] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Coordinated Sci. Lab., Univ. Illinois Urbana-Champaign, Champaign, IL, USA, Tech. Rep. UILU-ENG-09-2215, 2009.

[32] G. Liu *et al.*, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[33] T. Guo *et al.*, "Target detection in hyperspectral imagery via sparse and dense hybrid representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 716–720, Apr. 2020.

[34] Y. Zhang, B. Du, and L. Zhang, "Regularization framework for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 313–317, Jan. 2014.

[35] T. Wang, B. Du, and L. Zhang, "A kernel-based target-constrained interference-minimized filter for hyperspectral sub-pixel target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 626–637, Apr. 2013.

[36] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," *in Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 233–240.

[37] X. Sun, N. Li, and H.-J. Zhao, "Performance evaluation for hyperspectral target detection algorithms," *in Proc. SPIE*, 2008, pp. 712725–712726.

[38] T. Guo, L. Zhang, X. Tan, L. Yang, and Z. Liang, "Data induced masking representation learning for face data analysis," *Knowl.-Based Syst.*, vol. 177, pp. 82–93, Aug. 2019.

[39] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jul. 2019.

[40] T. Guo *et al.* "Learning robust weighted group sparse graph for discriminant visual analysis," *Neural Process. Lett.*, vol. 49, no. 1, pp. 203–226, 2019.

[41] F. Luo, L. Zhang, X. Zhou, T. Guo, Y. Cheng, and T. Yin, "Sparse-adaptive hypergraph discriminant analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1082–1086, Jun. 2020.

[42] T. Guo, L. Zhang, and X. Tan, "Neuron pruning-based discriminative extreme learning machine for pattern classification," *Cognitive Comput.*, vol. 9, no. 4, pp. 581–595, 2017.

[43] S. Wang, L. Zhang, W. Zuo, and B. Zhang, "Class-specific reconstruction transfer learning for visual recognition across domains," *IEEE Trans. Image Process.*, vol. 29, pp. 2424–2438, 2020.

[44] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.2963848.

[45] H. Su, Z. Wu, Q. Du, and P. Du, "Hyperspectral anomaly detection using collaborative representation with outlier removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 5029–5038, Dec. 2018.

[46] H. Su, Y. Yu, Q. Du, and P. Du, "Ensemble learning for hyperspectral image classification using tangent collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3778–3790, Jun. 2020.

[47] H. Su, B. Zhao, Q. Du, and P. Du, "Kernel collaborative representation with local correlation features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1230–1241, Feb. 2019.

[48] Y. Zhang, B. Du, L. Zhang, and T. Liu, "Joint sparse representation and multitask learning for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 894–906, Feb. 2017.

[49] H. Su, B. Zhao, Q. Du, P. Du, and Z. Xue. "Multi-feature dictionary learning for collaborative representation classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2467–2484, Apr. 2018.

[50] Y. Zhang, K. Wu, B. Du, and X. Hu, "Multitask learning-based reliability analysis for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2135–2147, Jul. 2019.

[51] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

Tan Guo received the M.S. degree in signal and information processing and the Ph.D. degree in communication and information systems from Chongqing University (CQU), Chongqing, China, in 2014 and 2017, respectively.

He is currently a Faculty Member with the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications (CQUPT), Chongqing, China. His research interests include hyperspectral target detection, pattern recognition, and machine learning.

Fulin Luo (Member, IEEE) received the M.S. and Ph.D. degrees in instrument science and technology from the Chongqing University, Chongqing, China, in 2013 and 2016, respectively.

He is currently an Associate Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China. His research interests are in hyperspectral image classification, image processing, sparse representation, and manifold learning in general.

Lei Zhang (Senior Member, IEEE) received the Ph.D. degree in circuits and systems from the College of Communication Engineering, Chongqing University, Chongqing, China, in 2013.

He worked as a Postdoctoral Fellow with The Hong Kong Polytechnic University, Hong Kong, from 2013 to 2015. He is currently a Professor/Distinguished Research Fellow with Chongqing University. He has authored more than 90 scientific papers in top journals, such as *IEEE T-NNLS*, *IEEE T-IP*, *IEEE T-MM*, *IEEE TIM*, *IEEE T-SMCA*, etc., and top conferences such as *ICCV*, *AAAI*, *ACM MM*, *ACCV*, etc. His current research interests include machine learning, pattern recognition, computer vision, and intelligent systems.

Dr. Zhang serves as an Associate Editor for IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT. He was the recipient of the Best Paper Award of CCBR2017, the Outstanding Reviewer Award of many journals such as *Pattern Recognition*, *Neurocomputing*, *Information Sciences*, etc., Outstanding Doctoral Dissertation Award of Chongqing, China, in 2015, Hong Kong Scholar Award, in 2014, Academy Award for Youth Innovation, in 2013, and the New Academic Researcher Award for Doctoral Candidates from the Ministry of Education, China, in 2012.

Bob Zhang (Senior Member, IEEE) received the B.A. degree in computer science from York University, Toronto, ON, Canada, in 2006, the M.A.Sc. degree in information systems security from Concordia University, Montreal, QC, Canada, in 2007, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2011.

He was with the Center for Pattern Recognition and Machine Intelligence, University of Waterloo. He was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently an Associate Professor with the Department of Computer and Information Science, University of Macau, Taipa, Macau. His research interests focus on biometrics, pattern recognition, and image processing.

Dr. Zhang is a Technical Committee Member of the IEEE Systems, Man, and Cybernetics Society and an Editorial Board Member of the *International Journal of Information*. He is an Associate Editor for the *International Journal of Image and Graphics*.

Xiaoheng Tan received the B.E. and Ph.D. degrees in electrical engineering from Chongqing University, Chongqing, China, in 1998 and 2003, respectively.

From 2008 to 2009, he was a Visiting Scholar with The University of Queensland, Brisbane, QLD, Australia. He is currently a Professor with the School of Microelectronics and Communication Engineering, Chongqing University. His current research interests include modern communications technologies and systems, communications signal processing, pattern recognition, and machine learning.

Xiaocheng Zhou, photograph and biography not available at the time of publication.