Hyperspectral Anomaly Detection via Merging Total Variation Into Low-Rank Representation

Linwei Li, *Student Member, IEEE*, Ziyu Wu, and Bin Wang, *Senior Member, IEEE*

***Abstract*—Anomaly detection (AD) aiming to locate targets distinct from the surrounding background spectra remains a challenging task in hyperspectral applications. The methods based on low-rank decomposition utilize the inherent low-rank characteristic of hyperspectral images (HSIs), which has attracted great interest and achieved many advances in recent years. In order to fully consider the characteristics of HSIs, more appropriate constrains need to be added to the low-rank model. However, there are too many regularizations and mutual constraints between regularizers, which would result in a reduction in detection accuracy, while an increasing number of tradeoff parameters complicates parameter tuning. To address the above problems, we propose a novel method based on merging total variation into low-rank representation (MTVLRR) for hyperspectral AD in this article, using a regularizer to reflect the low-rankness and smoothness of the background component of HSIs simultaneously, which can significantly decrease the mutual influence of regularizers and the difficulty of parameter tuning. Experimental results on both simulated and real hyperspectral datasets demonstrate that the proposed MTVLRR has an excellent AD performance in terms of detection accuracy compared with other state-of-the-art methods.**

***Index Terms*—Anomaly detection (AD), hyperspectral images (HSIs), low-rank, sparsity, total variation (TV).**

I. Introduction

HYPERSPECTRAL remote sensing techniques utilize tens to hundreds of narrow-band electromagnetic waves to image objects of interest, thereby obtaining abundant spectral information of objects [1]. Hyperspectral images (HSIs) are acquired by hyperspectral imaging sensors in real scenes, which contain tens to hundreds of narrow-band spectra can be visualized as a three-dimensional (3-D) data cube, including one spectral dimension and two spatial dimensions. Because of their substantial spatial information and high spectral resolution, HSIs are applicable to determine subtle differences between similar materials and can be utilized in several fields, such as classification and target detection (TD) [2], [3], [4], [5].

Received 31 March 2024; revised 31 May 2024 and 19 July 2024; accepted 19 August 2024. Date of publication 22 August 2024; date of current version 5 September 2024. This work was supported by the National Natural Science Foundation of China under Grant 62371140 and Grant 61971141. *(Corresponding author: Bin Wang.)*

The authors are with the Key Laboratory for Information Science of Electromagnetic Waves, Fudan University, Shanghai 200433, China, and also with the Image and Intelligence Laboratory, School of Information Science and Technology, Fudan University, Shanghai 200433, China (e-mail: wangbin@fudan.edu.cn).

In addition, our code is available at: https://github.com/I3ab/MTVLRR.

Digital Object Identifier 10.1109/JSTARS.2024.3447896

Hyperspectral TD (HTD) locates targets by classifying each pixel of HSIs into background and targets. Based on whether the target has prior knowledge or not, it can be categorized into supervised TD and unsupervised anomaly detection (AD). Although HTD and hyperspectral AD (HAD) are employed in different situations, the HTD method can be converted into the HAD method [6], [7], [8]. Due to the difficulty in obtaining a priori information about the target spectra and the variability of the spectra in real scenarios, AD should be more practical in hyperspectral applications and has been widely adopted in a variety of fields, such as surveillance monitoring, environmental and mineral exploration, etc. Anomalous targets in HSIs are usually unknown, with a low probability of occurrence, and markedly different spectra from the surrounding background pixels, such as rare minerals in mineral exploration [2]. Anomaly detectability and background suppression are two key metrics for the performance of AD [9].

Over the years, many methods for HAD have been advanced [10], [11], [12], [13]. In particular, the statistical-theory-based Reed–Xiaoli global algorithm (RX) [14] has been used as the benchmark method, which presumes that the background follows the multivariate Gaussian distribution, and the Mahalanobis distance between the measured pixel and the surrounding background is adopted as the final decision rule. The RX algorithm has been further extended to many versions. The local RX [15] is the typical version that have been extensively studied. Considering the difficulty of describing complex backgrounds with a single Gaussian distribution, kernel RX [16] is proposed. Besides, there have been many methods for the improvement of the RX [17], [18], [19]. Since the hypothesis that the background follows a Gaussian distribution is difficult to meet in complex and realistic environments, the detection accuracies of statistically based methods is often limited in practice.

The sparse representation (SR) based detector (SRD) [20] for HAD does not require any other specific hypothesis about the distribution of background component and only assumes that the background dictionary can linearly represent the background component, while the abnormal pixels cannot be represented. The background joint SR [21] for HAD selects the most important atoms to represent background adaptatively. The AD method based on constrained SR [22] adds the coefficients' non-negativity constraint and the coefficients' sum-to-one constraint to SR. The methods based on SR have been successfully applied to the HAD [23], [24]. The collaborative representation (CR) based detector (CRD) [25] for HAD is similar to SRD. The difference between SRD and CRD is that the SRD is based

on an assumption that sparse combination of spatially adjacent background pixels can represent themselves and constrains the representation coefficients by ℓ_1-norm, while the CRD uses all the atoms in the dictionary and the cooperation between atoms is enhanced by ℓ_2-norm. However, CRD constructs the dictionary via a pixel-by-pixel two-window approach is unreliable. The relaxed CRD [26] for HAD can improve this problem through the use of a new nonglobal dictionary for HAD. Recently, the combination of SR and CR (SRCR) [27] for HAD has the benefits of both SRCR.

Unlike the SRD and CRD which detect anomalies pixel by pixel, the low-rank decomposition-based methods not only consider the global structure of the data but also do not require a specific hypothesis regarding the distribution of background component. Robust principle component analysis (RPCA) [28], [29], [30] presumes that background is in the same subspace and anomalies can be separated by sparse constraints. However, the background does not lie in a single subspace for complex HSIs. The low-rank representation (LRR) [31], [32] introduces a background dictionary that distributes hyperspectral data across multiple subspaces. The low-rank and sparse matrix decomposition based models of [33], [34], and [35] divide the HSIs into three parts: background, anomaly, and noise. Considering that a single distribution cannot represent complex noise, the low-rank, and sparse decomposition with mixture of Gaussian method [36] introduces a mixed noise model combined with LRR to characterize the complex distribution of HSIs and designs a Manhattan distance based global detector. To improve the background suppression performance of detectors based on low-rank and sparse representations (LRaSR), the effective anomaly space (EAS) for HAD [37] introduces independent component analysis and sparsity cardinality into LRaSR-based methods. And to improve the anomalous target detectability, the iterative spectral–spatial for HAD (ISSHAD) [38] captures spectral and spatial information through iterative optimization. Since EAS and ISSHAD cannot solve the problem of reduced interband correlation, the band sampling for HAD [39] in EAS combines band sampling [40], EAS [37], and ISSHAD [38]. Further, the methods based on the combination of LRR and CR have been proposed [41], [42], [43] to improve the performance of LRR.

Due to ignoring the meaningful potential spatial similarity between each pixel and its surrounding neighbors, these LRR-based methods may result in inaccurate separating background from anomaly. To avoid breaking the intrinsic connection of the 3-D HSIs, the tensor low-rank sparse representation (TLRSR) method [44] utilizes tensor nuclear norm to characterize the low-rank background. The enhanced tensor low-rank representation [45] divides the HSIs into background, anomaly, and noise components while preserving the intrinsic connection of the 3-D HSIs. The prior-based tensor approximation (PTA) [46] unfolds the 3-D HSIs along different dimensions, with the smoothing constraint being imposed in the spatial dimension and the low-rank constraint being imposed in the spectral dimension.

Furthermore, the low average rank with total variation (TV) regularization method for AD (LARTVAD) [47] utilizes the low-rank of tensor and TV to constrain the background of 3-D HSIs. The tensor ring (TR) decomposition with factors TV regularization model for AD (TRDFTVAD) [48] fully utilizes the smoothing property of the background in 3-D HSIs by the TR decomposition with factors TV regularization. Many AD methods have emerged that are based on tensor representations [49], [50], [51] in recent years. However, the spatial and spectral dimensions have distinct physical meanings. The graph and total variation regularized LRR [52] add TV regularity with graph regularity to the LRR to preserve the geometric structure and correlation between pixels in the data, where graph regularity can represent structural information. Other methods based on graph theory for HAD are proposed in [53] and [54]. Moreover, the local spatial constraint and TV (LSC–TV) method [55] employs superpixel segmentation and uses TV to obtain the local spatial structure of the background. However, in the above methods, too many regularizers result in increased interaction between regularizers and difficult parameter tuning.

On the other hand, the 3-D correlated TV improved RPCA (3DCTV–RPCA) [56] joints low-rank and local smoothness as a regularizer. This regularizer performs differential and low-rank operations on both spatial and spectral dimensions, and achieves good results on the denoising task of HSIs. It should be noted that although both denoising and HAD task can benefit from obtaining a clean background, a method used for the denoising task is not fully applicable to the HAD task due to the characteristic difference between anomalies and noise.

In recent years, many deep-learning-based AD algorithms have emerged, and usually the anomalies they acquire come from reconstruction errors. The autonomous HAD network [57] uses a fully convolutional autoencoder (AE) and designs an adaptive-weighted loss function. To account for the geometry of the data, the robust graph AE (RGAE) [58] detector introduces a graph constraint, and to better distinguish between anomalies and noise, RGAE introduces $\ell_{2,1}$-norm. To reduce the reconstruction of anomalous features, the guided AE-based detector [59] adds a guided module. Besides, many other deep-learning-based methods have been proposed, such as [60], [61], and [62]. However, there are several problems with deep-learning-based methods. First, their lack of physical interpretability leads to poor generalization performance, and second, due to the relatively small size of HSI datasets (compared to the size of the networks) used for training (in the published articles [57], [58], [59], [60], [61], and [62]) may result in overfitting, making their experimental results less convincing.

Inspired by the 3DCTV–RPCA, this article proposes a novel LRR-based method for HAD. Due to that HAD task is essentially different from the denoising task, the differential operation on spectral dimension will affect the anomalous degree of the pixels, resulting in the loss of anomalies, which in turn results in the decrease of detection accuracy. Further, the anomalies coming from different subspaces are substantially different from the noise. Unlike RPCA-based 3DCTV–RPCA, a novel AD method based on LRR, by merging TV into LRR (MTVLRR) to constrain the coefficient matrix, is proposed.

Although LRR and TV have been extensively studied in previous articles, exploiting these two properties simultaneously requires the inclusion of more regularizers in the optimization

objective, thereby increasing both the numbers of regularizers and the tradeoff parameters. Since the mutual constraints between the regularizers will result in a decrease in detection accuracy, and an increase in the number of tradeoff parameters will complicate parameter adjustment, the proposed MTVLRR uses only one regularizer to make full advantage of the low-rankness and smoothness of the background to learn a more accurate representation (i.e., a more accurate background component), thus helping to improve the detection accuracy for HAD. Especially, the proposed MTVLRR reduces the number of tradeoff parameters and the constraints between the regularizations while retaining the use of spectral and spatial information in HSIs. Thus, it could better reconstruct the background, and better detect anomalous targets from HSIs. Experimental results on both simulated and real hyperspectral datasets consistently demonstrate that the proposed method has an excellent AD performance in terms of detection accuracy compared with other state-of-the-art (SOTA) methods.

The main contributions of this article can be concisely summarized as follows.

1) By merging TV into the regularizer of LRR, the proposed method can fully utilize the spatial similarity between adjacent pixels, and the low-rank property of background component. Moreover, it can reduce the complexity of mutual constraints and parameter adjustment among the regularizers, which makes the model simpler and also makes it more stable and practical.
2) We utilize the alternating direction method of multiplier (ADMM) [63] algorithm to resolve the proposed model, thus constructing a new algorithm, which can improve the efficiency and accuracy of the solution. The superiority and effectiveness of the proposed MTVLRR have been consistently verified on both simulated and real hyperspectral datasets.

The rest of this article is organized as follows. The basic concepts of RPCA, LRR, and 3DCTV–RPCA are reviewed in Section II. The proposed algorithm is presented in Section III. Its effectiveness is demonstrated by the experimental results on both simulated and real datasets in Section IV. Finally, Section V concludes this article.

II. Related Works

A. RPCA and LRR

The principle component analysis (PCA) [64] presumes that the noise conforms to a Gaussian distribution, which leads to the fact that PCA will not get the desired results when the data have large noise and abnormal values. However, the low-rank decomposition-based RPCA [28], [29], [30] considers that the background component lies in a single subspace, and decomposes the original data into a sparse part that contains anomalies and noise, and a low-rank background part, which makes the principle components in the low-rank part robust to the anomalies and noise in the sparse part. RPCA can be described as

$$\min_{\boldsymbol{L},\boldsymbol{E}} \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{E}\|_0, \quad \text{s.t.}\, \boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{E} \tag{1}$$

where $\boldsymbol{Y}$ is the original data and hypothesized to be decomposable into the sparse part $\boldsymbol{E}$ and the low-rank part $\boldsymbol{L}$. $\|\boldsymbol{L}\|_*$ is the nuclear norm of $\boldsymbol{L}$, which is used to constrain the low-rank of the matrix $\boldsymbol{L}$. $\|\cdot\|_0$denotes the ℓ_0-norm of the matrix, which is nonconvex.

To facilitate obtaining the optimal solution, the convex optimization problem of RPCA is expressed as follows:

$$\min_{\boldsymbol{L},\boldsymbol{E}} \|\boldsymbol{L}\|_* + \lambda\|\boldsymbol{E}\|_1, \quad \text{s.t.}\, \boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{E}. \tag{2}$$

In HAD, $\boldsymbol{Y} \in \mathbb{R}^{B\times N}$can be considered as hyperspectral data with N pixels and each pixel has B bands. The error $\boldsymbol{E}$ represents the sparse anomaly component, and the matrix $\boldsymbol{L}$ represents the low-rank background. RPCA does not easily distinguish anomalies and noise, resulting in a high false alarm rate.

The methods based on LRR [31], [32], [33], [34], [35], [36] consider that different anomalies come from their corresponding subspaces, and thus LRR can more accurately recover the subspace structure by finding the dictionary $\boldsymbol{A}$ corresponding to a hyperspectral data $\boldsymbol{Y} \in \mathbb{R}^{B\times N}$ with N pixels and B bands. The objective function of LRR model can be written as

$$\min_{\boldsymbol{X},\,\boldsymbol{E}} \|\boldsymbol{X}\|_* + \lambda\|\boldsymbol{E}\|_{2,\,1}, \quad \text{s.t.}\, \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{E} \tag{3}$$

where $\boldsymbol{AX}$ denotes the background component, meaning that the background component consists of a background dictionary $\boldsymbol{A}$ and its corresponding matrix of coefficient $\boldsymbol{X} \in \mathbb{R}^{m\times N}$($m$ denotes the number of atoms); if $\boldsymbol{A} = \boldsymbol{I}$, the model is consistent with RPCA; and$\|\,\boldsymbol{E}\,\|_{2,\,1}$ denotes the $\ell_{2,\,1}$-norm of the anomaly matrix $\boldsymbol{E}$.

B. 3DCTV–RPCA

3DCTV–RPCA [56] is a new RPCA model that joints local smoothness and low-rankness, which utilizes ℓ_1-norm constraint on the sparse part and achieves good results on the denoising task of HSIs. The objective function of 3DCTV–RPCA can be written as

$$\min_{\boldsymbol{L},\boldsymbol{E}} \|\boldsymbol{L}\|_{\text{3DCTV}} + \lambda\|\boldsymbol{E}\|_1, \quad \text{s.t.}\, \boldsymbol{Y} = \boldsymbol{L} + \boldsymbol{E} \tag{4}$$

$$\|\boldsymbol{L}\|_{\text{3DCTV}} = \sum_{i=1}^{3} \|\nabla_i \boldsymbol{L}\|_* \tag{5}$$

where $\boldsymbol{L}$ is the background component, representing the low-rank term of the observed data; and $\boldsymbol{E}$ is the sparse matrix, representing the sparse term of the observed data. Due to that the gradient graph also has a low-rank structure, the low-rank constraint can be employed on the gradient graph. Formula (5) represents the low-rank sum of the matrices after differentiation of $\boldsymbol{L}$ in the horizontal, vertical, and spectral dimensions.

III. Proposed Method

A. Merge TV Into LRR

LRR and TV represent the most important low-rank and smoothing properties of HSIs, respectively. LRR exploits the correlation of spectral dimensions of HSIs to present the global structure of the data, and it can reduce the dimensionality of the

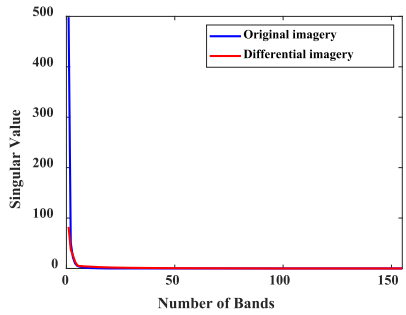

Fig. 1. Singular value curves of hyperion datasets.

data and extract potential features of the data by representing the data as low-rank structured data. The TV leverages the smoothness of the background of HSIs to achieve the dual goals of image smoothness and edge information preservation. To exploit LRR and TV simultaneously requires the inclusion of more regularizers in the optimization objective, thereby increasing both the number of regularizers and the tradeoff parameters, which leads to the interaction between regularizers, increases the difficulty of parameter tuning, and ultimately reduces the detection accuracy.

To address the above problems, the proposed MTVLRR uses a regularizer that merges TV into LRR to enable simultaneous utilization of the low-rankness and smoothness of HSIs without adding any other regularizer. The objective function of MTVLRR is written as

$$\min_{\boldsymbol{X},\boldsymbol{E}} \|\mathbf{H}\boldsymbol{X}\|_* + \lambda\|\boldsymbol{E}\|_{2,1}, \quad \text{s.t. } \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{E} \tag{6}$$

where $\|\boldsymbol{E}\|_{2,1}$ denotes the sum of ℓ_2-norm for each column of $\boldsymbol{E}$, which encourages the pixels belonging to the background in $\boldsymbol{E}$ to be zero, i.e., achieving column sparsity. λ is the only tradeoff parameter for balancing the regularizer and the error term.

Fig. 1 illustrates the singular value curves of the original imagery and the differential imagery, from which it can be seen that the original imagery and the differential imagery obtained by performing TV are all low-rank. Therefore, a low-rank constraint can be applied to the differential imagery. Based on this, a novel regularizer $\|\mathbf{H}\boldsymbol{X}\|_*$ is designed by MTVLRR term and can be calculated as follows:

$$\|\mathbf{H}\boldsymbol{X}\|_* = \left\| \begin{bmatrix} \mathbf{H}_h\boldsymbol{X} \\ \mathbf{H}_v\boldsymbol{X} \end{bmatrix} \right\|_* \tag{7}$$

where $\mathbf{H}_h$ (or $\mathbf{H}_v$) denotes the linear operation to calculate the representation coefficient difference of adjacent pixels in the horizontal (or vertical) direction, $\mathbf{H}_h\boldsymbol{X} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N]$ (or $\mathbf{H}_v\boldsymbol{X} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N]$), where $\boldsymbol{d}_i = \boldsymbol{x}_i - \boldsymbol{x}_{i_h}$ (or $\boldsymbol{v}_i = \boldsymbol{x}_i - \boldsymbol{x}_{i_v}$) and i_hth (or i_vth) pixel indicates the neighboring pixels of the ith pixel in the horizontal (or vertical) direction.

In our method, we construct the dictionary $\boldsymbol{A}$ following the approach described in [52]. This approach first classifies all pixels of HSIs into M categories. Calculate the Mahalanobis distance between each pixel and the mean vector of the current category, and then select S pixels that give the smallest results in every cluster to be the elements of $\boldsymbol{A}$.

B. Optimization Process Using ADMM

The optimal solution of the objective function (4) can be obtained by using the ADMM optimization algorithm [62].

First, introducing auxiliary variables $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, which makes (6) equivalent to

$$\begin{aligned} &\min_{\boldsymbol{X},\,\boldsymbol{E},\,\boldsymbol{P}_1,\,\boldsymbol{P}_2} \|\boldsymbol{P}_2\|_* + \lambda\|\boldsymbol{E}\|_{2,1} \\ &\text{s.t. } \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{E}, \quad \boldsymbol{P}_1 = \boldsymbol{X}, \quad \boldsymbol{P}_2 = \mathbf{H}\boldsymbol{P}_1. \end{aligned} \tag{8}$$

With the introduction of Lagrange multipliers, (8) can be further transformed as follows:

$$\begin{aligned} &\mathcal{L}(\boldsymbol{P}_1,\ \boldsymbol{P}_2,\ \boldsymbol{X},\ \boldsymbol{E},\ \boldsymbol{G}_1,\ \boldsymbol{G}_2,\ \boldsymbol{G}_3,\ \mu) \\ &= \|\boldsymbol{P}_2\|_* + \lambda\|\boldsymbol{E}\|_{2,1} + \frac{\mu}{2}\left(\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X} - \boldsymbol{E} - \boldsymbol{G}_1\|_F^2\right) \\ &\quad + \frac{\mu}{2}\left(\|\boldsymbol{P}_1 - \boldsymbol{X} - \boldsymbol{G}_2\|_F^2 + \|\boldsymbol{P}_2 - \mathbf{H}\boldsymbol{P}_1 - \boldsymbol{G}_3\|_F^2\right) \end{aligned} \tag{9}$$

where $\boldsymbol{G}_1, \boldsymbol{G}_2$, and $\boldsymbol{G}_3$ are the Lagrange multipliers, respectively, and $\mu > 0$ is the penalty parameter. Since there are multiple variables, (9) can be divided into subproblems based on the different update variables. After fixing the other variables separately, each subproblem can be optimized alternately for the final optimization solution.

In order to optimize $\boldsymbol{X}$, fixing other variables to optimize the following subproblems:

$$\begin{aligned} \boldsymbol{X}^{(v+1)} = \arg\min_{\boldsymbol{X}} &\left\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X} - \boldsymbol{E}^{(v)} - \boldsymbol{G}_1^{(v)}\right\|_F^2 \\ &+ \left\|\boldsymbol{P}_1^{(v)} - \boldsymbol{X} - \boldsymbol{G}_2^{(v)}\right\|_F^2. \end{aligned} \tag{10}$$

The solution to the subproblem about $\boldsymbol{X}$ can be described as

$$\begin{aligned} \boldsymbol{X}^{(v+1)} = &\left(\boldsymbol{A}^T\boldsymbol{A} + \boldsymbol{I}\right)^{-1}\left[\boldsymbol{A}^T\left(\boldsymbol{Y} - \boldsymbol{E}^{(v)} - \boldsymbol{G}_1^{(v)}\right)\right. \\ &\left. + \left(\boldsymbol{P}_1^{(v)} - \boldsymbol{G}_2^{(v)}\right)\right]. \end{aligned} \tag{11}$$

To optimize $\boldsymbol{P}_1$, the following equations can be considered:

$$\begin{aligned} \boldsymbol{P}_1^{(v+1)} = \arg\min_{\boldsymbol{P}_1} &\left\|\boldsymbol{P}_1 - \boldsymbol{X}^{(v+1)} - \boldsymbol{G}_2^{(v)}\right\|_F^2 \\ &+ \left\|\boldsymbol{P}_2^{(v)} - \mathbf{H}\boldsymbol{P}_1 - \boldsymbol{G}_3^{(v)}\right\|_F^2. \end{aligned} \tag{12}$$

Thus, the solution for updating $\boldsymbol{P}_1$ is (13). And it can be solved by the discrete Fourier transform diagonalization [65]

$$\begin{aligned} \boldsymbol{P}_1^{(v+1)} = &\left(\mathbf{H}^T\mathbf{H} + \boldsymbol{I}\right)^{-1} \\ &\times \left(\boldsymbol{X}^{(v+1)} + \boldsymbol{G}_2^{(v)} + \mathbf{H}^T\left(\boldsymbol{P}_2^{(v)} - \boldsymbol{G}_3^{(v)}\right)\right). \end{aligned} \tag{13}$$

To update $\boldsymbol{P}_2$, fixing other variables to optimize the subproblem about $\boldsymbol{P}_2$

$$\begin{aligned} \boldsymbol{P}_2^{(v+1)} = \arg\min_{\boldsymbol{P}_2} &\|\boldsymbol{P}_2\|_* \\ &+ \frac{\mu^{(v)}}{2}\left\|\boldsymbol{P}_2 - \mathbf{H}\boldsymbol{P}_1{}^{(v+1)} - \boldsymbol{G}_3^{(v)}\right\|_F^2. \end{aligned} \tag{14}$$

By applying the singular value thresholding (SVT) operator [66], the solution to this subproblem can be written as

$$\boldsymbol{P}_2^{(v+1)} = \Theta_{1/\mu^{(v)}}\left(\mathbf{H}\boldsymbol{P}_1{}^{(v+1)} + \boldsymbol{G}_3^{(v)}\right) \tag{15}$$

where Θ denotes the SVT operator.

To update $\boldsymbol{E}$, the subproblem of fixing other variables can be described as

$$\boldsymbol{E}^{(v+1)} = \arg\min_{\boldsymbol{E}} \lambda\|\boldsymbol{E}\|_{2,1} + \frac{\mu^{(v)}}{2}\left\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}^{(v+1)} - \boldsymbol{E} - \boldsymbol{G}_1^{(v)}\right\|_F^2. \tag{16}$$

The closed-form solution for the subproblem about $\boldsymbol{E}$ can be given as follows:

$$\boldsymbol{E}^{(v+1)} = \Omega_{\lambda/\mu^{(v)}}\left(\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}^{(v+1)} - \boldsymbol{G}_1^{(v)}\right) \tag{17}$$

where Ω denotes the $\ell_{2,1}$-minimization operator [31].

Updating $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, $\boldsymbol{G}_3$, and μ according to

$$\boldsymbol{G}_1^{(v+1)} = \boldsymbol{G}_1^{(v)} - \left(\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}^{(v+1)} - \boldsymbol{E}^{(v+1)}\right) \tag{18a}$$

$$\boldsymbol{G}_2^{(v+1)} = \boldsymbol{G}_2^{(v)} - \left(\boldsymbol{P}_1^{(v+1)} - \boldsymbol{X}^{(v+1)}\right) \tag{18b}$$

$$\boldsymbol{G}_3^{(v+1)} = \boldsymbol{G}_3^{(v)} - \left(\boldsymbol{P}_2^{(v+1)} - \mathbf{H}\boldsymbol{P}_1^{(v+1)}\right) \tag{18c}$$

$$\mu^{(v+1)} = \min\left(\rho\mu^{(v)}, \mu_{\max}\right). \tag{18d}$$

The above alternating iteration process finishes when the residuals are satisfied or the maximum number $v_{\max}$ of iterations is reached

$$\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X} - \boldsymbol{E}\|_F + \|\boldsymbol{P}_1 - \boldsymbol{X}\|_F + \|\boldsymbol{P}_2 - \mathbf{H}\boldsymbol{P}_1\|_F \le \tau. \tag{19}$$

Finally, the anomalous degree of each pixel is calculated as follows, and the proposed MTVLRR can be summarized in Algorithm 1:

$$D = \left\|[\boldsymbol{E}^*]_{:,i}\right\|_2. \tag{20}$$

Algorithm 1: Pseudocode of MTVLRR.

Input: The hyperspectral data $\boldsymbol{Y} \in \mathbb{R}^{B\times N}$, background dictionary $\boldsymbol{A}$, parameter $\lambda > 0$.

Initialize: Set the anomaly matrix $\boldsymbol{E}$, coefficient matrix $\boldsymbol{X}$, variables $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, and $\boldsymbol{G}_3$ as zero matrices, $\mu_0 = 1e-6$, $\tau = 1e-4$, $v_{\max} = 200$, $\rho = 1.5$, $\mu_{\max} = 1e+10$, and $k = 0$.

Repeat:

1) Update $\boldsymbol{X}$ with (11);
2) Update $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ with (13) and (15), respectively;
3) Update $\boldsymbol{E}$ with (17);
4) Update $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, $\boldsymbol{G}_3$, and μ with (18a)–(18d), respectively;
5) $v = v + 1$;

Until (19) is satisfied or $v > v_{\max}$.

Output: Anomaly matrix $\boldsymbol{E}$, and coefficient matrix $\boldsymbol{X}$.

IV. Experimental Results and Analysis

In this section, the detection performance of the proposed MTVLRR is evaluated on both simulated datasets and six real datasets. We compare the proposed MTVLRR with recent advanced methods, LRR–EAS [37], LSC–TV [55], PTA [46], PCA–TLRSR [44], LARTVAD [47], and TRDFTVAD [48]. Since the $\ell_{2,1}$-norm can better express the column sparse structure of the data than the ℓ_1-norm, the ℓ_1-norm in 3DCTV–RPCA [56] is modified to the $\ell_{2,1}$-norm when applied in the HAD task. To demonstrate the phenomenon that the differential in the spectral dimension would degrade the accuracy, 3DCTV–RPCA is transformed into 2DCTV–RPCA after removing the differential in the spectral dimension.

The detection accuracy of each method is measured by the 3-D receiver operating characteristic (3-D ROC) curve as described in [67] and [68], and the importance of 3-D ROC for performance evaluation is shown in [9]. Since the 3-D ROC curve described in [67] is used for TD and the 3-D ROC curve in [68] is utilized for AD, this article adopts the 3-D ROC curve from [68]. To quantitatively display the performance of each method, the area under the ROC(P_D, P_F), ROC(P_D, τ), and ROC(P_F, τ) curve, namely AUC$_{(D,\ F)}$, AUC$_{(D,\ \tau)}$, and AUC$_{(F,\ \tau)}$) are used, in which P_D and P_F are the probability of detection and false alarm, respectively. Moreover, the overall performance is measured by AUC$_{\mathrm{OADP}}$ [68], which is calculated by

$$\mathrm{AUC}_{\mathrm{OADP}} = \mathrm{AUC}_{(D,\ F)} + \mathrm{AUC}_{(D,\ \tau)} + \left(1 - \mathrm{AUC}_{(F,\ \tau)}\right). \tag{21}$$

Further, we measure anomaly detectability with background suppressibility by using the signal-to-noise-probability ratio (AUC$_{\mathrm{SNRP}}$) [68], which is defined by

$$\mathrm{AUC}_{\mathrm{SNRP}} = \frac{\mathrm{AUC}_{(D,\ \tau)}}{\mathrm{AUC}_{(F,\ \tau)}}. \tag{22}$$

A computer with Intel Core, 3.50 GHz CPU, and 32.0 GB RAM serves as a platform to implement all experiments in MATLAB R2021a.

A. Experiments on Simulated Datasets

Simulated Dataset Generation: The real dataset obtained by the AVIRIS sensor in the San Diego Airport area is utilized to generate the background of the simulated data. It has 224 bands, after discarding those affected by water vapor absorption and bands with low signal-to-noise ratio (SNR), 186 bands are left for experiments. Moreover, the simulated data are produced by cropping an area of size 100 × 100 from the original image. Fig. 2(a) displays the pseudocolor image of the original data, where the red square denotes the cropped area. We embed 16 in anomalous blocks into the simulated data with sizes of 1 × 1, 2 × 1, 1 × 2, and 2 × 2, and each size has four blocks with different mixing fractions α, which belongs to $[0.1, 0.4, 0.8, 1.0]$. We generate anomalous pixels based on a linear spectral mixture [69], and the simulated anomalous pixel $\boldsymbol{y}$ can be represented as

$$\boldsymbol{y} = \alpha \cdot \boldsymbol{t} + (1 - \alpha) \cdot \boldsymbol{b} \tag{23}$$

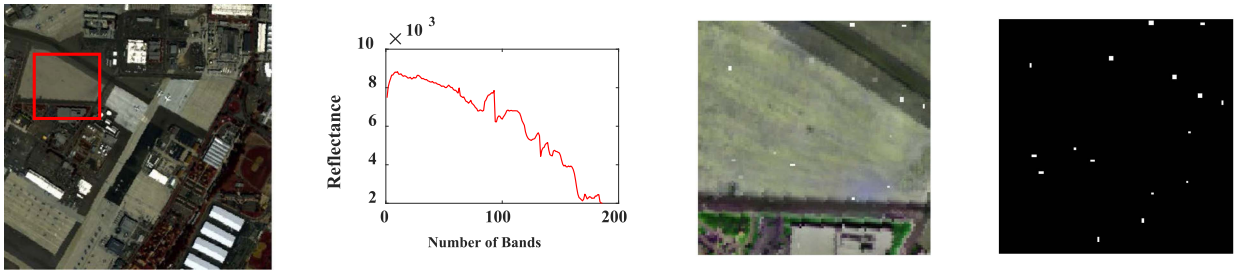

Fig. 2. Simulated dataset. (a) Pseudocolor image of the original data. (b) Anomaly spectrum. (c) Pseudocolor image of the simulated dataset. (d) Ground truth.

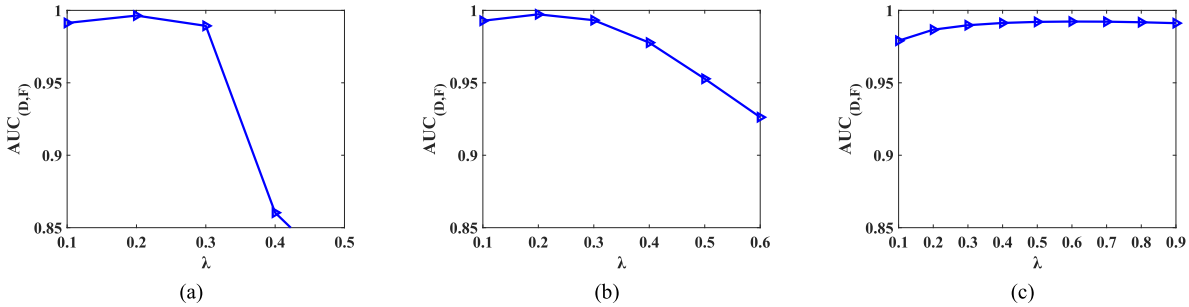

Fig. 3. Impact of the changes in parameter λ on the detection accuracy of 3DCTVS–RPCA, 2DCTV–RPCA, and MTVLRR on simulated dataset. (a) 3DCTVS–RPCA. (b) 2DCTV–RPCA. (c) MTVLRR.

where $\boldsymbol{t}$ is the anomalous spectrum, which is shown in Fig. 2(b). $\boldsymbol{b}$ denotes the background spectrum. The pseudocolor image of the simulated dataset and the corresponding ground truth are displayed in Fig. 2(c) and (d), respectively.

Parameter Experiments: The parameter experiments are conducted on the synthetic dataset to find the optimal values of the parameter λ for 3DCTV–RPCA, 2DCTV–RPCA, and MTVLRR. Since there is only one parameter λ in the above models, the optimal values can be confirmed by observing the variation of the detection accuracy with the different values of parameter λ. Fig. 3(a) and (b) shows that the accuracies of 3DCTV–RPCA and 2DCTV–RPCA change smoothly while λ in [0.1, 0.3], and therefore we set $\lambda = 0.2$ for them. From Fig. 3(c), it can be found that the detection performance of MTVLRR is insensitive to the variations in λ values when λ in [0.4, 0.8], so we set $\lambda = 0.7$ for MTVLRR. In addition to the unique tradeoff parameter λ in the optimization objective function of MTVLRR, the ADMM algorithm introduces a penalty parameter μ. As depicted in Fig. 4, the detection results of MTVLRR exhibit insensitivity to variations in μ over a relatively broad range when λ equals 0.7. The parameters of the comparison methods are all set in accordance with the recommended settings in the relevant references [37], [44], [46], [47], [48], [56].

In comparison with other methods, the proposed method has only one tradeoff parameter λ and its value is determined and selected only by performing parameter adjustment experiments on the simulated dataset, and then adopted for experiments on all other datasets (also including all real datasets).

Detection Performance: Fig. 5 shows the ROC curves of each method for the simulated dataset, from which it can be

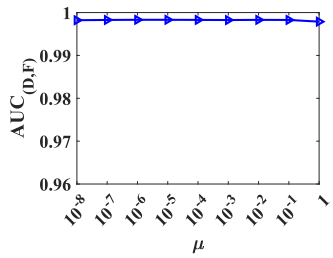

Fig. 4. Impact of the changes in parameter μ on the detection accuracy of MTVLRR when $\lambda = 0.7$.

observed that MTVLRR possesses superior background suppression and detectability, and the detection rate of MTVLRR is highest among all the comparison methods when the false alarm rate is greater than 10^{-3}. Table I quantitatively presents the experimental results on the simulated dataset, indicating that LRR–EAS has the shortest runtime, but the MTVLRR method outperforms all the comparative methods in detection accuracy, albeit with a slightly longer runtime.

Noise Robustness Experiment: To investigate the robustness of MTVLRR, the experiment was repeated 20 times for random burial locations of the anomalous target spectra with different levels of zero-mean Gaussian noise. The SNR determines the noise level which is calculated by

$$\mathrm{SNR} = 10\log_{10}\frac{\mathbb{E}\left[\boldsymbol{y}^T\boldsymbol{y}\right]}{\mathbb{E}\left[\boldsymbol{e}^T\boldsymbol{e}\right]} \tag{24}$$

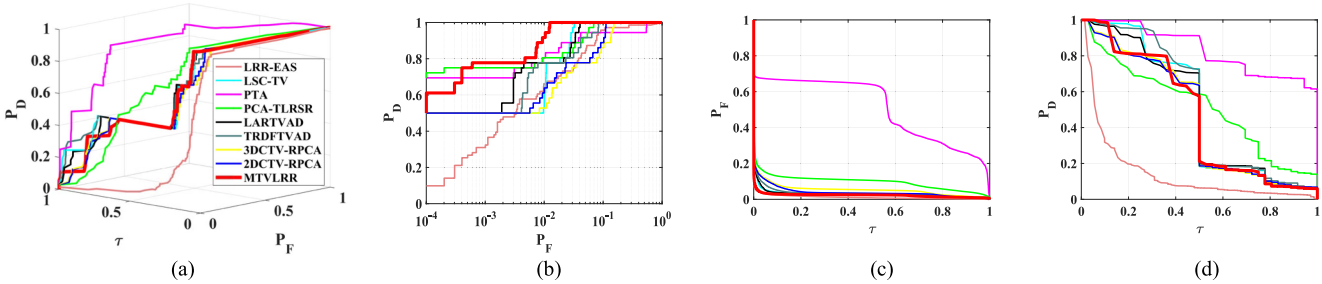

Fig. 5. ROC curves of different methods on the simulated dataset. (a) 3-D ROC curves. (b) ROC(P_D, P_F) curves. (c) ROC(P_F, τ) curves. (d) ROC(P_D, τ) curves.

TABLE I
FIVE AUC VALUES AND RUNNING TIME OF DIFFERENT METHODS ON THE SIMULATED DATASET

Methods	LRR–EAS (2022)	LSC–TV (2022)	PTA (2022)	PCA–TLRSR (2023)	LARTVAD (2023)	TRDFTVAD (2023)	3DCTV–RPCA (2023)	2DCTV–RPCA (2023)	MTVLRR (ours)
$AUC_{(D,F)}$ ↑	0.9627	0.9911	0.9656	0.9909	0.9921	0.9872	0.9633	0.9745	**0.9980**
$AUC_{(D,\tau)}$ ↑	0.1435	0.5009	**0.8350**	0.5158	0.4926	0.5181	0.4656	0.4632	0.4670
$AUC_{(F,\tau)}$ ↓	**0.0114**	0.0298	0.5064	0.0961	0.0316	0.0343	0.0529	0.0398	0.0229
AUC_{OADP} ↑	2.0948	2.4622	2.2942	2.4106	2.4531	**2.4710**	2.3760	2.3979	2.4421
AUC_{SNPR} ↑	12.5877	16.8087	1.6489	5.3673	15.5886	15.0833	8.8015	11.6382	**20.3930**
Time (s) ↓	**1.56**	356.18	16.20	80.07	15.59	39.52	13.04	15.65	80.80

Bold and underlined indicate optimal and suboptimal, respectively.

TABLE II
$AUC_{(D,\ F)}$ VALUES (AVERAGE ± STD) FOR 20 RUNS OF DIFFERENT METHODS ON THE SIMULATED DATASET WITH DIFFERENT LEVELS OF ZERO-MEAN GAUSSIAN NOISE

Methods	LRR–EAS (2022)	LSC–TV (2022)	PTA (2022)	PCA–TLRSR (2023)	LARTVAD (2023)	TRDFTVAD (2023)	3DCTV–RPCA (2023)	2DCTV–RPCA (2023)	MTVLRR (ours)
30 dB	0.9556±0.0197	0.9806±0.0112	0.9636±0.0074	0.9887±0.0094	0.9954±0.0018	0.9856±0.0032	0.9714±0.0065	0.9749±0.0020	**0.9972**±0.0014
25 dB	0.9309±0.0364	0.9509±0.0164	0.9524±0.0181	0.9595±0.0145	0.9914±0.0072	0.9845±0.0142	0.9555±0.0196	0.9706±0.0103	**0.9954**±0.0063
20 dB	0.9131±0.0699	0.9344±0.0256	0.9291±0.0268	0.9354±0.0264	0.9501±0.0315	0.9725±0.0271	0.9374±0.0295	0.9511±0.0265	**0.9783**±0.0237

Bold and underlined indicate optimal and suboptimal, respectively.

TABLE III
IN THE CASE OF WEAK TARGETS, FIVE AUC VALUES (AVERAGE ± STD) FOR 20 RUNS OF DIFFERENT METHODS ON THE SIMULATED DATASET

Methods	LRR–EAS (2022)	LSC–TV (2022)	PTA (2022)	PCA–TLRSR (2023)	LARTVAD (2023)	TRDFTVAD (2023)	3DCTV–RPCA (2023)	2DCTV–RPCA (2023)	MTVLRR (ours)
$AUC_{(D,F)}$ ↑	0.9836±0.0239	0.9607±0.0335	0.9456±0.0598	0.9738±0.0342	0.9395±0.0296	0.9584±0.0285	0.9508±0.0266	0.9626±0.0174	**0.9896**±0.0162
$AUC_{(D,\tau)}$ ↑	0.3590±0.0098	0.2319±0.0034	**0.7521**±0.0289	0.6652±0.0295	0.3266±0.0005	0.3744±0.0280	0.4068±0.0233	0.4118±0.0200	0.4030±0.0093
$AUC_{(F,\tau)}$ ↓	0.0360±0.0042	0.0484±0.0014	0.5990±0.0080	0.4984±0.0041	0.1313±0.0020	0.0590±0.0063	0.0614±0.0052	0.0415±0.0039	**0.0320**±0.0008
AUC_{OADP} ↑	2.2976±0.0168	2.0942±0.0266	2.0987±0.0886	2.2379±0.0589	2.1348±0.0011	2.2739±0.0300	2.2962±0.0323	2.3329±0.0182	**2.3606**±0.0155
AUC_{SNPR} ↑	9.9722±0.0329	4.7986±0.0126	1.2556±0.0514	1.3346±0.0423	2.4874±0.0127	6.3927±0.0875	4.7709±0.0075	5.6467±0.0048	**12.5980**±0.0025

Bold and underlined indicate optimal and suboptimal, respectively.

where $\boldsymbol{e}$ denotes the pixel's additive noise and $\mathbb{E}[\cdot]$ denotes the expectation operator.

Table II illustrates the $AUC_{(D,F)}$ values of the above repeated experiments, indicating that MTVLRR outperforms other methods with higher accuracy and more stable detection results, which could be explained by the use of LRR and TV in enhancing the distinction between the noise and the intrinsic structure of the data.

To compare the performance of all methods in the weak target case, a smaller mixing fraction α (= 0.05, 0.1, 0.2, 0.4) is adopted. The above experiment is repeated 20 times and the noise level is set to 30. Table III shows the results of all AD methods, which indicates that MTVLRR significantly outperforms the comparison method and is also more stable in terms of detection results. Compared with Table II, MTVLRR has less reduction in detection accuracy compared with other methods, which indicates the superior detection performance of MTVLRR in the case of weak anomalies.

B. Experiments on Real Hyperspectral Datasets

In this subsection, the performance of the proposed method is evaluated and validated on the six real datasets.

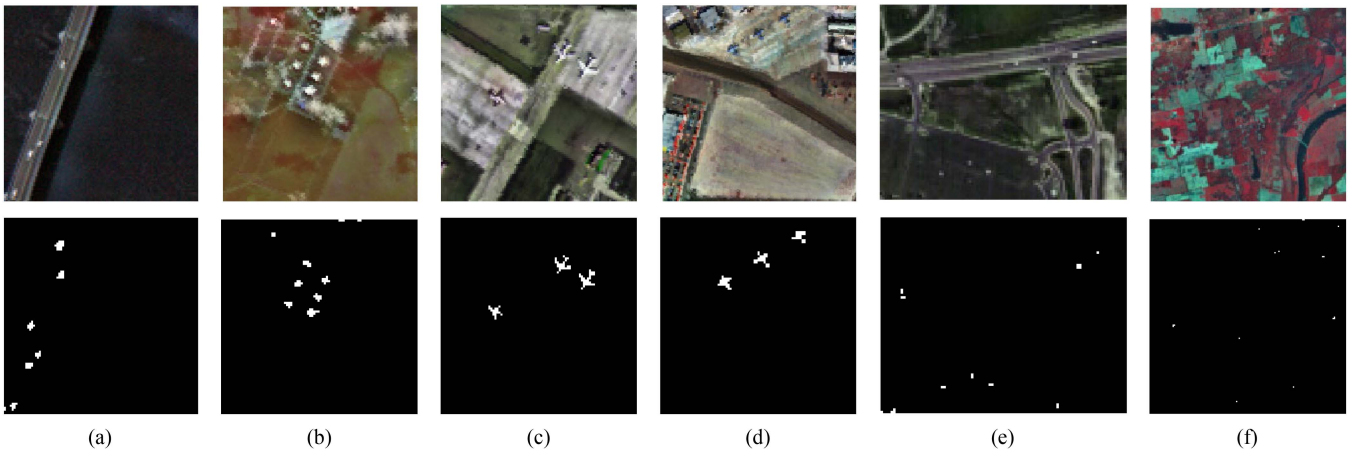

Fig. 6. Pseudocolor images of the six real datasets (first row) and ground truth (second row). (a) Pavia. (b) ABU-Urban. (c) AVIRIS-1. (d) AVIRIS-2. (e) HYDICE. (f) Hyperion.

1) Datasets Description: The first dataset was captured by ROSIS sensor, retained spectral information for 102 bands, and selected a size of 100 × 100 with 71 anomalous pixels subimages for the experiments. The second dataset called the Airport–Beach–Urban dataset (ABU-Urban) was collected by the AVIRIS sensor from the Texas Coast, TX, USA. We selected a subimage which has a size of 100 × 100 with 204 bands and 67 anomalous pixels for the experiments. The third and fourth datasets (AVIRIS-1 and AVIRIS-2) are both subimages obtained by the AVIRIS sensor from the San Diego airport. They all have a size of 100 × 100 with 186 bands. AVIRIS-1 and AVIRIS-2 contain 77 anomalous pixels and 72 anomalous pixels, respectively. The fifth dataset captured by the HYDICE sensor has a subimage of size 80 × 100 with 174 bands and 21 anomalous pixels for the experiments. The sixth dataset captured by the Hyperion sensor is 150 × 150 in size with 155 bands, containing 17 anomalous pixels. The ground truths and pseudocolor images of the above six real datasets are illustrated in Fig. 6. In [9], the importance of experimental dataset selection and the phenomenon that the performance of HAD methods can be improved in an appropriate dataset are investigated and discussed. Therefore, the selected six real datasets include anomalous targets of different sizes, allowing the performance of different detectors can be compared and evaluated more comprehensively and objectively.

2) Detection Performance: Fig. 7 shows the 2-D detection results of each method for the six real datasets, respectively, indicating that MTVLRR can suppress the background and highlight the target well on the AVIRIS-1, AVIRIS-2, Pavia, HYDICE, and Hyperion datasets. For the ABU-Urban dataset, although the MTVLRR did not suppress the background as well as LRR–EAS, the anomalies are more obvious in the MTVLRR detection map. Overall, our method can effectively suppress the background while detecting the anomalies of different sizes across the datasets.

Fig. 8 shows the 3-D ROC curves, 2-D ROC(P_D, P_F) curves, and 2-D ROC(P_F, τ) curves obtained by each method on the six datasets, respectively. From Fig. 8(a), it can be seen that the 3-D ROC of the MTVLRR is much closer to the top-right corner than other detectors, implying that MTVLRR has the best detection performance. For the AVIRIS-2, HYDICE, and Hyperion datasets, it can be observed from Fig. 8(b) that MTVLRR has the highest P_D among all the comparison methods in all P_F. For other datasets, the P_D of MTVLRR is slightly lower than that of PCA–TLRSR, and LARTVAD only in a small range because of the logarithmic abscissas, and MTVLRR has the highest P_D in most of P_F. From Fig. 8(c), it can be seen that MTVLRR has the best background suppression on all the datasets except the ABU-Urban dataset. For the ABU-Urban dataset, the background suppression of MTVLRR is weaker than that of LRR–EAS. In summary, Fig. 8 indicates that MTVLRR has superior overall detection performance on the different real datasets.

Fig. 9 shows the separation maps of background and anomaly for the different methods on all real datasets. And it is generated from the corresponding values of the background and anomalies by the different algorithms, where the boxes of background (i.e., blue boxes) and the boxes of anomaly (i.e., red boxes) are shown by 10% to 90% of the corresponding data, respectively, the height of the boxes indicates how well different algorithms suppress the background and detect the anomalies, and the rest are the deviations above and below the box, i.e., the part of dotted line. As can be seen from the Fig. 9, the overlap of the deviations between the anomaly part and the background part of all methods is inevitable, but the overlap of the proposed method is the least severe. Further, from which it can be observed that the MTVLRR completely separates the background boxes and the anomalous boxes on each dataset, and the gap between the background boxes and the anomalous boxes is relatively large, implying that the background and anomalous targets can be better distinguished. It can also be found that our method has an advantage in terms of suppressing the background. Therefore, MTVLRR demonstrates stronger background and anomaly separation capabilities on the six real datasets.

In order to quantitatively demonstrate the accuracy of each method, the five AUC values and running time of each method on the real datasets are shown in Table IV, from which it can be seen that MTVLRR achieves the highest $\mathrm{AUC}_{(D,\ F)}$ and $\mathrm{AUC}_{\mathrm{SNRP}}$ on all datasets. For the ABU-Urban dataset, although

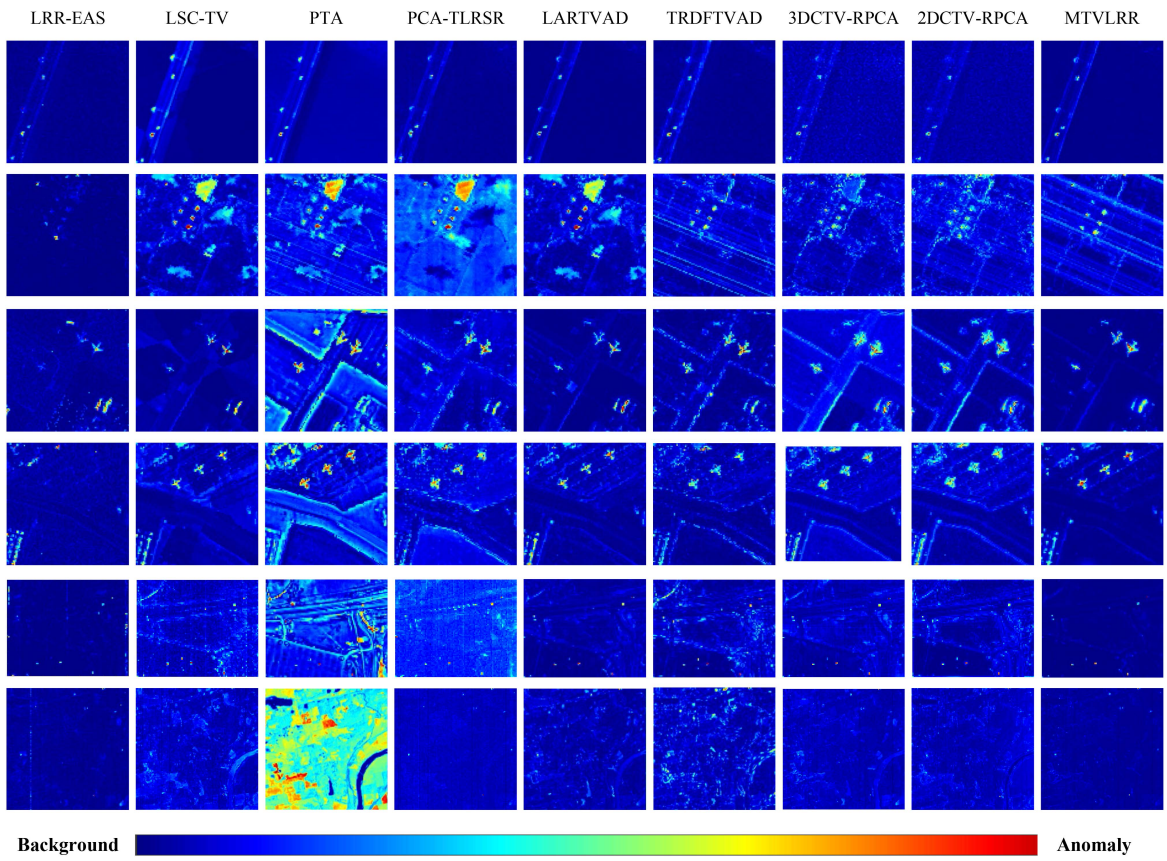

Fig. 7. Detection results of the different methods on the six real datasets. Each row from top to bottom: Pavia, ABU-urban, AVIRIS-1, AVIRIS-2, HYDICE, and Hyperion.

the $\mathrm{AUC}_{(F,\ \tau)}$ and $\mathrm{AUC}_{\mathrm{SNRP}}$ of MTVLRR are lower than those of the LRR–EAS method, the $\mathrm{AUC}_{\mathrm{OADP}}$ and $\mathrm{AUC}_{(D,\ F)}$ are higher than those of the LRR–EAS method, indicating that MTVLRR exhibits better overall performance. Although the $\mathrm{AUC}_{\mathrm{OADP}}$ of LARTVAD is higher than that of MTVLRR on the HYDICE and Hyperion datasets, its ability to suppress the background is poor, resulting in an increased false alarm rate. In terms of runtime, LRR–EAS has the shortest running time due to its simple model and computational efficiency. In summary, MTVLRR has superior AD ability with better background suppression ability. However, the running time of MTVLRR is still long, and it may be difficult for the current MTVLRR to achieve real-time detection, which should be a common problem faced by most model-based HAD algorithms implemented through iterative optimization.

Combining the qualitative and quantitative results, it can be concluded that MTVLRR has the best detection accuracy, but still requires a long runtime. Also, it can be found that the detection accuracy with TV regular term is better than without, due to that TV can characterize the segmental smoothness of HSIs and preserve the edge variations in HSIs.

C. Computational Complexity and Convergence Analysis

Computational Complexity Analysis: According to Algorithm 1, the main computational complexity of MTVLRR comes from fast Fourier transform (FFT) and SVT operator, which is the update of the auxiliary variables P_1 and P_2. Since the input data size is $m \times N$, the computational complexity of FFT and SVD operations is $\mathcal{O}(mN\log(N))$ and $\mathcal{O}(m^2N)$, respectively. During each iteration, Algorithm 1 includes one FFT operation and two SVD operations, so the overall computational complexity is $\mathcal{O}(mN(\log(N)+2m))$. Since $m > \log(N)$, the computational complexity of Algorithm 1 is $\mathcal{O}(m^2N)$.

Convergence Analysis: We verified the convergence of the proposed method on all the real datasets. From Fig. 10, it can be seen that the residuals rise at the beginning of the iteration, which may be due to that the auxiliary variables may change drastically during the initial iteration phase. However, as the number of iterations increases, the variables approach the optimal solution, which makes the residuals eventually tend to zero, where the residue is calculated by (19). Therefore, it can be concluded from Fig. 10 that MTVLRR is convergent.

D. Discussion

On both simulated and real hyperspectral datasets, extensive experiments have been conducted and the advantages of the proposed method have been demonstrated. To sum up, the proposed method has mainly four advantages as follows.

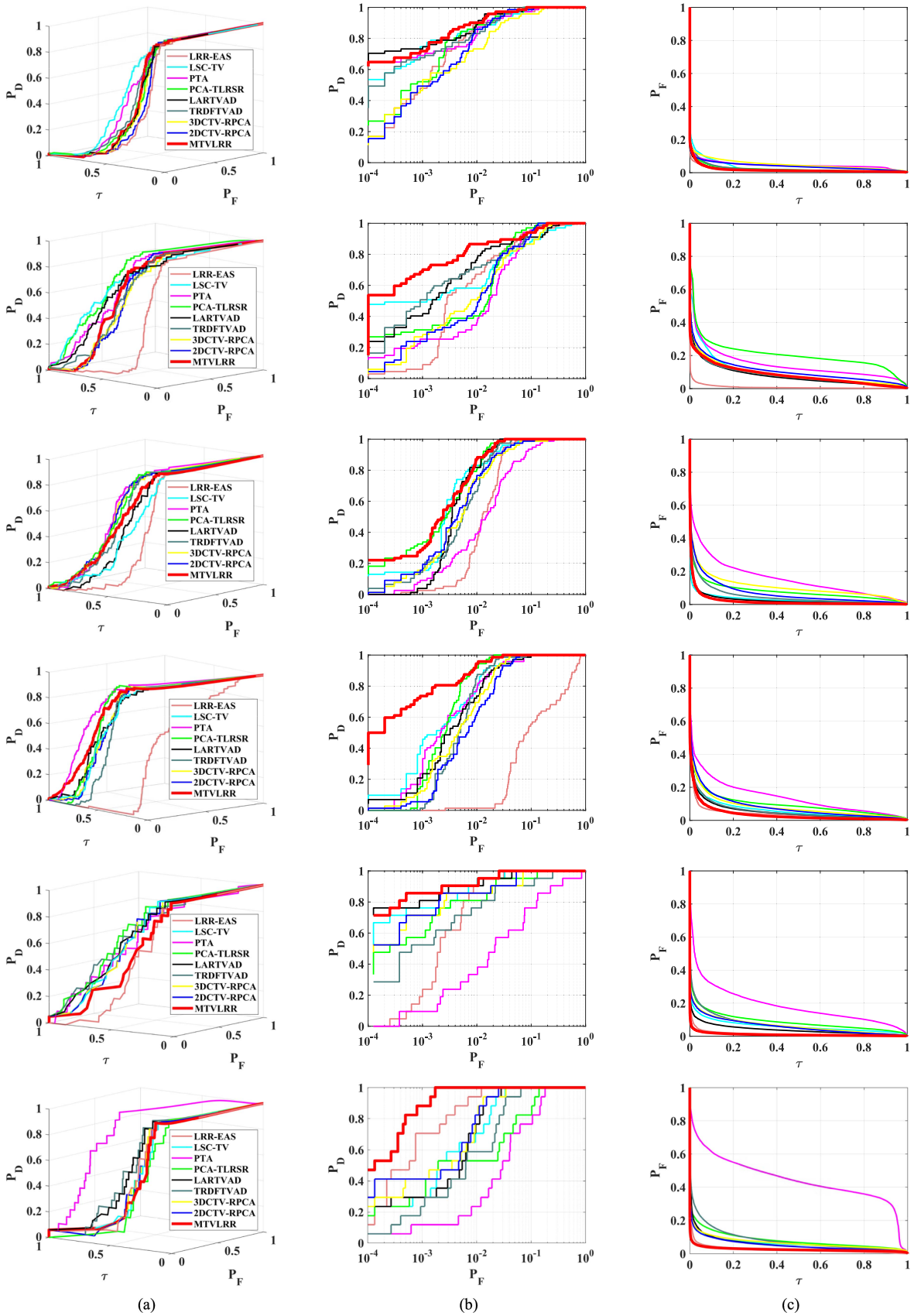

Fig. 8. ROC curves of different methods on the real datasets. (a) 3-D ROC curves. (b) ROC(P_D, P_F) curves. (c) ROC(P_F, τ) curves. Each row from top to bottom corresponds to Pavia, ABU-urban, AVIRIS-1, AVIRIS-2, HYDICE, and Hyperion datasets, respectively.

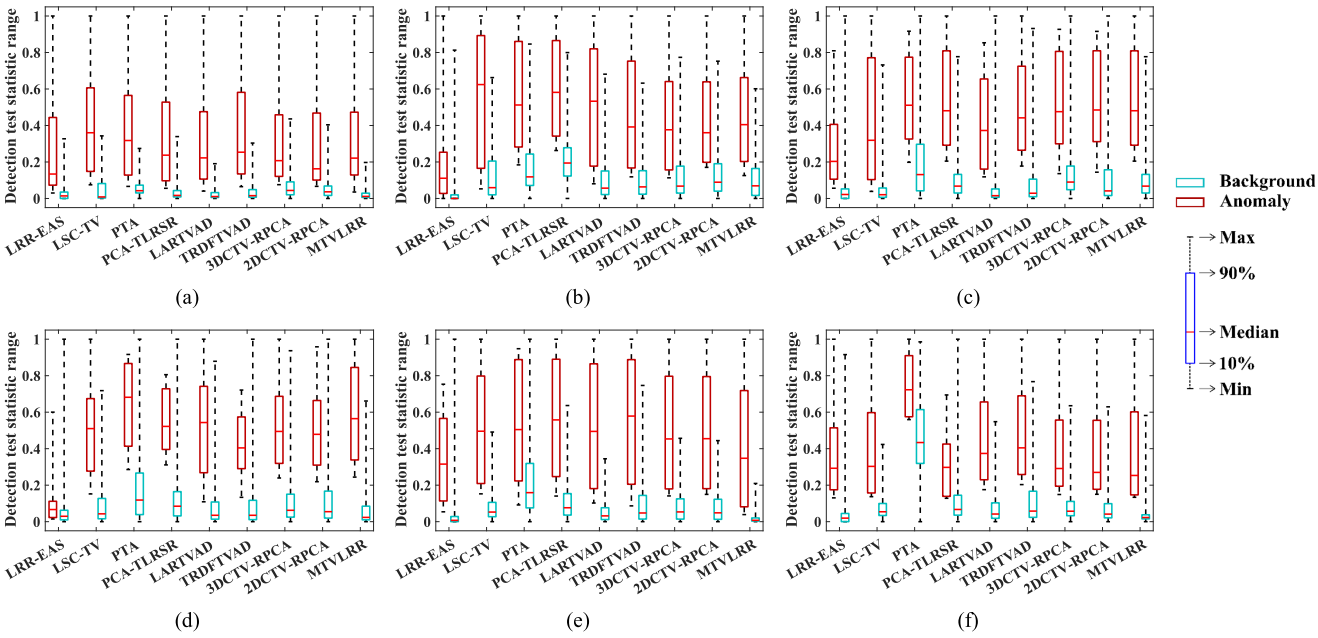

Fig. 9. Background-anomaly separation maps of the different methods on the six real datasets. (a) Pavia. (b) ABU-Urban. (c) AVIRIS-1. (d) AVIRIS-2. (e) HYDICE. (f) Hyperion.

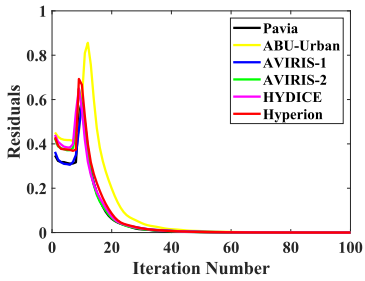

Fig. 10. Residuals versus iterations on the real datasets.

1) *Detection accuracy:* The proposed MTVLRR by MTVLRR can take full advantage of the inherent properties in HSIs, i.e., the low-rankness and smoothness of the background in HSIs. Moreover, the proposed MTVLRR explicitly reduces the mutual constraints between the regularizers. Also, the experimental results consistently shows that the proposed MTVLRR has obvious advantages over HAD in terms of detection accuracy.
2) *Convenient parameter setting:* The proposed MTVLRR has only one tradeoff parameter λ, and therefore, greatly reduces the difficulty of parameter tuning. Besides, the experiment results on simulated datasets have indicated that the detection accuracy of MTVLRR is not sensitive to the changes of λ value in a relatively wide range. And the satisfactory results were achieved although this unique parameter λ was fixed at 0.7 for all the datasets.
3) *Robustness to the noise:* Owing to that TV is introduced to suppress noise by means of the smoothness of background and LRR is utilized to distinguish the anomalies belonging to different subspaces from noise, which greatly enhances the distinction between the noise and the intrinsic structure of data, MTVLRR has strong noise robustness. Moreover, the experimental results have shown that the proposed method can obtain satisfactory detection results even in the case of high noise levels.
4) *Adaptability of anomalies:* The experimental results on both simulated and real datasets have demonstrated that our method can achieve excellent detection accuracy on different anomalies with various sizes. Further, the experimental results on both simulated and real datasets have also indicated that the proposed MTVLRR still has good detection accuracy even in the case of weak anomalies in subpixels, since the effectiveness of AD methods may mainly depend on the detection accuracy for subpixels.

However, due to the fact that the optimization process of the proposed method requires tens to hundreds of iterations, each of which undergoes one FFT operation and two SVD operations, there is a problem of relatively long runtime. This may be the drawback of most model-based HAD methods implemented through iterative optimization. In the future, we consider designing a deep network based on MTVLRR by means of a deep unfolding network approach, so as to obtain the advantages of both deep-learning and model-based methods. As is known, model-based methods utilize the physical properties (physical knowledge) of HSIs, and deep learning can adaptively learn parameters and compute them efficiently. Therefore, combining model-based methods and deep learning via the deep unfolding network approach could improve both detection time and accuracy.

TABLE IV
EXPERIMENTAL RESULTS OF DIFFERENT METHODS ON THE SIX REAL DATASETS

Datasets	AUC values	LRR–EAS (2022)	LSC–TV (2022)	PTA (2022)	PCA–TLRSR (2023)	LARTVAD (2023)	TRDFTVAD (2023)	3DCTV–RPCA (2023)	2DCTV–RPCA (2023)	MTVLRR (ours)
Pavia	$AUC_{(D,F)}\uparrow$	0.9929	0.9932	0.9927	0.9944	0.9958	0.9943	0.9865	0.9934	**0.9963**
	$AUC_{(D,\tau)}\uparrow$	0.2209	**0.3935**	0.3414	0.2843	0.2766	0.3184	0.2711	0.2433	0.2850
	$AUC_{(F,\tau)}\downarrow$	0.0180	0.0259	0.0474	0.0233	**0.0165**	0.0240	0.0507	0.0405	0.0170
	$AUC_{ODP}\uparrow$	2.1957	**2.2842**	2.2867	2.2554	2.2558	2.2888	2.2069	2.1962	2.2643
	$AUC_{SNPR}\uparrow$	12.2703	15.1544	7.2038	12.2017	16.7202	13.2827	5.3471	6.0074	**16.7647**
	Time (s) $\downarrow$	**0.91**	343.98	9.01	40.03	17.92	24.79	4.71	3.39	21.73
ABU-Urban	$AUC_{(D,F)}\uparrow$	0.9718	0.9603	0.9711	0.9803	0.9828	0.9804	0.9677	0.9760	**0.9861**
	$AUC_{(D,\tau)}\uparrow$	0.1460	0.5939	0.5428	**0.6148**	0.5655	0.4342	0.4119	0.4125	0.4461
	$AUC_{(F,\tau)}\downarrow$	**0.0090**	0.0925	0.1456	0.2041	0.0766	0.0793	0.0889	0.1056	0.0673
	$AUC_{ODP}\uparrow$	2.1088	**2.4617**	2.3684	2.3910	1.4718	2.3353	2.2907	2.2829	2.3649
	$AUC_{SNPR}\uparrow$	**16.1464**	6.4205	3.7284	3.0122	7.3852	5.4735	4.6332	3.9063	6.6286
	Time (s) $\downarrow$	**2.48**	360.71	18.00	280.17	21.28	56.85	14.96	48.59	72.07
AVIRIS-1	$AUC_{(D,F)}\uparrow$	0.9832	0.9949	0.9708	0.9909	0.9941	0.9918	0.9879	0.9900	**0.9951**
	$AUC_{(D,\tau)}\uparrow$	0.2382	0.3629	0.5324	0.5250	0.3908	0.4762	**0.5334**	0.5269	0.4718
	$AUC_{(F,\tau)}\downarrow$	0.0318	0.0286	0.1559	0.1022	0.0270	0.0494	0.1422	0.0930	**0.0190**
	$AUC_{ODP}\uparrow$	2.1896	2.3292	2.3473	2.4137	2.3579	2.4186	2.3791	2.4239	**2.4479**
	$AUC_{SNPR}\uparrow$	7.4890	12.6889	3.4147	5.1370	14.4485	9.6479	3.7511	5.6656	**24.8316**
	Time (s) $\downarrow$	**1.68**	370.45	16.65	62.72	20.12	46.72	11.51	12.56	59.63
AVIRIS-2	$AUC_{(D,F)}\uparrow$	0.7798	0.9933	0.9918	0.9961	0.9910	0.9928	0.9914	0.9881	**0.9980**
	$AUC_{(D,\tau)}\uparrow$	0.0753	0.4901	**0.6597**	0.5432	0.5195	0.4225	0.5075	0.4941	0.5915
	$AUC_{(F,\tau)}\downarrow$	0.0397	0.0621	0.1400	0.0962	0.0541	0.0550	0.0807	0.0811	**0.0353**
	$AUC_{ODP}\uparrow$	1.8155	2.4213	2.5115	2.4431	2.4564	2.3603	2.4182	2.4011	**2.5539**
	$AUC_{SNPR}\uparrow$	1.8987	7.8921	4.7128	5.6466	9.5959	7.6870	6.2887	6.0555	**16.7564**
	Time (s) $\downarrow$	**1.38**	183.84	16.15	61.97	20.20	46.91	12.24	13.78	63.00
HYDICE	$AUC_{(D,F)}\uparrow$	0.9953	0.9971	0.9011	0.9910	0.9970	0.9791	0.9919	0.9937	**0.9981**
	$AUC_{(D,\tau)}\uparrow$	0.3294	0.5078	0.5370	**0.5700**	0.5287	0.5600	0.5043	0.5054	0.4093
	$AUC_{(F,\tau)}\downarrow$	0.0145	0.0617	0.1862	0.0884	0.0396	0.0673	0.0646	0.0606	**0.0107**
	$AUC_{ODP}\uparrow$	2.3102	2.4432	1.1862	2.4726	**2.4861**	2.4719	2.4316	2.4385	2.3967
	$AUC_{SNPR}\uparrow$	22.7164	8.2301	2.8843	6.4480	13.3552	8.3211	7.8065	8.3399	**38.2523**
	Time (s) $\downarrow$	**1.34**	277.47	12.48	41.37	18.95	44.10	9.48	8.49	48.77
Hyperion	$AUC_{(D,F)}\uparrow$	0.9981	0.9923	0.9477	0.9774	0.9935	0.9854	0.9939	0.9944	**0.9996**
	$AUC_{(D,\tau)}\uparrow$	0.3348	0.3165	**0.7333**	0.3123	0.4355	0.4478	0.3512	0.3372	0.3412
	$AUC_{(F,\tau)}\downarrow$	0.0258	0.0626	0.4443	0.1270	0.0556	0.0810	0.1006	0.0728	**0.0257**
	$AUC_{ODP}\uparrow$	2.3071	2.2462	1.4443	2.1627	**2.3733**	2.3521	2.2445	2.2588	2.3151
	$AUC_{SNPR}\uparrow$	12.9767	5.0559	1.6505	2.4590	7.8323	5.5259	3.4910	4.6318	**13.2762**
	Time (s) $\downarrow$	**2.11**	1736.79	30.89	114.12	43.90	86.33	23.30	15.92	96.66

Bold and underlined indicate optimal and suboptimal, respectively.

V. CONCLUSION

In this article, we have proposed a novel method (named MTVLRR) for HAD. Taking into account the degradation in accuracy due to the mutual constraints between the regularizers and the complexity of parameter tuning, the proposed MTVLRR merges TV into LRR. In this way, the proposed method not only retains the advantages of LRR, which fully exploits the low-rank properties of HSIs without any statistical assumptions, but also employs the local spatial smoothness in HSIs without increasing the number of the constraints and the tradeoff parameters. The experimental results have shown that MTVLRR has an excellent performance in detecting anomalies while suppressing the background well, and significantly achieves more accurate detection of anomalous targets even in the case of weak anomalies, and has better robustness compared with other SOTA methods.

However, MTVLRR still has a high computational cost. In future article, we may concentrate on deeply unfolding MTVLRR as a learnable deep network that leverages the advantages of both model-based methods and deep learning to achieve more effective and efficient HAD.

REFERENCES

[1] C. Zhou, Z. He, A. Lou, and A. Plaza, "RGB-to-HSV: A frequency-spectrum unfolding network for spectral super-resolution of RGB videos," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Feb. 2024, Art. no. 5609318.

[2] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022.

[3] J. Peng et al., "Low-rank and sparse representation for hyperspectral image processing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 10–43, Mar. 2022.

[4] Y. Xu, L. Zhang, B. Du, and L. Zhang, "Hyperspectral anomaly detection based on machine learning: An overview," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3351–3364, Apr. 2022.

[5] B. Chen, L. Liu, Z. Zou, and Z. Shi, "Target detection in hyperspectral remote sensing image: Current status and challenges," *Remote Sens.*, vol. 15, no. 13, Jun. 2023, Art. no. 3223.

[6] C.-I. Chang, "Hyperspectral anomaly detection: A dual theory of hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jun. 2022, Art. no. 5511720.

[7] C.-I. Chang, "Target-to-anomaly conversion for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5540428.

[8] C.-I. Chang, "Constrained energy minimization anomaly detection for hyperspectral imagery via dummy variable trick," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5517119.

[9] C.-I. Chang, S. Chen, S. Zhong, and Y. Shi, "Exploration of data scene characterization and 3D ROC evaluation for hyperspectral anomaly detection," *Remote Sens.*, vol. 16, no. 1, 2024, Art. no. 135.

[10] B. Tu, Z. Wang, X. Yang, J. Li, and A. Plaza, "Hyperspectral anomaly detection using quantum potential clustering," *IEEE Trans. Instrum. Meas.*, vol. 71, Nov. 2022, Art. no. 5025913.

[11] C.-I. Chang, "Hyperspectral anomaly detection by data sphering and sparsity density peaks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5526321.

[12] B. Tu, X. Yang, W. He, J. Li, and A. Plaza, "Hyperspectral anomaly detection using reconstruction fusion of quaternion frequency domain analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 6, pp. 8358–8372, Jun. 2024.

[13] Q. Xiao, L. Zhao, S. Chen, and X. Li, "Hyperspectral anomaly detection via MERA decomposition and enhanced total variation regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Apr. 2024, Art. no. 5514919.

[14] 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.

[15] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.

[16] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.

[17] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "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.

[18] 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.

[19] C. Zhao, X.-F. Yao, and Y. Yan, "Modified kernel RX algorithm based on background purification and inverse-of-matrix-free calculation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 544–548, Apr. 2017.

[20] 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.

[21] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, Jun. 2015.

[22] Q. Ling, Y. Guo, Z. Lin, and W. An, "A constrained sparse representation model for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2358–2371, Apr. 2019.

[23] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection via sparse dictionary learning method of capped norm," *IEEE Access*, vol. 7, pp. 16132–16144, 2019.

[24] G. Zhao, F. Li, X. Zhang, K. Laakso, and J. C.-W. Chan, "Archetypal analysis and structured sparse representation for hyperspectral anomaly detection," *Remote Sens.*, vol. 13, no. 20, Oct. 2021, Art. no. 4102.

[25] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[26] Z. Wu et al., "Hyperspectral anomaly detection with relaxed collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2022, Art. no. 5533417.

[27] S. Lin, M. Zhang, X. Cheng, K. Zhou, S. Zhao, and H. Wang, "Hyperspectral anomaly detection via sparse representation and collaborative representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 946–961, Dec. 2023.

[28] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 1–37, May 2011.

[29] P. Netrapalli, U. N. Niranjan, S. Sanghavi, A. Anandkumar, and P. Jain, "Nonconvex robust PCA," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1107–1115.

[30] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Geosci. Remote Sens. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018.

[31] 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.

[32] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in HSIs based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.

[33] W. Sun, C. Liu, J. Li, Y. M. Lai, and W. Li, "Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery," *J. Appl. Remote Sens.*, vol. 8, no. 1, pp. 1–18, 2014.

[34] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.

[35] Y. Yang, J. Zhang, S. Song, C. Zhang, and D. Liu, "Low-rank and sparse matrix decomposition with orthogonal subspace projection-based background suppression for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1378–1382, Aug. 2020.

[36] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of Gaussian for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4363–4372, Sep. 2021.

[37] C.-I. Chang, "Effective anomaly space for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5526624.

[38] C.-I. Chang, C.-Y. Lin, P.-C. Chung, and P. F. Hu, "Iterative spectral–spatial hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 5504330.

[39] C.-I. Chang, C.-Y. Lin, and P. F. Hu, "Band sampling of hyperspectral anomaly detection in effective anomaly space," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Dec. 2024, Art. no. 5502729.

[40] C.-I. Chang, "Band sampling for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 5514024.

[41] H. Su, Z. Wu, A. X. Zhu, and Q. Du, "Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction," *ISPRS J. Photogrammetry Remote Sens.*, vol. 169, pp. 195–211, Nov. 2020.

[42] S. Feng, S. Tang, C. Zhao, and Y. Cui, "A hyperspectral anomaly detection method based on low-rank and sparse decomposition with density peak guided collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 5501513.

[43] S. Lin, M. Zhang, X. Cheng, K. Zhou, S. Zhao, and H. Wang, "Dual collaborative constraints regularized low-rank and sparse representation via robust dictionaries construction for hyperspectral anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2009–2024, Oct. 2023.

[44] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, "Learning tensor low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 679–691, Jan. 2023.

[45] Q. Xiao, L. Zhao, S. Chen, and X. Li, "Enhanced tensor low-rank representation learning for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Nov. 2023, Art. no. 5512005.

[46] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022.

[47] S. Sun, J. Liu, X. Chen, W. Li, and H. Li, "Hyperspectral anomaly detection with tensor average rank and piecewise smoothness constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8679–8692, Nov. 2023.

[48] M. Feng, W. Chen, Y. Yang, Q. Shu, H. Li, and Y. Huang, "Hyperspectral anomaly detection based on tensor ring decomposition with factors TV regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5514114.

[49] X. He, J. Wu, Q. Ling, Z. Li, Z. Lin, and S. Zhou, "Anomaly detection for hyperspectral imagery via tensor low-rank approximation with multiple subspace learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5509917.

[50] Q. Xiao, L. Zhao, S. Chen, and X. Li, "Robust tensor low-rank sparse representation with saliency prior for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Nov. 2023, Art. no. 5529920.
[51] S. Liu, C. Zhu, D. Ran, and G. Wen, "Anomaly detection via tensor multisubspace learning and nonconvex low-rank regularization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8178–8190, Sep. 2023.
[52] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jan. 2020.
[53] B. Tu, Z. Wang, H. Ouyang, X. Yang, J. Li, and A. Plaza, "Hyperspectral anomaly detection using the spectral–spatial graph," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5542814.
[54] X. Yang, B. Tu, Q. Li, J. Li, and A. Plaza, "Graph evolution-based vertex extraction for hyperspectral anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2023.3303273.
[55] R. Feng, H. Li, L. Wang, Y. Zhong, L. Zhang, and T. Zeng, "Local spatial constraint and total variation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 5512216.
[56] J. Peng, Y. Wang, H. Zhang, J. Wang, and D. Meng, "Exact decomposition of joint low rankness and local smoothness plus sparse matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5766–5781, May 2023.
[57] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5503314.
[58] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2022, Art. no. 5511314.
[59] P. Xiang, S. Ali, S. K. Jung, and H. Zhou, "Hyperspectral anomaly detection with guided autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 5538818.
[60] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Deep low-rank prior for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 5527017.
[61] L. Wang, X. Wang, A. Vizziello, and P. Gamba, "RSAAE: Residual self-attention-based autoencoder for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5510614.
[62] Z. Li, Y. Wang, C. Xiao, Q. Ling, Z. Lin, and W. An, "You only train once: Learning a general anomaly enhancement network with random masks for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Mar. 2023, Art. no. 5506718.
[63] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[64] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018.
[65] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
[66] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[67] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, Jun. 2021.
[68] C.-I. Chang, "Comprehensive analysis of receiver operating characteristic (ROC) curves for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5541124.
[69] M. S. Stefanou and J. P. Kerekes, "A method for assessing spectral image utility," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1698–1706, Jun. 2009.

Linwei Li (Student Member, IEEE) received the B.S. degree in information engineering from the Xidian University, Xi'an, China, in 2021. She is currently working toward the Ph.D. degree in electronic science and technology with the School of Information Science and Technology, Fudan University, Shanghai, China.

Her research interests include hyperspectral target detection, machine learning, and pattern recognition.

Ziyu Wu received the B.S. degree in electronic engineering from the Fudan University, Shanghai, China, in 2019. He is currently working toward the Ph.D. degree in electronic science and technology with the School of Information Science and Technology, Fudan University, Shanghai, China.

His research interests include hyperspectral target detection, machine learning, and pattern recognition.

Bin Wang (Senior Member, IEEE) received the B.S. degree in electronic engineering and the M.S. degree in communication and electronic systems from the Xidian University, Xi'an, China, in 1985 and 1988, respectively, and the Ph.D. degree in system science from the Kobe University, Kobe, Japan, in 1999.

After his graduation in 1988, he was with the Xidian University, as a Teacher. From 1999 to 2000, he was with the Communications Research Laboratory, Ministry of Posts and Telecommunications, Kobe, Japan, as a Research Fellow, working on magnetoencephalography signal processing and its application for brain science. Then, as a Senior Supervisor, he was with the Department of Etching, Tokyo Electron AT Ltd., Tokyo, Japan, from 2000 to 2002, dealing with the development of advanced process control systems for etching semiconductor equipment. Since September 2002, he has been with the Department of Electronic Engineering, Fudan University, Shanghai, China, where he is currently a Full Professor and the Leader of the Image and Intelligence Laboratory. He has authored/coauthored more than 150 scientific papers in important domestic and international periodicals. He is the holder of several patents. His research interests include multispectral/hyperspectral image analysis, automatic target/object detection and recognition, pattern recognition, signal detection and estimation, and machine learning.

Dr. Wang is an Associate Editor of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.