Citation:
C. -I. Chang, H. Cao, S. Chen, X. Shang, C. Yu and M. Song, "Orthogonal Subspace Projection-Based Go-Decomposition Approach to Finding Low-Rank and Sparsity Matrices for Hyperspectral Anomaly Detection," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 3, pp. 2403-2429, March 2021, doi: 10.1109/TGRS.2020.3002724.

DOI:
https://doi.org/10.1109/TGRS.2020.3002724

Orthogonal Subspace Projection-Based Go-Decomposition Approach to Finding Low-Rank and Sparsity Matrices for Hyperspectral Anomaly Detection

Chein-I Chang, *Life Fellow, IEEE*, Hongju Cao, Shuhan Chen, *Student Member, IEEE*, Xiaodi Shang, *Student Member, IEEE,* Chunyan Yu, *Member, IEEE*, and Meiping Song

***Abstract*— Low-rank and sparsity-matrix decomposition (LRaSMD) has received considerable interests lately. One of effective methods for LRaSMD is called go decomposition (GoDec), which finds low-rank and sparse matrices iteratively subject to the predetermined low-rank matrix order m and sparsity cardinality k. This article presents an orthogonal subspace-projection (OSP) version of GoDec to be called OSP-GoDec, which implements GoDec in an iterative process by a sequence of OSPs to find desired low-rank and sparse matrices. In order to resolve the issues of empirically determining $p = m + j$ and k, the well-known virtual dimensionality (VD) is used to estimate p in conjunction with the Kuybeda *et al.* developed minimax-singular value decomposition (MX-SVD) in the maximum orthogonal complement algorithm (MOCA) to estimate k. Consequently, LRaSMD can be realized by implementing OSP-GoDec using p and k determined by VD and MX-SVD, respectively. Its application to anomaly detection demonstrates that the proposed OSP-GoDec coupled with VD and MX-SVD performs very effectively and better than the commonly used LRaSMD-based anomaly detectors.**

***Index Terms*— Go decomposition (GoDec), low-rank and sparsity-matrix decomposition (LRaSMD), minimax-singular value decomposition (MX-SVD), orthogonal subspace projection GoDec (OSP-GoDec), Reed and Xiaoli anomaly detector (RX-AD), virtual dimensionality (VD).**

I. Introduction

LOW-RANK and sparsity-matrix decomposition (LRaSMD) has received considerable interest in hyperspectral imaging in recent years. Its idea is to model a hyperspectral image (HSI) as a linear mixture of background (BKG), sparse signal, and noise spaces, which can be characterized by low-rank, sparse, and noise matrices, respectively, referred to as LRaSMD. The idea of LRaSMD originated from [1] and [2], where a data matrix $\mathbf{X}$ could be decomposed into a sum of a low-rank component $\mathbf{L}$ and a sparse component $\mathbf{S}$ as $\mathbf{X} = \mathbf{L} + \mathbf{S}$ [3]–[14]. Unlike the classical principal component analysis (PCA), the entries in $\mathbf{S}$ can have arbitrarily large magnitudes and their supports are assumed to be sparse but unknown. In order to solve this problem, a tractable convex optimization method called principal component pursuit is used. When one or two of $\mathbf{L}$ and $\mathbf{S}$ were relaxed, a more adaptive model, $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$, was also mentioned in [1], where $\mathbf{N}$ was a dense noise and considered as small perturbation. However, this model was not studied in [1] until the work in [15], which represents the low-rank component, the gross sparse errors, and the small entrywise noise, and guaranteed the stable and accurate recovery in the presence of the componentwise noise $\mathbf{N}$ in the $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$ model. Ever since, the $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$ model was studied extensively [16]–[22] and later interpreted by the concept of compressed sensing [23]. In this case, an approach called go decomposition (GoDec) developed in [24] can be used to find the low-rank matrix $\mathbf{L}$ and the sparse matrix $\mathbf{S}$.

This article is particularly interested in investigating LRaSMD from the perspective of the linear spectral mixture model arising in hyperspectral unmixing in the sense that

Manuscript received March 21, 2020; accepted June 10, 2020. The work of Chein-I Chang was supported by the Fundamental Research Funds for Central Universities under Grant 3132019341. The work of Hongju Cao was supported by the Nature Science Foundation of Liaoning Province under Grant 20180550018. The work of Meiping Song was supported by the National Nature Science Foundation of China under Grant 61601077, Grant 61971082, and Grant 61890964. *(Corresponding author: Hongju Cao.)*

Chein-I Chang is with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information and Technology College, Dalian Maritime University, Dalian 116026, China, also with Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD21250 USA, and also with the Department of Computer Science and Information Management, Providence University, Taichung 02912, Taiwan (e-mail: cchang@umbc.edu).

Hongju Cao is with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information and Technology College, Dalian Maritime University, Dalian 116026, China, and also with the School of Software, Dalian University of Foreign Languages, Dalian 116044, China (e-mail: hongjucao1980@sina.com).

Shuhan Chen is with the Department of Electrical Engineering, Zhejiang University, Hangzhou 310000, China, and also with the Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD 21250 USA (e-mail: 11410057@zju.edu.cn).

Xiaodi Shang, Chunyan Yu, and Meiping Song are with the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Information and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: shangxiaodi@qq.com; 36202986@qq.com; 49322784@qq.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.3002724

these two LRaSMD models can find their corresponding counterparts in linear spectral unmixing [25], [26], that is, assume that $\mathbf{r}$ is a data sample vector that can be described as a linear mixture of a signature matrix $\mathbf{M}$ and its associated abundance vector $\boldsymbol{\alpha}$ as $\mathbf{r} = \mathbf{M}\boldsymbol{\alpha}$. Then, the nonnegative matrix factorization (NMF) [27] used to solve this linear mixture model is similar to the robust PCA (RPCA) in [1] that is used to solve $\mathbf{X} = \mathbf{L} + \mathbf{S}$. Furthermore, if $\mathbf{r}$ is described by $\mathbf{r} = \mathbf{M}\alpha + \mathbf{N}$ with $\mathbf{N}$ introduced as noise, then the fully constrained least squares (FCLS) developed in [28] used to solve the fully abundance constrained least-squares problem is also similar to GoDec used to solve $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$. Coincidentally, the two iterative processes used by GoDec to solve the low-rank matrix $\mathbf{L}$ and the sparse matrix $\mathbf{S}$ can also find their counterparts in FCLS that also makes use of the two iterative processes specified by the nonnegativity constrained least squares (NCLS) developed [29] and the sum-to-one constrained least squares (SCLS) [25] to perform fully constrained linear spectral unmixing.

Inspired by the striking similarities between FCLS and GoDec, this article revisits GoDec in the context of the linear mixture model used in the linear spectral unmixing to reinterpret GoDec. In particular, we rederive GoDec using the well-established approach in hyperspectral imaging, called orthogonal subspace projection (OSP) [30], to reinterpret GoDec. More specifically, the two iterative processes implemented in GoDec to find the low-rank and sparse matrices can be rederived as the two processes implemented by OSP iteratively. To be more precise, the decomposition of the low-rank matrix and the sparsity matrix can be simply carried out by a series of OSPs in an iterative manner.

However, upon implementing GoDec, two crucial parameters must be predetermined, i.e., the order of a low-rank matrix $\mathbf{L}$, rank($\mathbf{L}$), m and sparsity cardinality k of a sparse matrix $\mathbf{S}$, denoted by card($\mathbf{S}$) $= k$. Specifically, the value of m must be readjusted and updated at each iteration during the iterative processes. If we further assume that the rank of the sparse matrix is specified by j, then k can be determined by j with $k = j \times N$ according to [20], where N is the total number of data samples. Consequently, the issue in determining k is reduced to determining j. With this interpretation, the well-known virtual dimensionality (VD) [25], [26], [31]–[34] can be used to estimate p, which is the total number of spectrally distinct signatures present in an HSI that accounts for the basis vectors used to represent the low-rank matrix $\mathbf{L}$ and the sparse matrix $\mathbf{S}$, i.e., $p = m + j$. Such VD only solves half a problem. We need to solve the other half by breaking p into m and j. Fortunately, an algorithm, called minimax-singular value decomposition (MX-SVD), developed in the maximum orthogonal complement algorithm (MOCA) [35] fits exactly what we need where MX-SVD was used to estimate the rank of the rare vectors in the high-dimensional noisy signals. Such rare vectors can be considered as sparse signal sources. In this case, the rank of the sparse-signal matrix can be simply estimated to be j by MX-SVD. This implies that the rank of the low-rank matrix m can be, therefore, calculated by $m = p - j$. If we further interpret each pixel as a single signal source of the sparsity dimensionality specified by j, then the cardinality of the sparsity matrix $\mathbf{S}$, k turns out to be $k = j \times N$.

Furthermore, according to [35], MOCA used a binary hypothesis testing detection theory to estimate the number of signal sources p. In this case, the VD-estimated p can be replaced by the MOCA-estimated p, which is the sum of the number of BKG signal sources m and the number of rare signal sources j, i.e., $p = m + j$, where MX-SVD is then used to find the rare signal space separated from the BKG space with its dimensionality $j = p - m$. As also shown in [36] and [37], MX-SVD is, indeed, an OSP-based algorithm, called the automatic target-generation process (ATGP) developed in [38], which also makes use of a sequence of OSPs to find the target pixels in an exact manner that MX-SVD finds singular vectors. Consequently, the LRaSMD problem can be eventually solved by implementing OSP-GoDec in conjunction with VD and MX-SVD. Such resulting GoDec is called OSP-GoDec. Unfortunately, GoDec does not have algorithms similar to MX-SVD to solve m as well as VD/MOCA to determine p. By now, it is clear to see the primary reason why we rederive GoDec as the OSP-based GoDec (OSP-GoDec), because VD, MOCA, and MX-SVD are also the OSP-estimation techniques that are based on eigenanalysis.

In order to demonstrate and illustrate the utility of OSP-GoDec in hyperspectral imaging, anomaly detection is investigated for its potential application, where [18] is believed to be the first work applying LRaSMD to hyperspectral anomaly detection. Since then, many works on hyperspectral anomaly detection using two different LRaSMD models described previously have been reported [9]–[11] for the $\mathbf{L} + \mathbf{S}$ model such as robust PCA (RPCA) in [1] and [18]–[22] for the $\mathbf{L}$+$\mathbf{S}$+$\mathbf{N}$ model such as the LRaSMD-based anomaly detection in [20] along with a commonly used sparse-representation-based collaborative anomaly detection [39]. This article offers a rather different view from conventional LRaSMD-based approaches to develop OSP-GoDec.

II. Related Works

An earlier development of LRaSMD was derived from the video-security-surveillance system [40]–[53]. One of the major problems is how to find the motion objects or the anomalous targets quickly and precisely. For example, Cevher *et al.* [40] and Hale *et al.* [41] proposed to use compressive sensing for BKG subtraction based on the fact that BKG generally changes slowly in a short time period, which means that the number of the frames is not large. In this case, the BKG matrix should be a low-rank matrix due to its steady environment. If we further assume that the foreground is composed of a few moving objects or the anomalous targets, then the foreground matrix can be considered as a sparsity matrix. Because of that, the LRaSMD model has been widely used for this purpose. Specifically, a low-rank matrix captures the global BKG information of the videos, while a sparsity matrix captures the local information of the anomalous targets. Accordingly, many works have used the $\mathbf{L} + \mathbf{S}$ model to solve the problem of finding targets in video surveillance. The main differences among these methods are the loss functions used for optimization and solvers. Since the original

formulation can be NP-hard, these methods mainly focused on how to develop efficient relaxed models and find solvers. Vaswani *et al*. [54], [56] and Bouwmans *et al*. [55] provided reviews of these recent developments.

In hyperspectral data exploitation, LRaSMD has found its potential in many applications, such as denoising [4], [10], [12], [17], target detection [6], unmixing [5], [7], pan-sharpening [57], [58], compressive sensing and recovery [8], [40], [41], and anomaly detection [9]–[11], [18]–[22]. In these applications, two models of LRaSMD described as above, $\mathbf{X} = \mathbf{L} + \mathbf{S}$ and $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$, along with their variants are generally being used. As for $\mathbf{X} = \mathbf{L} + \mathbf{S}$, the randomized approximate matrix decomposition [58] and RPCA [1] can be used to solve the decomposition problem. Specifically, Halko *et al*. [3] shown that a matrix can be well approximated by its projection onto the column space of its random projections and provides a fast approximation of singular value decomposition (SVD)/PCA. On the other hand, Candes *et al*. [1] proved that RPCA can be used as a blind-source-separation technique to decompose the data matrix into the low-rank and sparse components as the "low rank + sparse" (i.e., $\mathbf{X} = \mathbf{L} + \mathbf{S}$) matrix decomposition. On the other hand, as for $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$, an approach called GoDec developed in [24] can be used to find the low-rank matrix $\mathbf{L}$ and the sparse matrix $\mathbf{S}$. Since then, different variants have also been developed for GoDec. For example, Zhou and Tao [59] proposed a greedy bilateral smoothing (GreBsmo) to formulate $\mathbf{L}$ with bilateral factorization and regulate the L_1 norm of the entries of $\mathbf{S}$. In addition, the l_1 regularization induces soft thresholding in updating $\mathbf{S}$, which is faster than the sorting caused by the cardinality constraint in GoDec. In [60], semisoft GoDec (SSGoDec) replaces the hard-thresholding cardinality constraint used in GoDec with a soft threshold, which significantly reduces the computational time and facilitates the parameter tuning. In order to be suitable for high-resolution or long-time video real-time processing, Chen *et al*. [61] developed an incremental low-rank and sparse decomposition that can effectively handle large-scale video sequences without much performance loss.

Although GoDec is promising in hyperspectral data exploitation, there are still two issues in using the $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$ model. One is how to determine the rank of the low-rank matrix m, which represents BKG. The other is how to determine the cardinality of the sparse signals, k, which is the sparsity of the signal sources. With the given values of m and k, GoDec breaks up solving two simultaneous optimization problems for finding the low-rank matrix and the sparsity matrix into two iterative processes, which find low-rank matrices and sparse matrices iteratively instead of finding low-rank and spare matrices simultaneously. Consequently, to utilize effectively GoDec, the method of determining these two values is a key success to its use. Until now, this issue remains unsolved and presents a great challenge in applications. One goal of this article is to provide a feasible solution to resolving this issue.

III. GO DECOMPOSITION

Let $\{\mathbf{r}_i\}_{i=1}^N$ be the set of data sample vectors and $\mathbf{X}_{N\times L} = [\mathbf{r}_1\mathbf{r}_2\cdots\mathbf{r}_N]^T$ be its data matrix. In addition, let the total number of spectral bands be denoted by $L \leq N$. The LRaSMD is modeled by

$$\mathbf{X}_{N\times L} = \mathbf{L}_{N\times L} + \mathbf{S}_{N\times L} + \mathbf{N}_{N\times L} \tag{1}$$

where $\mathbf{L}_{N\times L}$ is an $N \times L$ low-rank matrix, $\mathbf{S}_{N\times L}$ is an $N \times L$ sparsity matrix, and $\mathbf{N}_{N\times L}$ is an $N \times L$ noise matrix. Such a "low-rank + sparse" decomposition problem in (1) can be solved by minimizing the following constrained optimization decomposition error:

$$\min_{\mathbf{L}_{N\times L},\mathbf{S}_{N\times L}} ||\mathbf{X}_{N\times L} - \mathbf{L}_{N\times L} - \mathbf{S}_{N\times L}||_F^2 \tag{2}$$

subject to $\text{rank}(\mathbf{L}_{N\times L}) \leq m$ and $\text{card}(\mathbf{S}_{N\times L}) \leq k$.

The GoDec developed in [24] was designed to solve (2) by iteratively solving the following two optimization problems:

$$\mathbf{L}_{N\times L}^{(k)} = \arg\left\{\min_{\text{rank}(\mathbf{L}_{N\times L})\leq m} ||\mathbf{X}_{N\times L} - \mathbf{L}_{N\times L} - \mathbf{S}_{N\times L}^{(k-1)}||_F^2\right\} \tag{3}$$

$$\mathbf{S}_{N\times L}^{(k)} = \arg\left\{\min_{\text{card}(\mathbf{S}_{N\times L})\leq k} ||\mathbf{X}_{N\times L} - \mathbf{L}_{N\times L}^{(k)} - \mathbf{S}_{N\times L}||_F^2\right\}. \tag{4}$$

GoDec develops a low-rank approximation based on the bilateral random projection (BRP) as well as its power-scheme modification [24]. It first generates two random matrices, $\mathbf{\Psi}_{L\times m}$ and $\mathbf{\Phi}_{N\times m}$ to construct $\mathbf{Y}_{N\times m}$ and $\mathbf{Z}_{L\times m}$ given by

$$\mathbf{Y}_{N\times m} = \mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m} \tag{5}$$

$$\mathbf{Z}_{L\times m} = \mathbf{X}_{N\times L}^T\mathbf{\Phi}_{N\times m}. \tag{6}$$

Then, the BRP-based m-rank approximation of $\mathbf{X}_{N\times L}$ can be obtained by

$$\begin{aligned}\mathbf{L}_{N\times L} &= \mathbf{Y}_{N\times m}\left(\mathbf{\Phi}_{N\times m}^T\mathbf{Y}_{N\times m}\right)^{-1}\mathbf{Z}_{L\times m}^T \\ &= \mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m}\left(\mathbf{\Phi}_{N\times m}^T\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m}\right)^{-1}\left(\mathbf{X}_{N\times L}^T\mathbf{\Phi}_{N\times m}\right)^T.\end{aligned} \tag{7}$$

IV. DERIVATION OF OSP-GO DECOMPOSITION

In order to improve the approximation precision of $\mathbf{L}_{N\times L}$ in (7), we can use the obtained right random projection (5) to build a better left projection matrix $\mathbf{\Phi}_{N\times m}$ and use (6) to build a better $\mathbf{\Psi}_{L\times m}$. In particular, after (5), we update $\mathbf{\Phi}_{N\times m} = \mathbf{Y}_{N\times m}$ and calculate the left random projection $\mathbf{Z}_{L\times m} = \mathbf{X}_{N\times L}^T\mathbf{\Phi}_{N\times m}$, and then, we update $\mathbf{\Psi}_{L\times m} = \mathbf{Z}_{L\times m}$ and calculate the right random projection $\mathbf{Y}_{N\times m} = \mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m}$. A better low-rank approximation $\mathbf{L}_{N\times L}$ will be obtained when the new $\mathbf{Y}_{N\times m}$ and $\mathbf{Z}_{L\times m}$ are applied to (7).

Using the update, $\mathbf{\Phi}_{N\times m} = \mathbf{Y}_{N\times m}$ and $\mathbf{\Psi}_{L\times m} = \mathbf{Z}_{L\times m}$ (7) can be further expressed as

$$\begin{aligned}\mathbf{L}_{N\times L} &= \mathbf{Y}_{N\times m}((\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m})^T\mathbf{Y}_{N\times m})^{-1}\mathbf{Z}_{L\times m}^T \\ &= \mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m}((\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m})^T\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m})^{-1} \\ &\quad\times\left(\mathbf{X}_{N\times L}^T\mathbf{\Phi}_{N\times m}\right)^T \\ &= \mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m}((\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m})^T\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m})^{-1} \\ &\quad\times\left(\mathbf{X}_{N\times L}^T\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m}\right)^T \\ &= \mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m}((\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m})^T\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m})^{-1} \\ &\quad\times(\mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m})^T\mathbf{X}_{N\times L}.\end{aligned} \tag{8}$$

Let

$$\mathbf{U}_{N\times m} = \mathbf{X}_{N\times L}\mathbf{\Psi}_{L\times m}. \tag{9}$$

Then, (8) becomes

$$\begin{aligned}\mathbf{L}_{N\times L} &= \mathbf{U}_{N\times m}\left(\mathbf{U}_{N\times m}^T\mathbf{U}_{N\times m}\right)^{-1}\mathbf{U}_{N\times m}^T\mathbf{X}_{N\times L}\\ &= P_{\mathbf{U}_{N\times m}}\mathbf{X}_{N\times L}\end{aligned} \tag{10}$$

where $P_{\mathbf{U}_{N\times m}}$ is an $N\times N$ OSP matrix [30] defined by

$$P_{\mathbf{U}_{N\times m}} = \mathbf{U}_{N\times m}\left(\mathbf{U}_{N\times m}^T\mathbf{U}_{N\times m}\right)^{-1}\mathbf{U}_{N\times m}^T. \tag{11}$$

Then

$$\begin{aligned}\mathbf{S}_{N\times L} &= P_{\Omega}(\mathbf{X}_{N\times L} - \mathbf{L}_{N\times L}) = P_{\Omega}(\mathbf{X}_{N\times L} - P_{\mathbf{U}_{N\times m}}\mathbf{X}_{N\times L})\\ &= P_{\Omega}((\mathbf{I} - P_{\mathbf{U}_{N\times m}})\mathbf{X}_{N\times L}) = P_{\Omega}\left(P^{\perp}_{\mathbf{U}_{N\times m}}\mathbf{X}_{N\times L}\right)\end{aligned} \tag{12}$$

where $P_{\Omega}(\cdot)$ is the projection of a matrix onto an entry set Ω, which is the nonzero subset of the first k largest entries of $P^{\perp}_{\mathbf{U}_{N\times m}}\mathbf{X}_{N\times L}$. Furthermore

$$\begin{aligned}\mathbf{L}_{N\times L} &= P_{\mathbf{U}_{N\times m}}(\mathbf{X}_{N\times L} - \mathbf{S}_{N\times L})\\ &= P_{\mathbf{U}_{N\times m}}\left(\mathbf{X}_{N\times L} - P_{\Omega}\left(P^{\perp}_{\mathbf{U}_{N\times m}}\mathbf{X}_{N\times L}\right)\right)\\ &= P_{\mathbf{U}_{N\times m}}\mathbf{X}_{N\times L} - P_{\mathbf{U}_{N\times m}}P_{\Omega}\left(P^{\perp}_{\mathbf{U}_{N\times m}}\mathbf{X}_{N\times L}\right)\\ &= P_{\mathbf{U}_{N\times m}}\mathbf{X}_{N\times L}\end{aligned} \tag{13}$$

since $P_{\mathbf{U}_{N\times m}}P_{\Omega}(P^{\perp}_{\mathbf{U}_{N\times m}}\mathbf{X}_{N\times L}) = 0$.

Algorithm 1 OSP-GoDec

1. Initial conditions:
   Prescribe an error threshold, ε
   p determined by MOCA developed in Section V or VD developed in Section VI.
   m estimated by MX-SVD developed in Sections V. $j = p - m$ with $k = j \times N$
   Randomly generate $\mathbf{\Psi}_{L\times m}$
   Let $\mathbf{L}^{(0)}_{N\times L} = \mathbf{X}_{N\times L}$, $\mathbf{S}^{(0)}_{N\times L} = \mathbf{O}_{N\times L}$
2. Calculate for $i \geq 1$

$$\mathbf{U}^{(i)}_{N\times m} = (\mathbf{X}_{N\times L} - \mathbf{S}^{(i-1)}_{N\times L})\mathbf{\Psi}_{L\times m} \tag{14}$$

3. In order to implement two iterative loops using (13) and (12), we need include $\mathbf{S}^{(i-1)}_{N\times L}$ in (13) and $\mathbf{L}^{(i)}_{N\times L}$ in (12) as follows.

$$\mathbf{L}^{(i)}_{N\times L} = P_{\mathbf{U}^{(i)}_{N\times m}}(\mathbf{X}_{N\times L} - \mathbf{S}^{(i-1)}_{N\times L}) \tag{15}$$

$$\mathbf{S}^{(i)}_{N\times L} = P_{\Omega}\,(\mathbf{X}_{N\times L} - \mathbf{L}^{(i)}_{N\times L}) \tag{16}$$

   where $P_{\Omega}(\cdot)$ is the projection of a matrix onto an entry set Ω which is the nonzero subset of the first k largest entries of $P^{\perp}_{\mathbf{U}^{(i)}_{N\times m}}(\mathbf{X}_{N\times L})$. In other words, these two iterative loops are carried out by beginning the initial condition $\mathbf{L}^{(0)}_{N\times L}$ and then iterating (13) and (12) as follows.

$$\begin{aligned}\mathbf{L}^{(0)}_{N\times L} &\overset{(12)}{\Rightarrow} \mathbf{S}^{(0)}_{N\times L} \overset{(15)}{\Rightarrow} \mathbf{L}^{(1)}_{N\times L} \overset{(16)}{\Rightarrow} \mathbf{S}^{(1)}_{N\times L}\cdots\\ &\overset{(15)}{\Rightarrow} \mathbf{L}^{(i)}_{N\times L} \overset{(16)}{\Rightarrow} \mathbf{S}^{(i)}_{N\times L}\end{aligned} \tag{17}$$

4. Stopping rule
   If $||\mathbf{X}_{N\times L} - \mathbf{L}^{(i)}_{N\times L} - \mathbf{S}^{(i)}_{N\times L}||^2_F/||\mathbf{X}_{N\times L}||^2_F > \varepsilon$, go to step 2. Otherwise, the algorithm is terminated.

Therefore, when the GoDec [24] can be implemented as OSP-GoDec, the parameter q in GoDec is set to 0. In this case, there is no need to determine the value of q as GoDec does. A step-by-step implementation of OSP-GoDec is given algorithm 1.

It is worth noting that in OSP-GoDec, the values of m and j are given. Unfortunately, there is no procedure included in the original GoDec regarding how to find their values, which must be adjusted empirically. Specifically, m and j represent the number of BKG signal sources and sparse-signal sources, respectively. In the following two sections, instead of determining m and j directly, we determine the sum of m and j, $p = m + j$ first and then determine m later. As a result, k can be determined by $j = p - m$, i.e., $k = j \times N$ [20].

V. Determination of Low-Rank Matrix and Sparse Matrix by Maximum Orthogonal Subspace Projection

Interestingly, a recent approach developed by Kuybeda *et al.* [35], called MOCA can be used to determine p followed by the MX-SVD to determine m. Its detailed process is described as follows.

Let $\mathbf{X}$ be a data matrix formed by data sample vectors $\{\mathbf{r}_i\}_{i=1}^N$, i.e., $\mathbf{X} = [\mathbf{r}_1\mathbf{r}_2\cdots\mathbf{r}_N]$. Define the norm of the data matrix $\mathbf{X}$ by

$$||\mathbf{X}|| = \max_{1\leq i\leq N}||\mathbf{r}_i|| \tag{18}$$

where $||\mathbf{r}_i||$ is the length of the vector $\mathbf{r}_i = (r_{i1}, r_{i2}, \ldots, r_{iL})^T$ defined by $||\mathbf{r}_i||^2 = \sum_{l=1}^L r_{il}^2$. Assume that

$$i^* = \arg\{\max_{1\leq i\leq N}||\mathbf{r}_i||\}. \tag{19}$$

The norm of matrix $\mathbf{X}$ in (18) can be further expressed by

$$||\mathbf{X}|| = ||\mathbf{r}_{i^*}|| \tag{20}$$

which is exactly the brightest pixel $\mathbf{r}_{i^*}$ whose norm has the maximum vector length.

A. MX-SVD

We first assume that p is given and then determine j. Initially, let $j = 1$, i.e., $\mathbf{T}_j = \emptyset$. We then use SVD to find the first p principal left singular vectors of the data matrix $\mathbf{X} = [\mathbf{r}_1\mathbf{r}_2\cdots\mathbf{r}_N]$ denoted by $\mathbf{s}_1^{(0)}, \mathbf{s}_2^{(0)}, \ldots, \mathbf{s}_p^{(0)}$, in which case the p-dimensional signal subspace is given by $\mathbf{S}_p = [\mathbf{s}_1^{(0)}\mathbf{s}_2^{(0)}\cdots\mathbf{s}_p^{(0)}]$. Let $\mathbf{t}_1 = \arg\{\max_{\mathbf{r}\in\{\mathbf{r}_i\}_{i=1}^N}[||\mathbf{P}^{\perp}_{\mathbf{S}_p}\mathbf{r}||]\}$. Next, we increase j by 1, i.e., $j = 1$, and let $\mathbf{T}_1 = [\mathbf{t}_1]$. We once again use SVD to find the first $p-1$ principal left singular vectors $\mathbf{s}_1^{(1)}, \mathbf{s}_2^{(1)}, \ldots, \mathbf{s}_{p-1}^{(1)}$ of the matrix $\mathbf{P}^{\perp}_{\mathbf{T}_1}\mathbf{X}$ obtained by operating an orthogonal subspace projector $\mathbf{P}^{\perp}_{\mathbf{T}_1}$ via (11) on the original data matrix $\mathbf{X}$ to form $\mathbf{S}_p = [\mathbf{B}_{p-1}|\mathbf{T}_1]$, where $\mathbf{B}_{p-1} = [\mathbf{s}_1^{(1)}\mathbf{s}_2^{(1)}\cdots\mathbf{s}_{p-1}^{(1)}]$. Then, we further find $\mathbf{t}_2 = \arg\{\max_{\mathbf{r}\in\{\mathbf{r}_i\}_{i=1}^N}[||\mathbf{P}^{\perp}_{[\mathbf{B}_{p-1}|\mathbf{T}_1]}\mathbf{r}||]\}$ and add it to the current target subspace $\mathbf{T}_1$ to form a new target subspace $\mathbf{T}_2 = [\mathbf{t}_1\mathbf{t}_2]$. We repeat the same procedure until $j = p$ and find $\min_{1\leq j\leq p}[||\mathbf{P}^{\perp}_{[\mathbf{B}_{p-j}|\mathbf{T}_j]}\mathbf{t}_j||]$, which is the minimum of the maximum orthogonal projections $\{\mathbf{t}_j\}_{j=1}^p$ found in each previous stage. The $j^* = \arg\{\min_{1\leq j\leq p}[||\mathbf{t}_j||]\}$, which yields

Algorithm 2 MX-SVD

1. Initialization: Let p be given and $j = 1$, i.e., $\mathbf{T}_1 = \emptyset$.
2. Use SVD to find the first p principal left singular vectors of the data matrix $\mathbf{X} = [\mathbf{r}_1\mathbf{r}_2\cdots\mathbf{r}_N]$, denoted by $\mathbf{s}_1^{(0)}, \mathbf{s}_2^{(0)}, \ldots, \mathbf{s}_p^{(0)}$ and let $\mathbf{t}_1 = \arg\left\{\max_{\mathbf{r}\in\{\mathbf{r}_i\}_{i=1}^N}\left[||\mathbf{P}^{\perp}_{\mathbf{S}_p}\mathbf{r}||\right]\right\}$.
3. Let $j \leftarrow j+1$. Form $\mathbf{T}_j = \left[\mathbf{T}_{j-1}\mathbf{r}_j^*\right]$. Use SVD to find the first $p-j$ principal left singular vectors, $\mathbf{s}_1^{(j)}, \mathbf{s}_2^{(j)}, \ldots, \mathbf{s}_{p-j}^{(j)}$ of the matrix $\mathbf{P}^{\perp}_{\mathbf{T}_j}\mathbf{X}$ and find
$$\mathbf{t}_j = \arg\left\{\max_{\mathbf{r}\in\{\mathbf{r}_i\}_{i=1}^N}\left[||P^{\perp}_{[\mathbf{B}_{p-j}|\mathbf{T}_j]}\mathbf{r}||\right]\right\} \quad (21)$$
4. Repeat step 3 until $j = p$. Continue
5. At this stage, calculate $j^{\text{MX-SVD}}$ by
$$j^{\text{MX-SVD}} \equiv \arg\left\{\min_{1\le j\le p}\left[||\mathbf{P}^{\perp}_{[\mathbf{B}_{p-j}|\mathbf{T}_j]}\mathbf{t}_j||\right]\right\} \quad (22)$$

the minimum, turns out to be the rank of the desired target subspace, and $m = p - j^*$ is the rank of the BKG subspace.

The key idea of the above process is to work on (18) and (19) backward to find the target pixels by a sequence of orthogonal complement subspace projections $\{\mathbf{P}^{\perp}_{\mathbf{T}_j}\mathbf{r}_i\}_{j=1}^p$ within the data space until it reaches p. Then, j^* is determined by $j^* = \arg\{\min_{1\le j\le p}[||\mathbf{t}_j||]\}$, referred to as $j^{\text{MX-SVD}}$. The set of target signal sources $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{j^*}$ form a target matrix $\mathbf{T}_{j^*} = [\mathbf{t}_1\mathbf{t}_2\cdots\mathbf{t}_{j^*}]$ as the desired target signal subspace, and the complement subspace within $\mathbf{S}_p$ orthogonal to $\mathbf{T}_{j^*}$ is the BKG subspace $\mathbf{B}_{m^*}$ with $m^* = p - j^*$.

The process described above is called MX-SVD named by Kuybeda *et al.* [35]. The details of MX-SVD implementation can be summarized as follows.

It should be noted that $j^{\text{MX-SVD}}$ in (22) is obtained by assuming that the rank of a signal subspace p is given *a priori*. Unfortunately, finding an appropriate p is a challenging issue, as described in the following section. As an alternative to (22), it will be nice to implement MX-SVD without actually determining p, that is, we can assume that there are L dimensions of an HSI cube (i.e., band dimensionality) and then carry out MX-SVD by making the rank of a signal subspace l ranging from 1 to L, in which case the rank of a signal subspace can be examined by all possible subspaces in the data space of dimensionality L. Since the sequence $\{||\mathbf{P}^{\perp}_{[\mathbf{B}_{l-j}|\mathbf{T}_j]}\mathbf{t}_j||\}_{j=1}^l$ is monotonically decreasing at l, that is
$$\left|\left|\mathbf{P}^{\perp}_{[\mathbf{B}_{l-j}|\mathbf{T}_j]}\mathbf{t}_j\right|\right| \ge \left|\left|\mathbf{P}^{\perp}_{[\mathbf{B}_{(l+1)-j}|\mathbf{T}_j]}\mathbf{t}_j\right|\right| \text{ for } l = 1, \ldots, L-1 \quad (23)$$
we can plot the values of $\{||\mathbf{P}^{\perp}_{[\mathbf{B}_{l-j}|\mathbf{T}_j]}\mathbf{t}_j||\}_{l=1}^L$ against dimensionality l from 1 to L in a similar manner that an eigenvalue distribution is plotted for the eigenenergy-based methods so that the dimension p that causes a sudden drop is the desired rank of the signal subspace $\mathbf{S}_p$. We then use such obtained p in MX-SVD to find the rank of the desired target subspace as $j^{\text{MX-SVD}}$ and $p - j^{\text{MX-SVD}}$ as the rank of the BKG signal subspace by letting $m^{\text{MX-SVD}} = p - j^{\text{MX-SVD}}$.

B. MOCA

When MX-SVD is implemented, it assumes that p is given *a priori*. Therefore, for each p, which is supposed to be the rank of a signal subspace $\mathbf{S}_p$, the data sample vectors $\{\mathbf{r}_i\}_{i=1}^N$ are divided into two index classes: one is the target signal class $I_T(p)$ and the other is the BKG class $I_B(p)$. Now, we define
$$\nu_p = \max_{i\in I_B(p)}\left|\left|\mathbf{P}^{\perp}_{\mathbf{S}_p}\mathbf{r}_i\right|\right|^2 \quad (24)$$
$$\xi_p = \max_{i\in I_T(p)}\left|\left|\mathbf{P}^{\perp}_{\mathbf{S}_p}\mathbf{r}_i\right|\right|^2 \quad (25)$$
$$\eta_p = \max\{\xi_p, \nu_p\}. \quad (26)$$
Then, for each $1 \le p \le L$, we further define
$$\mathbf{t}_p = \arg\left\{\max_{\mathbf{r}}\left|\left|\mathbf{P}^{\perp}_{\mathbf{S}_p}\mathbf{r}\right|\right|\right\} \quad (27)$$
$$\eta_p = ||\mathbf{t}_p||^2. \quad (28)$$
It should be noted that $\{\mathbf{S}_p\}$ is monotonically increasing at p in the sense that $\mathbf{S}_0 \subset \mathbf{S}_1 \subset \cdots \subset \mathbf{S}_p$ and $\{\eta_p\}$ is monotonically decreasing at p in the sense that $\eta_0 \ge \eta_1 \ge \cdots \ge \eta_p$. As a consequence, a prescribed threshold is required to determine how many signal pixels are needed for an SVD to generate, i.e., p. MOCA casts its stopping rule as a binary composite hypothesis testing problem as follows:
$$\begin{aligned} &H_0 : \eta_p \approx p(\eta_p|H_0) = p_0(\eta_p) \\ &\text{versus} \\ &H_1 : \eta_p \approx p(\eta_p|H_1) = p_1(\eta_p) \\ &\quad \text{for } l = 1, 2, \ldots, L \end{aligned} \quad (29)$$
where the alternative hypothesis H_1 and the null hypothesis H_0 represent the target signal subspace and the BKG subspace, respectively. In order to make (29) work, we need to find the probability distributions under both hypotheses. Because the orthogonal complement subspace projections of the data sample vectors $\mathbf{P}^{\perp}_{\mathbf{S}_p}\mathbf{r}_i$ under H_0 are supposed to be the noise sample vectors, it is reasonable for an MOCA to assume that the vector $\mathbf{P}^{\perp}_{\mathbf{S}_p}\mathbf{r}_i$ under H_0 behaves like independent identically Gaussian random variables. Moreover, η_p is the maximum residuals of orthogonal projection obtained in $< \mathbf{S}_p >^{\perp}$ under H_0. By virtue of the extreme value theory [62], η_p can be modeled as a Gumbel distribution, i.e., $F_{\upsilon_p}(\eta_p)$ is the cumulative distribution function (cdf) of υ_p given by (30), as shown at the bottom of the next page

Furthermore, MOCA is made another assumption on η_p under the alternative hypothesis H_1, that is, for each pixel $\mathbf{r}_i$, $i \in I_T(p)$, the maximum of residuals of the orthogonal complement subspace projections can be modeled as a probability distribution $p_{\xi_p}(\eta_p)$, which is assumed to be uniformly distributed on $[0, \eta_{p-1}]$ due to the fact that there is no prior knowledge available about the distribution of the target pixels, and assuming η_p under H_1 uniformly distributed seems most reasonable according to the maximum entropy principle in Shannon's information theory [63].

Under these two assumptions, we can obtain
$$p(H_0, \eta_p) = p_{\upsilon_p}(\eta_p)F_{\xi_p}(\eta_p) = p_{\upsilon p}(\eta_p)(\eta_p/\eta_{p-1}) \quad (31)$$
$$p(H_1, \eta_p) = F_{\upsilon_p}(\eta_p)p_{\xi_p}(\eta_p) = F_{\upsilon_p}(\eta_p)(1/\eta_{p-1}). \quad (32)$$

Since $p_{\eta_p}(\eta_p) = p(H_0, \eta_p) + p(H_1, \eta_p) = (1/\eta_{p-1})[\eta_p p_{v_p}(\eta_p) + F_{v_p}(\eta_p)]$, we can obtain *a posteriori* probability distribution of $p(H_0|\eta_p)$ given by

$$p(H_0|\eta_p) = \frac{\eta_p p_{v_p}(\eta_p)}{\eta_p p_{v_p}(\eta_p) + F_{v_p}(\eta_p)} \tag{33}$$

and *a posteriori* probability distribution of $p(H_1|\eta_p)$ is given by

$$p(H_1|\eta_p) = \frac{F_{v_p}(\eta_p)}{\eta_p p_{v_p}(\eta_p) + F_{v_p}(\eta_p)}. \tag{34}$$

By virtue of (33) and (34), the rank of the desired signal subspace can be obtained by

$$p^* = \arg\{\min_{1\le p\le L} p(H_0|\eta_p) \ge p(H_1|\eta_p)\}. \tag{35}$$

Using (35), p determined by MOCA can be found by

$$p^{\text{MOCA}} = p^*. \tag{36}$$

Now, substituting p^{MOCA} in (36) for the p in (22), which is used to find $j^{\text{MX-SVD}}$, yields the rank of the BKG signal subspace m^{MOCA} given by

$$m^{\text{MOCA}} = p^{\text{MOCA}} - j^{\text{MX-SVD}}. \tag{37}$$

VI. Determination of Number of Spectrally Distinct Signals: Virtual Dimensionality

As an alternative to addressing the issue of determining the values of m and j, we look into the eigenvalue distribution of the data since all the eigenvectors (EVs) are mutually orthogonal and each EV can be used to specify one dimensionality. Let $\{\lambda_l\}_{l=1}^L$ be the eigenvalues of the data sample covariance matrix $\mathbf{K}_{L\times L}$. We can rearrange all the eigenvalues $\{\lambda_l\}_{l=1}^L$ according to their magnitudes in descending order

$$\underbrace{\{\lambda_1 \ge \cdots \ge \lambda_m\}}_{\text{BKG}} \ge \underbrace{\{\lambda_{m+1} \ge \cdots \ge \lambda_{p=m+j}\}}_{\text{sparse signals}} \ge \underbrace{\{\lambda_{p+1} \ge \cdots \ge \lambda_L\}}_{\text{noise}}. \tag{38}$$

According to factor analysis in [64] and [65], $\{\lambda_l\}_{l=1}^L$ can be decomposed into three categories: a primary set of eigenvalues $\{\lambda_l\}_{l=1}^m$ corresponding to m BKG sources and a secondary set of eigenvalues $\{\lambda_l\}_{l=m+1}^p$ with $j = p - m$ corresponding to sparse j signal sources and $\{\lambda_l\}_{l=p+1}^L$ corresponding to noise sources. The challenging issue associated with (38) is how to break up $\{\lambda_l\}_{l=1}^L$ into these three categories at the values of m and j.

This section presents an approach to determining the value of p, which is quite different from that determined by MOCA. It takes advantage of the VD recently developed to determine the number of spectrally distinct signatures p assumed to be present in hyperspectral imagery [25], [26], [31]–[34], which can be considered as the sum of the BKG signal sources and sparse signals, i.e., the value of p in (22) or (36). Its idea is to formulate the VD-determination problem as a binary hypothesis testing problem as follows [66].

Let $\{\hat{\lambda}_l\}_{l=1}^L$ and $\{\lambda_l\}_{l=1}^L$ be the eigenvalues of the data sample correlation matrix $\mathbf{R}_{L\times L}$ and the data sample covariance matrix $\mathbf{K}_{L\times L}$, respectively. If we assume that the noise in each spectral band has zero mean and variance $\sigma_{n_l}^2$, then $\hat{\lambda}_l = \mu_l^2 + \sigma_{n_l}^2$ and $\lambda_l = \sigma_{n_l}^2$, where μ_l is the sample mean of the lth spectral band. If there is a signal present in the lth band, then $\hat{\lambda}_l - \lambda_l = \mu_l^2 > 0$; otherwise, $\hat{\lambda}_l - \lambda_l = 0$. With this interpretation, we can formulate a binary composite hypothesis testing problem as

$$\begin{aligned} &H_0 : z_l = \hat{\lambda}_l - \lambda_l = s_l^2 = 0 \\ &\quad \text{versus} \\ &H_1 : z_l = \hat{\lambda}_l - \lambda_l = s_l^2 > 0 \\ &\qquad\qquad \text{for } l = 1, 2, \ldots, L \end{aligned} \tag{39}$$

where the null hypothesis H_0 indicates the absence of a signal source with the case that the correlation-eigenvalue $\hat{\lambda}_l$ is equal to its corresponding covariance eigenvalue λ_l and the alternative hypothesis H_1 represents the case that the correlation-eigenvalue $\hat{\lambda}_l$ is greater than its corresponding covariance eigenvalue λ_l, respectively. Using this formulation, a Neyman–Pearson detector, denoted by $\delta^{\text{NP}}(z_l)$, can be derived to test the difference between the correlation-eigenvalue and the covariance-eigenvalue $\hat{\lambda}_l - \lambda_l$ via (39) by maximizing the detection power P_D with a constraint β imposed on the false-alarm probability P_F. According to [66], $\delta^{\text{NP}}(z_l)$ can be obtained by

$$\delta^{\text{NP}}(z_l) = \begin{cases} 1; & \text{if } \Lambda(z_l) > \tau_l \\ 1\text{with probability } \kappa; & \text{if } \Lambda(z_l) = \tau_l \\ 0; & \text{if } \Lambda(z_l) < \tau_l \end{cases} \tag{40}$$

where the likelihood ratio test $\Lambda(z_l)$ is given by $\Lambda(z_l) = p_1(z_l)/p_0(z_l)$ given in [25] and [31], and the threshold value τ_l is determined by $\text{P}_\text{F} = \beta$. Finally, VD is determined by the number of times the detector $\delta^{\text{NP}}(z_l)$ in (40) fails the test where a test failure for a specific spectral band l indicates that there is a signal energy contributed to the eigenvalue $\hat{\lambda}_l$ in the lth spectral band. A method developed by Harsanyi–Farrend–Chang, called HFC method in [67] along with its noise-whitened version HFC (NWHFC), can be used to find VD for (40) via $\delta^{\text{NP}}(z_l)$. Such a resulting VD estimate for p in (36) is denoted by $p^{\text{HFC/NWHFC}}(\text{P}_\text{F})$ by specifying the constraint β imposed on P_F.

VII. Application to Anomaly Detection

In previous sections, a theory of OSP-GoDec was derived by introducing OSP into GoDec, where the number of spectral distinct signal sources, the order of the low-rank matrix m, and

$$F_{v_p}(x) \approx \exp\left\{-e^{-(2\log N)^{1/2}\left[\frac{x-\sigma^2(L-p)}{\sigma^2\sqrt{2(L-p)}} - (2\log N)^{1/2} + \frac{1}{2}(2\log N)^{-1/2}(\log\log N + \log 4\pi)\right]}\right\}. \tag{30}$$

the number of sparse-signal sources j can be determined by MX-SVD in (22) with p estimated by MOCA in (36) or VD in (40).

A. Anomaly Detectors

In order to demonstrate the utility of the developed OSP-GoDec along with MOCA/VD-determined p and MX-SVD-determined m *and* j in anomaly detection, the well-known anomaly detector developed in [68], referred to as Reed and Xiaoli anomaly detector (RX-AD), is used for benchmark comparison. It is defined by

$$\delta^{\text{RX-AD}}(\mathbf{r}) = (\mathbf{r}-\boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{r}-\boldsymbol{\mu}) \tag{41}$$

where $\mathbf{r} \in \Re^{L\times 1}$ is a data sample currently being processed, $\boldsymbol{\mu}$ is the sample mean of the image, and $\mathbf{K} \in \Re^{L\times L}$ is the global sample data covariance matrix given by $\mathbf{K} = (1/N)\sum_{i=1}^{N}(\mathbf{r}_i-\boldsymbol{\mu})(\mathbf{r}_i-\boldsymbol{\mu})^T$, where N is the total number of pixels and L is the total number of bands.

As an alternative, RX-AD can also be modified by replacing $\mathbf{K}$ and $\mathbf{r}-\boldsymbol{\mu}$ by $\mathbf{R}$ and $\mathbf{r}$, respectively, as follows:

$$\delta^{\text{R-AD}}(\mathbf{r}) = \mathbf{r}^T \mathbf{R}^{-1}\mathbf{r} \tag{42}$$

referred to as R-AD in [25] and [69], [70], where $\mathbf{r} \in \Re^{L\times 1}$ is a data sample currently being processed and $\mathbf{R} \in \Re^{L\times L}$ is the global sample correlation matrix given by $\mathbf{R} = (1/N)\sum_{i=1}^{N}\mathbf{r}_i\mathbf{r}_i^T$. For simplicity, we use $\delta^{\text{RX/R-AD}}(\mathbf{r})$ to represent two ADs specified by (41) and (42).

The two key issues to implement $\delta^{\text{RX/R-AD}}(\mathbf{r})$ are implemented. One is the use of inverting the covariance/correlation matrix **K/R** to suppress BKG to increase the contrast between the data sample vector $\mathbf{r}$ currently being processed and its surrounding BKG. The other is the matching of the data sample vector $\mathbf{r}$ with its suppressed $\mathbf{K}^{-1}\mathbf{r}/\mathbf{R}^{-1}\mathbf{r}$ by performing the spectral angle mapper (SAM) between $\mathbf{r}$ and $\mathbf{K}^{-1}\mathbf{r}/\mathbf{R}^{-1}\mathbf{r}$ to enhance anomaly detectability. In the past, many research works have been directed to adjusting the window size of **K/R** to limit the surrounding neighborhoods of the data sample vectors using local, adaptive, and sliding windows but a very little has been done with the data sample vector $\mathbf{r}$ itself. The LRaSMD model allows us to deal with both issues at the same time, since the low-rank matrix $\mathbf{L}_m$ and sparse matrix $\mathbf{S}_j$ can be used for both purposes. To explore fully the information provided by the LRaSMD model, six types of implementations for each AD are of particular interest according to the low-rank matrix $\mathbf{L}_m$ and the sparse matrix $\mathbf{S}_j$ found by OSP-GoDec, where p and $j = p-m$ are estimated by MOCA/VD and MX-SVD, respectively, discussed in Sections V and VI.

In what follows, we use the notation of $\boldsymbol{\delta}_{\mathbf{B}}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{A}})$ to define various types of implementing an anomaly detector for experiments, where $\mathbf{A}$ should include the sparse signal and $\mathbf{B}$ is used as BKG matrices, respectively.

B. Six Types of Implementations for AD

Since we are interested in the data sample vectors to be specified by the components containing anomalies, this leads to six types of $\boldsymbol{\delta}_{\mathbf{B}}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{A}})$, where the data sample vector $\mathbf{r}$ to be detected should include a signal component specified by $\mathbf{A}$, which has two cases, $\mathbf{S}_j$ and $\mathbf{L}_m+\mathbf{S}_j$, and the BKG to be suppressed specified by $\mathbf{B}$, which can have three cases, $\mathbf{S}_j$, $\mathbf{L}_m$, and $\mathbf{L}_m+\mathbf{S}_j$, that is, $(\mathbf{S}_j, \mathbf{S}_j)$, $(\mathbf{S}_j, \mathbf{L}_m)$, $(\mathbf{S}_j, \mathbf{L}_m+\mathbf{S}_j)$, $(\mathbf{L}_m+\mathbf{S}_j, \mathbf{S}_j)$, $(\mathbf{L}_m+\mathbf{S}_j, \mathbf{L}_m)$, and $(\mathbf{L}_m+\mathbf{S}_j, \mathbf{L}_m+\mathbf{S}_j)$.

TABLE I
VARIOUS TYPES OF ADS USING LRASMD MODEL

Various types of ADs	pixel	sample mean	covariance / correlation matrix
$\delta^{\text{RX/R-AD}}(\mathbf{r})$	$\mathbf{r}$	$\boldsymbol{\mu}$/NA	$\mathbf{K}_{L\times L}$ / $\mathbf{R}_{L\times L}$
$\delta_{\mathbf{S}_j}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{S}_j})$	$\mathbf{r}_{\mathbf{S}_j}$	$\boldsymbol{\mu}_{\mathbf{S}_j}$ /NA	$\mathbf{K}_{\mathbf{S}_j}$ / $\mathbf{R}_{\mathbf{S}_j}$
$\delta_{\mathbf{L}_m}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{S}_j})$	$\mathbf{r}_{\mathbf{S}_j}$	$\boldsymbol{\mu}_{\mathbf{L}_m}$ /NA	$\mathbf{K}_{\mathbf{L}_m}$ / $\mathbf{R}_{\mathbf{L}_m}$
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{S}_j})$	$\mathbf{r}_{\mathbf{S}_j}$	$\boldsymbol{\mu}_{\mathbf{L}_m+\mathbf{S}_j}$ /NA	$\mathbf{K}_{\mathbf{L}_m+\mathbf{S}_j}$ / $\mathbf{R}_{\mathbf{L}_m+\mathbf{S}_j}$
$\delta_{\mathbf{S}_j}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	$\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j}$	$\boldsymbol{\mu}_{\mathbf{S}_j}$ /NA	$\mathbf{K}_{\mathbf{S}_j}$ / $\mathbf{R}_{\mathbf{S}_j}$
$\delta_{\mathbf{L}_m}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	$\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j}$	$\boldsymbol{\mu}_{\mathbf{L}_m}$ /NA	$\mathbf{K}_{\mathbf{L}_m}$ / $\mathbf{R}_{\mathbf{L}_m}$
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	$\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j}$	$\boldsymbol{\mu}_{\mathbf{L}_m+\mathbf{S}_j}$ /NA	$\mathbf{K}_{\mathbf{L}_m+\mathbf{S}_j}$ / $\mathbf{R}_{\mathbf{L}_m+\mathbf{S}_j}$

Table I lists the traditional RX/R-AD and six types of AD using the LRaSMD model with the low-rank matrix $\mathbf{L}_m$ of rank m and the low-rank + sparsity matrix $\mathbf{L}_m+\mathbf{S}_j$ of rank p along with their corresponding sample means and covariance/correlation matrices, which will be used for the following experiments.

C. Anomaly Detectors to be Used for Comparison

Since $\mathbf{L}+\mathbf{S}$ and $\mathbf{L}+\mathbf{S}+\mathbf{N}$ are the two data representation models for LRaSMD, they cannot be implemented alone but need to be used in conjunction with an operator designed for a particular application, which is the anomaly detector in this article. To realize the $\mathbf{L}+\mathbf{S}$ model, RPCA [1]-based RX/R-AD is used as the candidate for comparison. On the other hand, to implement the $\mathbf{L}+\mathbf{S}+\mathbf{N}$ model, an LRaSMD-based Mahalanobis-distance anomaly detector (LSMAD) developed in [20] is used as a candidate for comparison due to the fact that LSMAD used GoDec and was shown to perform better than the subspace anomaly detector (SSAD) and the Euclidean distance-based LRaSMD. It should be noted that the LSMAD implemented two different versions of calculating the Mahalanobis distance by inverting the covariance matrix using either full EVs or m EVs, where m is the rank of the low-rank matrix $\mathbf{L}_m$. In addition to the LRaSMD models, an anomaly detector based on a sparse-representation-based model, referred to as l_2-norm minimization and the distance-weighted regularization matrix and the sum-to-one constraint (CRD-DW-STO), developed in [39] is also used as a candidate for comparison, because it was shown to outperform the sparse-representation-based detector (SRD), RPCA-based anomaly detector, and local RX-AD.

VIII. SYNTHETIC IMAGE EXPERIMENTS

The image scene to be used to simulate the synthetic images was real Cuprite image data shown in Fig. 1(a), which is available at the USGS website http://aviris.jpl.nasa.gov/. This

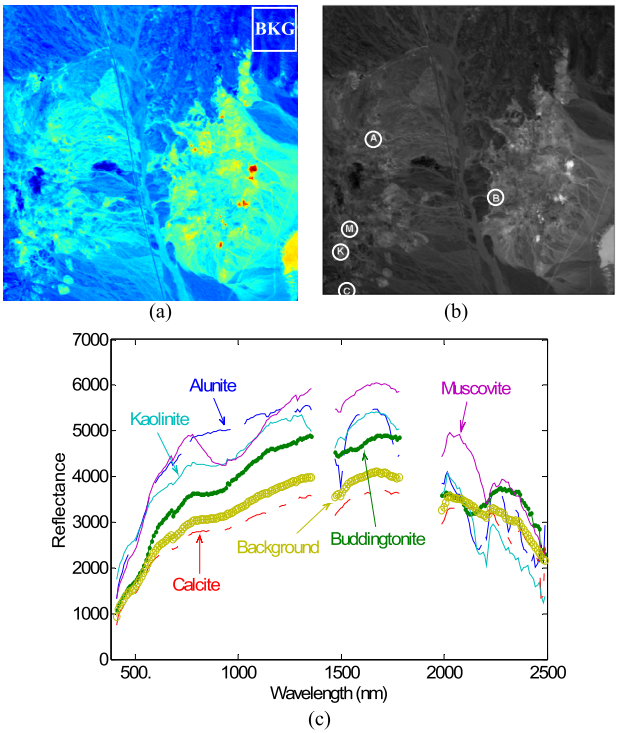


Fig. 1. (a) Cuprite AVIRIS image scene. (b) Spatial positions of five pure pixels corresponding to minerals: alunite (A), buddingtonite (B), calcite (C), kaolinite (K), and muscovite (M). (c) Five mineral reflectance spectra and BKG signature (b), which is the sample mean of the image in (a).

scene is a 224-band image with a size of 350×350 pixels and was collected over the Cuprite mining site, Nevada, in 1997. It is well understood mineralogically. As a result, a total of 189 bands were used for experiments, where bands 1–3, 105–115, and 150–170 have been removed prior to the analysis due to water absorption and low signal-to-noise ratio (SNR) in those bands.

Although there are more than five minerals in the data set, the ground truth available for this region only provides the locations of the pure pixels: alunite (A), buddingtonite (B), calcite (C), kaolinite (K), and muscovite (M). The locations of these five pure minerals are labeled as A, B, C, K, and M, respectively, and are shown in Fig. 1(b). Available from the image scene is a set of these reflectance spectra shown in Fig. 1(c) that will be used to simulate the synthetic images.

Since real images generally do not have complete ground truth about the targets to be detected, we must rely on the synthetic images simulated by complete knowledge to conduct quantitative analysis as the way was performed in [71] and [72] for the performance evaluation of various anomaly-detection algorithms. Fig. 2 shows a synthetic image simulated by 25 panels using five mineral signatures in Fig. 1(c).

Among 25 panels are five 4×4 pure-pixel panels for each row in the first column, five 2×2 pure-pixel panels for each row in the second column, five 2×2 mixed-pixel panels for each row in the third column, and five 1×1 subpixel panels for each row in both the fourth column and the fifth column, where the mixed and panel subpixels were simulated according to the legends in Fig. 2. Therefore, a total of 100 pure pixels (80 in the first column and 20 in the second column), referred

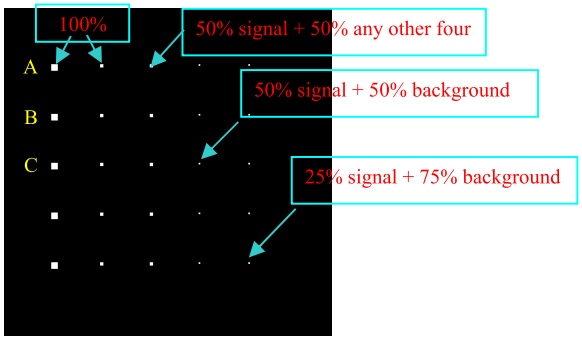


Fig. 2. Set of 25 panels simulated by A, B, C, K, and M.

TABLE II
m ESTIMATED FOR TI BY MOCA AND HFC/NWHFC USING MX-SVD

	p	*m* by MX-SVD
MOCA	19	14
HFC, $P_F \leq 10^{-3}$	13	9
NWHFC $P_F \leq 10^{-3}$	6	1

to as the endmember pixels, were simulated in the data by the five endmembers, A, B, C, K, and M. An area marked by "BKG" at the top-right corner of Fig. 1(a) was selected to find its sample mean, i.e., the average of all pixel vectors within the area "BKG," denoted by **b** and plotted in Fig. 1(c), to be used to simulate the BKG for the image scene in Fig. 2. The reason for this BKG selection is empirical, since the selected area "BKG" seemed more homogeneous than other regions. Nevertheless, other areas can also be selected for the same purpose. This **b**-simulated image BKG was further corrupted by an additive noise to achieve a certain SNR that was defined as 50% signature (i.e., reflectance/radiance) divided by the standard deviation of the noise in [30].

Once the target pixels and the BKG are simulated, two types of target insertion can be designed to simulate the experiments for various applications.

Target Implantation (TI): The first type of target insertion is target implantation (TI) that can be simulated by removing the BKG pixels and replacing them with clean target panels where the image BKG is corrupted by a Gaussian noise. In this case, TI can be used for endmember finding or extraction.

Target Embeddedness (TE): A second type of target insertion is target embeddedness (TE) that can also be simulated by embedding clean target panels into the clean image BKG with additive Gaussian noise by superimposing target pixels over the BKG pixels in which case TE can be used for signal-detection model.

A. TI Experiments

According to TI, there are five pure signatures considered as endmembers and one BKG signature considered to be a mixed signature. In this case, the value of p is set to 6. Table II tabulates the value of p estimated by the MOCA and HFC/NWHFC methods and the value of m estimated by MX-SVD for TI, where $n_{\mathrm{VD}} = p$ of TI is, indeed, also 6 if $P_F \leq 10^{-3}$.

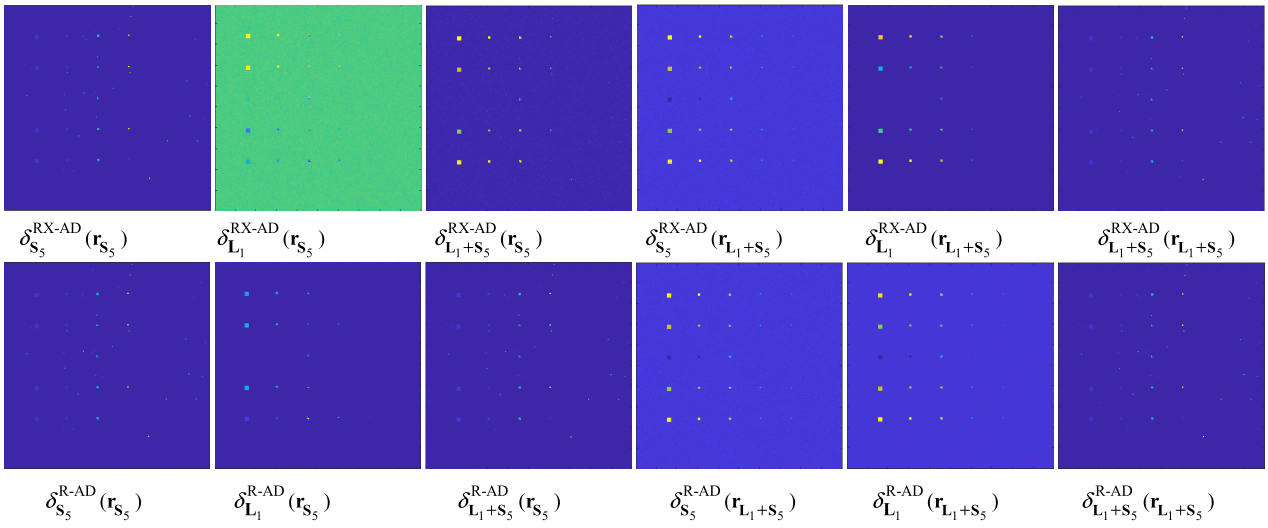


Fig. 3. TI: $p = 6$, $m = 1$, and $j = 5$.

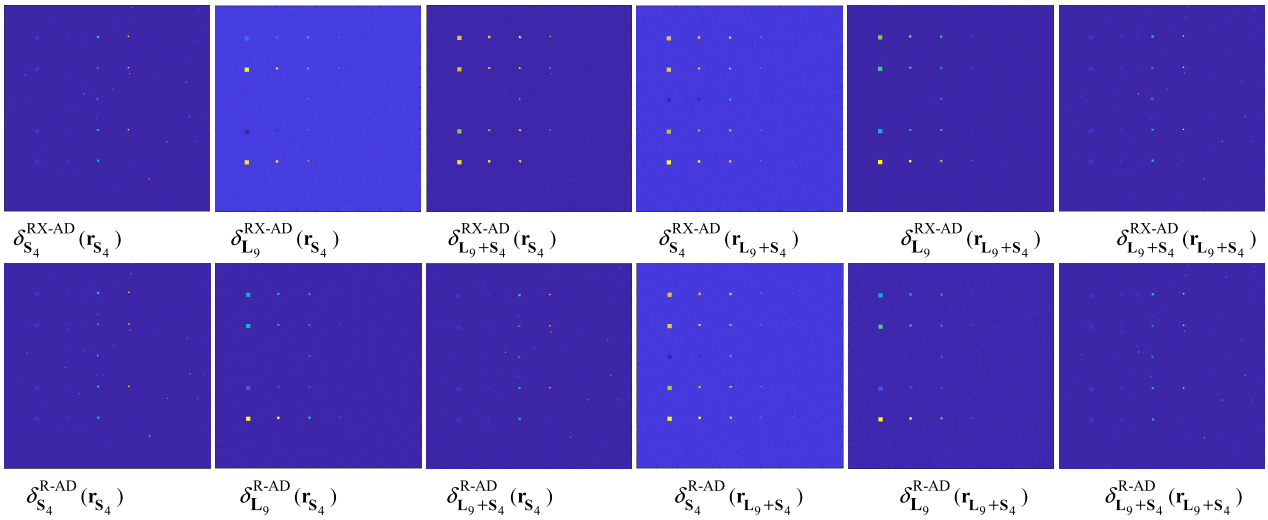


Fig. 4. TI: $p = 13$, $m = 9$, and $j = 4$.

By employing MX-SVD, the dimensionality of the BKG subspace is $m = 1$, which is consistent with the construction of TI data. In this case, the number of sparse signals is $j = 6 - 1 = 5$ and the cardinality of sparsity k is estimated by $k = 5 \times N$, where $N = 200 \times 200 = 40\,000$ is the number of total pixels. On the other hand, if MOCA was used, $p = 19$ and $m = 14$, which gave $j = 19 - 14 = 5$. In this case, $k = 5 \times N$ with $N = 40\,000$.

Based on the results obtained for p and m in Table II, Figs. 3–5 show the detection maps produced by six types of $\boldsymbol{\delta}^{\text{RX/R-AD}}(\mathbf{r})$ using $(p, m, j) = (6, 1, 5)$, $(13, 9, 4)$, and $(19, 14, 5)$. As shown in Figs. 3 and 4, $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$ and $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, which used the low-rank matrix $\mathbf{L}_m$ to suppress the BKG, detected panel pixels in the first three columns but missed most or all panel subpixels in the fourth and fifth columns. By contrast, $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ and $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ could detect mixed-panel pixels in the third column and panel subpixels in the fourth and fifth columns but missed all panel pixels in the first two columns. The most interesting finding is the case of $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, which showed that $\boldsymbol{\delta}^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ and $\boldsymbol{\delta}^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ performed completely differently. Specifically, $\boldsymbol{\delta}^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ detected panel pixels in the first three columns but missed most or all panel subpixels in the fourth and fifth columns, because $\boldsymbol{\delta}^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ removed the first-order statistics, sample mean from the data prior to AD. Conversely, without removing the sample mean, $\boldsymbol{\delta}^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ detected the mixed-panel pixels in the third column and the panel subpixels in the fourth and fifth columns but missed all panel pixels in the first two columns due to the fact that BKG is largely characterized by the sample means that was suppressed by using $\mathbf{L}_m + \mathbf{S}_j$. From the results in Figs. 3–5, we can conclude that $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\boldsymbol{\delta}^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ detected panel pixels in the first three columns but missed most or all panel subpixels in the fourth and fifth columns, while $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\boldsymbol{\delta}^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, and $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ detected mixed-panel pixels in the third column and panel subpixels in the fourth and fifth columns but missed all panel pixels in the first two columns. In addition, by visually inspecting the results in Figs. 3–5, the best one

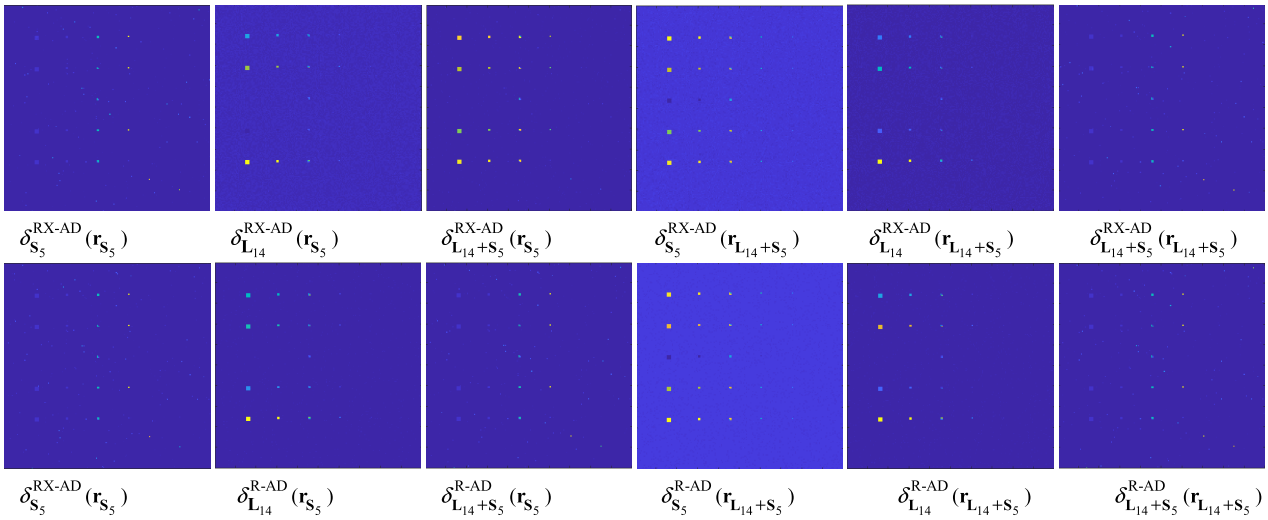
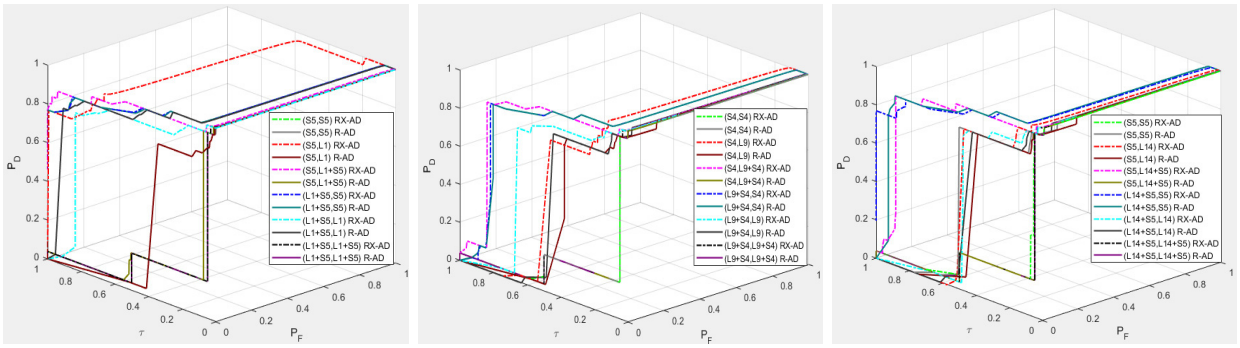


Fig. 5. TI: $p = 19$, $m = 14$, and $j = 5$.



p =6, m = 1 and j = 5 p =13, m = 9 and j = 4 p =19, m = 14 and j =5

Fig. 6. 3-D ROC curves of TI.

was produced by using $p = 13$, $m = 9$, and $j = 4$. It should be noted that since the signature, Calcite, used to simulate the panels was very close to the BKG signature, its detection in the third row was poor.

Using the 3-D ROC analysis developed in [26], [73], and [74], Fig. 6 plots the 3-D ROC curves and Table III tabulates the area under curve (AUC) values of their corresponding 2-D ROC curves [2-D ROC curves of (P_D, P_F), 2-D ROC curves of (P_D, τ), and 2-D ROC curves of (P_F, τ)], which are all between 0 and 1 where the best cases for various values of (p, m, j) are boldfaced. In order to take advantage of the three 2-D ROC curves of (P_D, P_F), (P_D, τ), and (P_F, τ), generated by the 3-D ROC curves, we further design a new metric as AUC_{OD} to measure the overall detection performance, which is defined by

$$\text{AUC}_{\text{OD}} = \text{AUC}(P_D, P_F) + \text{AUC}(P_D, \tau) - \text{AUC}(P_F, \tau) \tag{43}$$

where the values of AUC(P_D, P_F) and AUC(P_D, τ) are summed with the subtraction of AUC(P_F, τ). This is because a higher value of AUC(P_D, P_F) and AUC(P_D, τ) indicates a higher detection performance. Conversely, a lower value of AUC(P_F, τ) indicates a better BKG suppression, and thus a better detection performance. As a result, the AUC_{OD} defined in (43), indeed, combines the values calculated by the AUC values produced by these three 2-D ROC curves to produce a quantitative value of the overall detection performance for each of the test anomaly detectors where a higher value of an AUC_{OD} produced by an anomaly detector is a better and more effective the anomaly detector.

Comparing the quantitative results in Table III with the qualitative results in Figs. 3–5 by visual inspection, both results agreed and were consistent. As we can see, the best performance was $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, which produced the highest $\text{AUC}_{\text{OD}} = 1.8564$ for the case of (p, m, j) = (6, 1, 5) in Table III and also the best visual inspection in Fig. 3. In addition, agreed on Table III and Figs. 3–5 were the best cases, $\text{AUC}_{\text{OD}} = 1.7864$ produced by $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ for (p, m, j) = (13, 9, 4) in Fig. 4 and $\text{AUC}_{\text{OD}} = 1.8232$ produced by $\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ for (p, m, j) = (19, 14, 5) in Fig. 5.

A comment is noteworthy. An anomaly is generally invisible and cannot be inspected visually. In other words, if a target is visible, it indicates that it can be detected by visual inspection. In this case, it should be called target detection but not anomaly detection. Using this definition, the panel pixels in the first three columns in TI and TE should not be called anomalies, because we can visually see these panel pixels. The only targets that can be called anomalies are those ten panel subpixels in the fourth and fifth columns. In terms of this interpretation, $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ are actually target detectors as opposed

TABLE III
COMPARISON OF VARIOUS ANOMALY DETECTORS FOR TI USING AUC VALUES CALCULATED BY (43)

	p =6, m = 1 and j = 5				p =13, m = 9 and j = 4				p =19, m = 14 and j =5			
Detector	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
$\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9959	0.1542	0.0020	1.1481	0.9960	0.1558	0.0020	1.1498	0.9955	0.1659	0.0022	1.1592
$\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9959	0.1542	0.0020	1.1481	0.9958	0.1204	0.0020	1.1142	0.9936	0.4468	0.0095	1.4309
$\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$	0.9590	0.9169	0.6111	1.2648	0.9958	0.4468	0.1316	1.311	0.9974	0.4419	0.0359	1.4034
$\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$	0.9974	0.3297	0.0005	1.3266	0.9909	0.3681	0.0081	1.3509	0.9897	0.3812	0.0076	1.3633
$\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9951	0.8770	0.0157	**1.8564**	0.9973	0.8038	0.0147	**1.7864**	0.9973	0.8366	0.0107	**1.8232**
$\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9965	0.1424	0.0031	1.1358	0.9959	0.1545	0.0020	1.1484	0.9950	0.1544	0.0022	1.1472
$\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9887	0.6631	0.0775	1.5743	0.9974	0.7916	0.0807	1.7083	0.9968	0.8978	0.0725	1.8221
$\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9887	0.6593	0.0776	1.5704	0.9974	0.7842	0.0805	1.7011	0.9974	0.8748	0.0825	1.7897
$\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9974	0.7551	0.0000	1.7525	0.9974	0.6029	0.0087	1.5916	0.9974	0.4417	0.0095	1.4296
$\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9974	0.8617	0.0732	1.7859	0.9957	0.4161	0.0169	1.3949	0.9947	0.4128	0.0127	1.3948
$\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9961	0.1213	0.0020	1.1154	0.9961	0.1553	0.0021	1.1493	0.9955	0.1550	0.0023	1.1482
$\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9959	0.1542	0.0020	1.1481	0.9958	0.1542	0.0022	1.1478	0.9947	0.1223	0.0025	1.1145

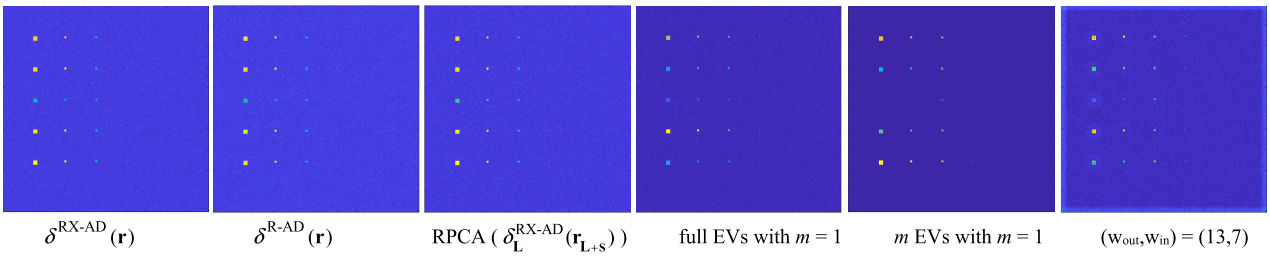


Fig. 7. Detection maps of RX/R-AD, RPCA, LSMAD, and CRD-DW-STO for TI.

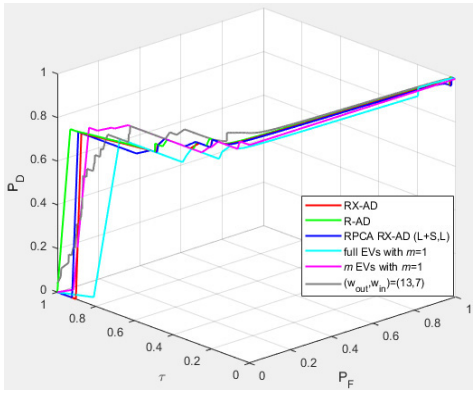


Fig. 8. 3-D ROC curves of Fig. 7.

to $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, which can be considered as the subpixel detectors and are, indeed, real anomaly detectors. Since the ROC analysis is designed to evaluate target detection, Table III was calculated based on all the 130 panel pixels considered as targets to be detected where the AUC values of their 2-D ROC curves of (P_D, P_F) of all six types of AD using $\mathbf{L}_m$ and $\mathbf{S}_j$ were better than those produced by the traditional RX/R-AD, but the AUC values of the 2-D ROC curves of (P_D, τ) and the 2-D ROC curves of (P_F, τ) were quite different. One most intriguing finding from Table III is the performance of the subpxiel/anomaly detectors: $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$. The AUC values of their 2-D ROC curves of (P_D, τ) were very low due to the fact that they could only detect the panel subpixels in the third, fourth, and fifth columns but missed all panel pixels in the first two columns. However, the AUC values of their 2-D ROC curves of (P_F, τ) were the smallest. This indicated that these detectors produced the smallest false-alarm probabilities meaning that they performed very well in BKG suppression. With such a very low P_D and the smallest P_F, the AUC values of their 2-D ROC curves of (P_D, P_F) were still higher than that produced by RX/R-AD. These results provide evidence that the classical 2-D ROC analysis was inapplicable to anomaly detection.

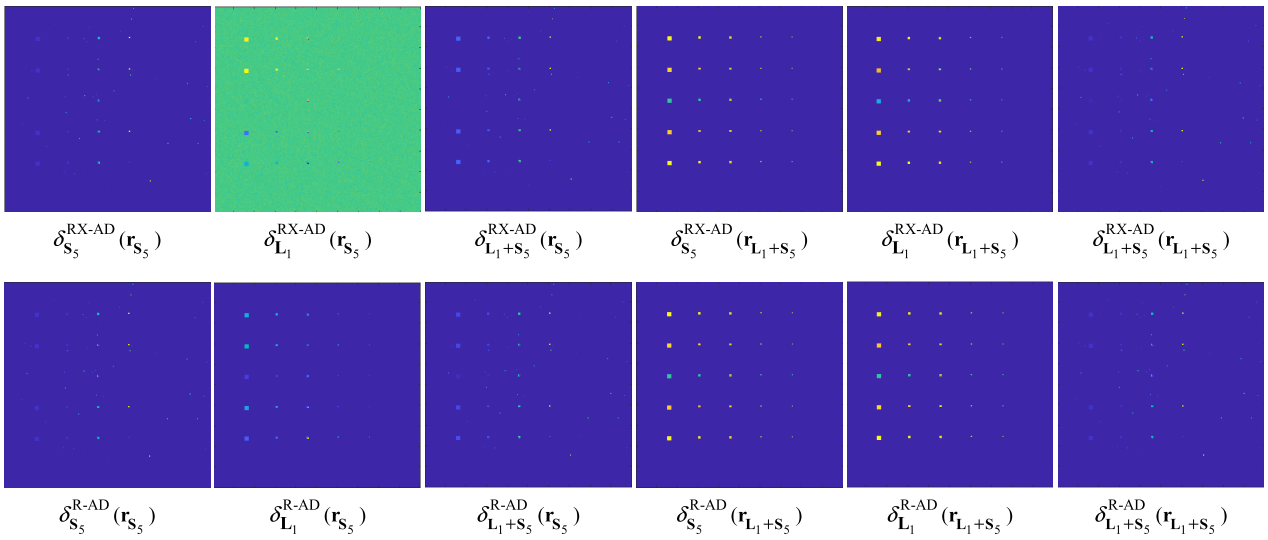


Fig. 9. TE: $p = 6$, $m = 1$, and $j = 5$.

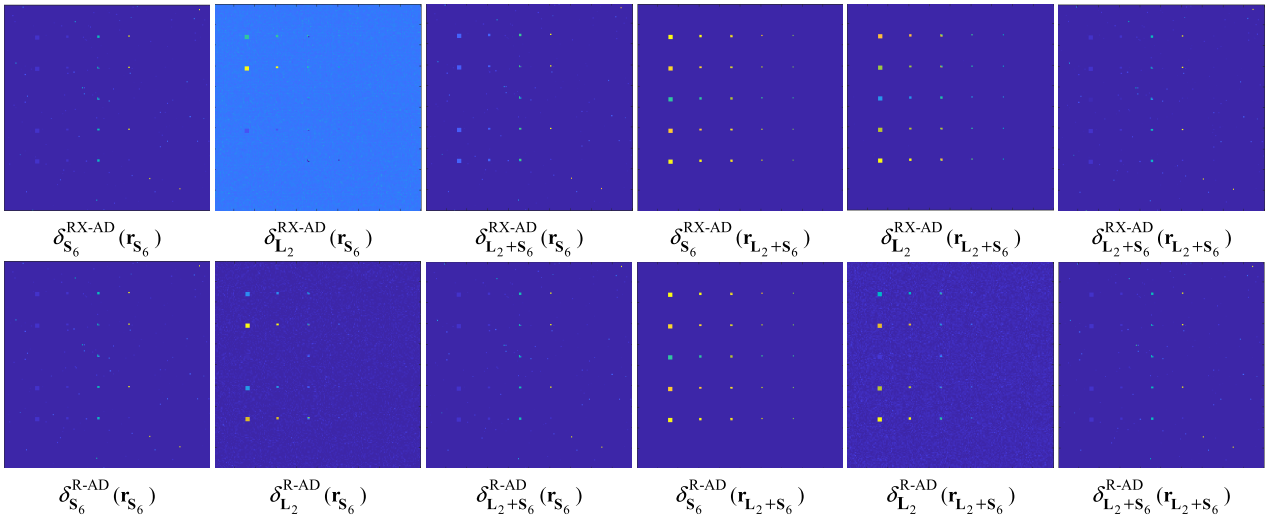


Fig. 10. TE: $p = 8$, $m = 2$, and $j = 6$.

To conduct a comparative analysis, we performed the traditional $\delta^{\text{RX/R-AD}}(\mathbf{r})$, RPCA, LSMAD using full EVs and m EVs with $p = 6$ and $m = 1$, and CRD-DW-STO using the size of outer window w_{out} and the size of the inner window w_{in}, which were specified by (13, 15, 17) and (1, 3, 5, 7), respectively, with 12 combinations of (w_{out}, w_{in}) and the sparsity parameter set to $\lambda = 10^{-6}$ implemented for experiments. Due to the limited space, we cannot include all the results but only the best results for each of the test cases in this article for comparison. Fig. 7 shows the detection-map results for TI using $\delta^{\text{RX/R-AD}}(\mathbf{r})$, LSMAD using full bands and m bands for the case $p = 6$ and $m = 1$, and CRD-DW-STO with (w_{out}, w_{in}) = (13, 7) along with their corresponding 3-D ROC curves shown in Fig. 8.

Table IV tabulates the AUC values of the 2-D ROC curves of (P_D, P_F), the 2-D ROC curves of (P_D, τ), the 2-D ROC curves of (P_F, τ) produced in Fig. 8, and their AUC_{OD} calculated by (43).

Like $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, which detected panel pixels in the first three columns but missed most or all panel subpixels in the fourth and fifth columns in Figs. 3–5, RX/R-AD, RPCA, LSMAD, and CRD-DW-STO could only detect panel pixels in the first three columns but failed to pick up the panel subpixels in the fourth and fifth columns in Fig. 7. Compared with Table III, the AUC_{OD} produced by RX/R-AD, RPCA,

TABLE IV
AUC OF RX/R-AD, RPCA, LSMAD, AND CRD-DW-STO FOR TI

			(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
RX-AD			0.9590	0.7665	0.0897	1.6358
R-AD			0.9590	0.8274	0.0960	1.6904
RPCA	$\delta^{\text{RX-AD}}_{\mathbf{L}}(\mathbf{r}_{\mathbf{L}+\mathbf{S}})$		0.9590	0.7752	0.0846	1.6497
LSMAD	p	m	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
full EVs	6	1	0.9652	0.6425	0.0337	1.5739
m EVs	6	1	0.9974	0.7829	0.0000	**1.7803**
CRD-DW -STO	w_{out}	w_{in}	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
	13	7	0.9962	0.7863	0.0503	1.7323

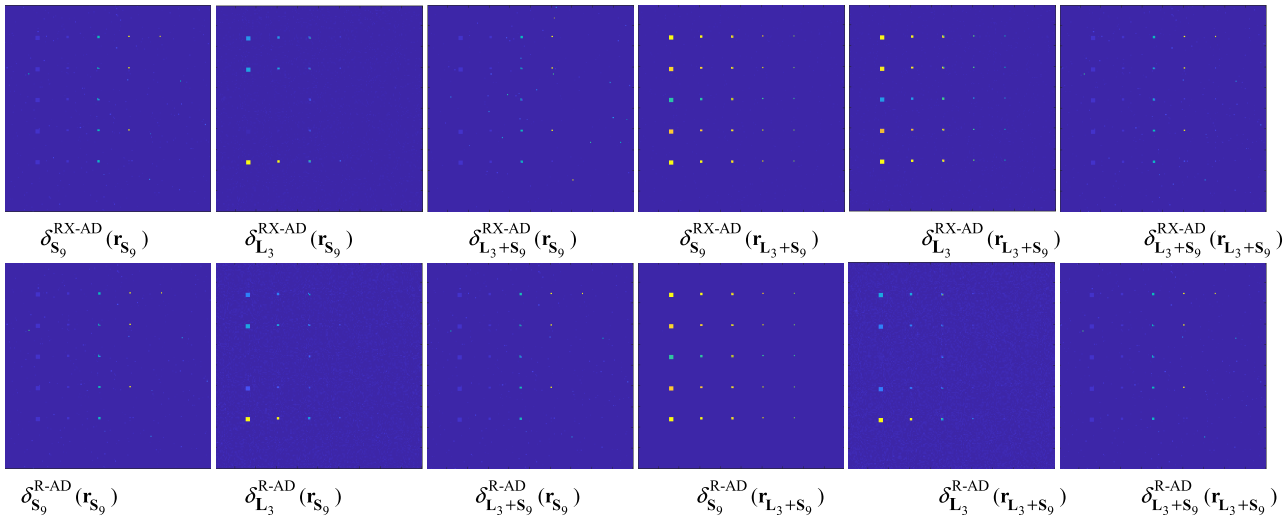
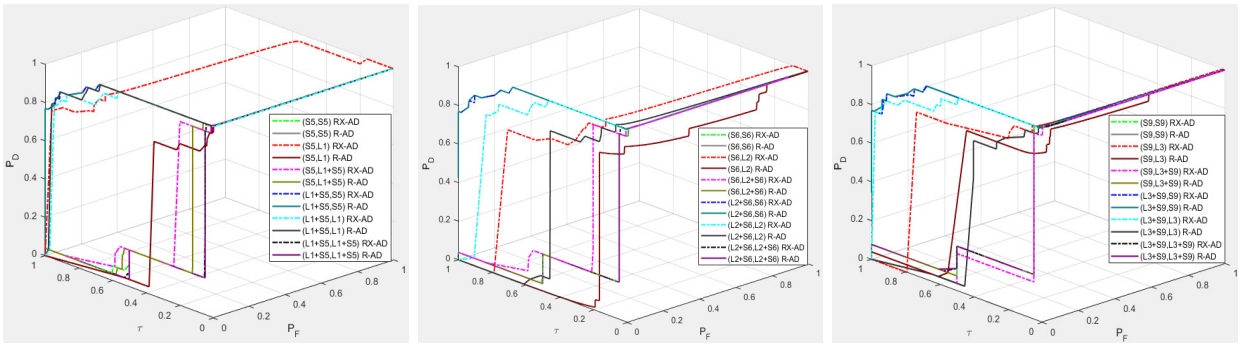


Fig. 11. TE: $p = 12$, $m = 3$, and $j = 9$.



$p = 6$, $m = 1$ and $j = 5$ $p = 8$, $m = 2$ and $j = 6$ $p = 12$, $m = 3$ and $j = 9$

Fig. 12. 3-D ROC curves of TE.

LSMAD, and CRD-DW-STO is far below the best results produced by OSP-GoDec in Table III for each of the three cases $(p, m, j) = (6, 1, 5)$, (13, 9, 4), and (19, 14, 5), which are 1.8564, 1.7864, and 1.8232, respectively.

B. TE Experiments

The same experiments conducted for TI were also performed for TE. Table V tabulates the value of p estimated by the MOCA and HFC/NWHFC methods and the value of m estimated by the MX-SVD for TE.

Figs. 9–11 show the detection maps produced by six types of ADs: $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, with $(p, m, j) = (6, 1, 5)$, (8, 2, 6), and (12, 3, 9) along with their corresponding 3-D ROC curves plotted in Fig. 12.

Table VI also tabulates the AUC values of the three 2-D ROC curves [2-D ROC curves of (P_D, P_F), 2-D ROC curves of (P_D, τ), and 2-D ROC curves of (P_F, τ)] obtained from the 3-D ROC curves in Fig. 12 along with their AUC_{OD} calculated by (43), where the best cases for various values of (p, m, j) are boldfaced.

Similar conclusion drawn for TI can be also applied to TE as well, that is, $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ detected panel pixels in the first three columns but missed most or all panel subpixels in the fourth and fifth columns but $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ detected the mixed-panel pixels in the third column and the panel subpixels in the fourth and fifth columns but missed all panel pixels in the first two columns. Interestingly, unlike TI experiments, the best results from Figs. 10 and 11 and Table VI were produced by using $(p, m, j) = (6, 1, 5)$. This is because TE is designed for the signal-detection model, where the target panels were superimposed over the BKG pixels. In this case, the target panel pixels are more pronounced than the panel pixels in the TI.

TABLE V
m ESTIMATED FOR TE BY MOCA AND HFC/NWHFC USING MX-SVD

	p	m by MX-SVD
MOCA	12	3
HFC, $P_F \le 10^{-3}$	8	2
NWHFC $P_F \le 10^{-3}$	6	1

Like TI, we also performed the traditional $\delta^{\text{RX/R-AD}}(\mathbf{r})$, RPCA, LSMAD using full EVs and m EVs with $p = 6$ and $m = 1$, and CRD-DW-STO using the size of the outer window w_{out} and the size of the inner window w_{in}, which were specified by (13, 15, 17) and (1, 3, 5, 7), respectively, with 12 combinations of $(w_{\text{out}}, w_{\text{in}})$ implemented for experiments. Fig. 13 shows the best detection-map results for TI using $\delta^{\text{RX/R-AD}}(\mathbf{r})$, RPCA, LSMAD using full EVs and m EVs for

TABLE VI
COMPARISON OF VARIOUS ANOMALY DETECTORS FOR TE USING AUC VALUES CALCULATED BY (43)

	p =6, m = 1 and j = 5				p =8, m = 2 and j = 6				p =12, m = 3 and j =9			
Detector	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
$\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9959	0.1586	0.0020	1.1525	0.9953	0.1542	0.0022	1.1473	0.9961	0.1600	0.0023	1.1538
$\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9959	0.1542	0.0020	1.1481	0.9951	0.1205	0.0022	1.1134	0.9958	0.1600	0.0023	1.1535
$\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$	0.9590	0.9052	0.5953	1.2689	0.9965	0.6441	0.1161	1.5245	0.9974	0.6788	0.0194	1.6568
$\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$	0.9974	0.3252	0.0007	1.3219	0.8676	0.1687	0.0411	0.9952	0.9649	0.4267	0.0148	1.3768
$\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9969	0.2730	0.0021	1.2678	0.9968	0.2611	0.0020	1.2559	0.9959	0.1542	0.0020	1.1481
$\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9968	0.1885	0.0025	1.1828	0.9951	0.1207	0.0022	1.1136	0.9959	0.1603	0.0023	1.1539
$\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9887	0.8486	0.0036	1.8337	0.9948	0.9297	0.0055	1.9190	0.9944	0.9196	0.0100	1.9040
$\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9973	0.9638	0.0036	**1.9575**	0.9955	0.9401	0.0051	**1.9305**	0.9960	0.9472	0.0073	**1.9359**
$\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9974	0.9180	0.0000	1.9154	0.9974	0.8284	0.0030	1.8228	0.9974	0.9306	0.0072	1.9208
$\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	**0.9974**	0.9408	0.0030	1.9352	0.9971	0.4205	0.0233	1.3943	0.9969	0.3870	0.0145	1.3694
$\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9963	0.1565	0.0020	1.1508	0.9956	0.1555	0.0022	1.1489	0.9959	0.1614	0.0022	1.1551
$\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9959	0.1542	0.0020	1.1481	0.9951	0.1203	0.0022	1.1132	0.9959	0.1621	0.0023	1.1557

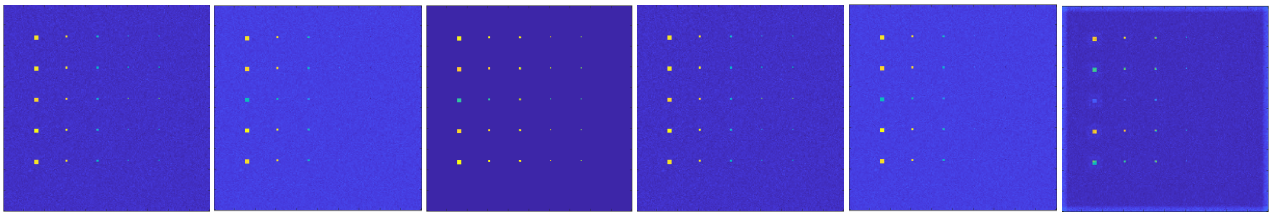


Fig. 13. Detection maps of RX/R-AD, RPCA, LSMAD, and CRD-DW-STO for TE.

the case $p = 6$ and $m = 1$, and CRD-DW-STO with $(w_{\text{out}}, w_{\text{in}}) = (13, 7)$ along with their corresponding 3-D ROC curves shown in Fig. 14.

Table VII tabulates the AUC values of the 2-D ROC curves of (P_D, P_F), 2-D ROC curves of (P_D, τ), and 2-D ROC curves of (P_F, τ) produced in Fig. 14 and their AUC_{OD} calculated by (43).

Like $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, which detected panel pixels in the first three columns but missed most or all panel subpixels in the fourth and fifth columns in Figs. 9–11, RX/R-AD, LSMAD, and CRD-DW-STO could only detect panel pixels in the first three columns but failed to pick up the panel subpixels in the fourth and fifth columns in Fig. 13. According to Table VII, the AUC_{OD} produced by RX/R-AD, LSMAD, and CRD-DW-STO are far below the best results produced by OSP-GoDec for each of three cases, $(p, m, j) = (6, 1, 5)$, $(8, 2, 6)$, and $(12, 3, 9)$, which are 1.9575, 1.9305, and 1.9359, respectively, and RPCA is 1.9369.

IX. REAL IMAGES TO BE STUDIED

Three real HSIs were used for experiments.

A. HYDICE 15-Panel Scene

The image scene shown in Fig. 15 was acquired by the airborne hyperspectral digital imagery collection experiment (HYDICE) sensor in August 1995 from a flight altitude of 10 000 ft. This scene has been studied extensively by many reports such as [25] and [26]. There are 15 square panels with three different sizes, 3 m × 3 m, 2 m × 2 m, and 1 m × 1 m, respectively, as shown in Fig. 15(a). Due to the approximately 1.56-m ground sampling distance, each of the panels in the first column except the first row contains two panel pixels highlighted by red, p_{211} and p_{221} in row 2, p_{311} and p_{312} in row 3, p_{411} and p_{412} in row 4, and p_{511} and p_{521} in row 5, as shown in Fig. 15(b), by the ground truth. All the 11 remaining panels in Fig. 15(b) contain one single-panel pixel for each panel also highlighted by red, p_{11}, p_{12}, and p_{13} in

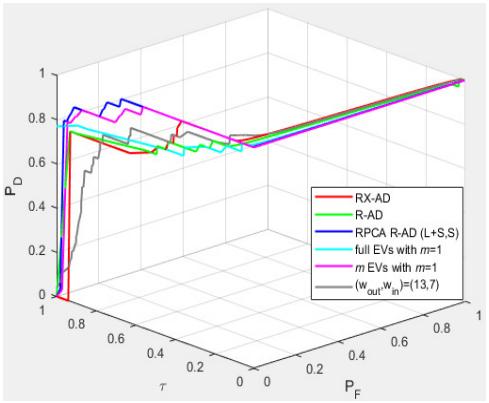


Fig. 14. 3-D ROC curves of Fig. 13.

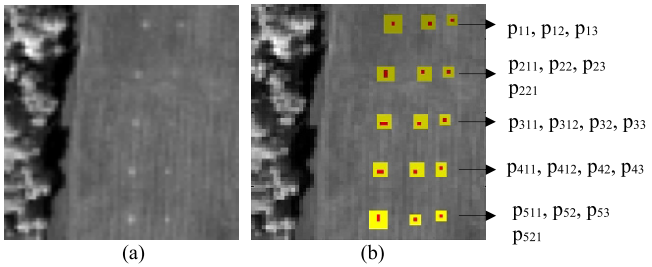


Fig. 15. (a) HYDICE panel scene which contains 15 panels. (b) Ground-truth map of spatial locations of the 15 panels.

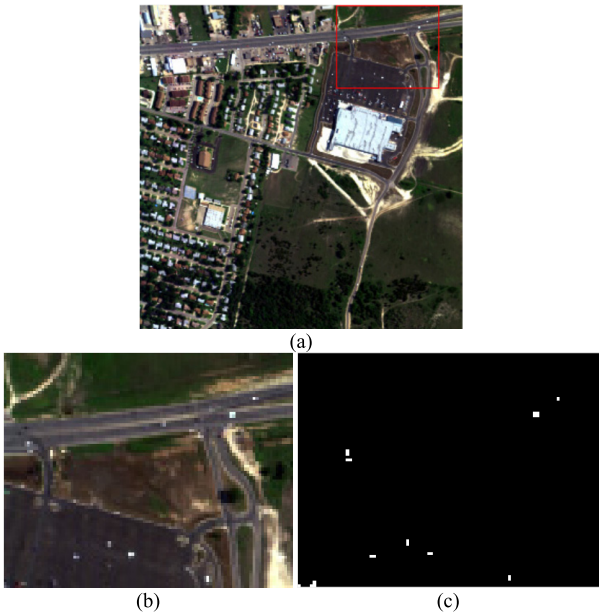


Fig. 16. HYDICE urban scene. (a) Pseudocolor image of the whole scene. (b) Pseudocolor image of the selected area. (c) Ground-truth map.

row 1, p_{22} and p_{23} in row 2, p_{32} and p_{33} in row 3, p_{42} and p_{43} in row 4, and p_{52} and p_{53} in row 5. Therefore, there are a total of 19 panel pixels. Fig. 15(b) shows their precise spatial locations with the pixels in yellow (Y pixels) indicating panel pixels mixed with the BKG.

B. HYDICE Urban Scene

A second data set is another HYDICE image with pseudocolor shown in Fig. 16(a). It is an urban scene and comprised 210 spectral bands with 174 bands being used for experiments

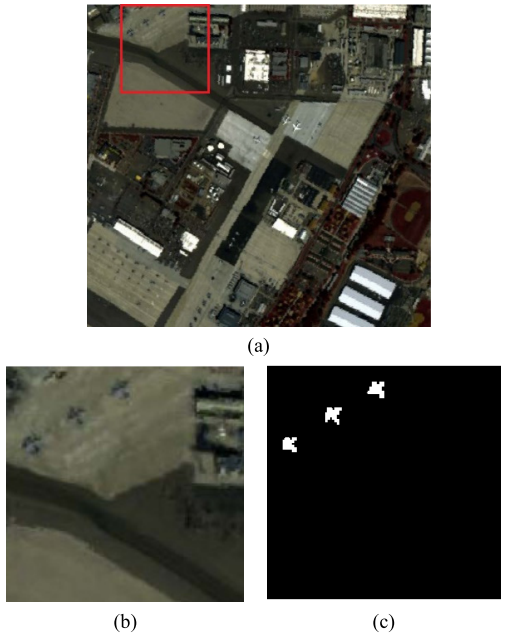


Fig. 17. AVIRIS San Diego Airport scene. (a) Pseudocolor image of the whole scene. (b) Pseudocolor image of the selected area. (c) Ground-truth map.

after the noise and water-absorption bands had been removed. A region with a size of 80 × 100 pixels located at the top right of the scene was selected as the test image shown in Fig. 16(b) along with the ground-truth map shown in Fig. 16(c), where 21 pixels were identified as anomalies, which were cars and roofs, because they had spectra that differ from the BKG.

C. AVIRIS San Diego Airport Scene

A third data set is an AVIRIS airport scene from San Diego, CA, USA, with pseudocolor shown in Fig. 17(a). It has a size of 400 × 400 pixels with a 3.5-m spatial resolution and 224 spectral channels in wavelengths ranging from 370 to 2510 nm. After removing the bands that correspond to the water-absorption regions, low SNR, and bad bands (1–6, 33–35, 97, 107–113, 153–166, and 221–224), 189 available bands of the data were retained in the experiments. An area of 100 × 100 pixels at the top-left corner of the scene was selected as the test image shown in Fig. 17(b) along with its ground-truth map shown in Fig. 17(c). It is an urban scene in which the main BKG materials are roof, shadow, and grass. There are three airplanes in the image, which consist of 85 pixels and account for 0.33% of the test image, as shown in Fig. 16(b). The spectral signatures of these planes are different from those of the main BKG, as shown in Fig. 17(c), and these pixels were, therefore, selected as the anomalies to be detected.

First, we calculated the value of p estimated by the MOCA and HFC/NWHFC methods and the value of m estimated by MX-SVD for the HYDICE 15-panel and the HYDICE urban and San Diego Airport scenes tabulated in Table VIII.

X. Experiments and Discussion

The experiments presented in this section consist of three parts. The first part conducts a comprehensive study and

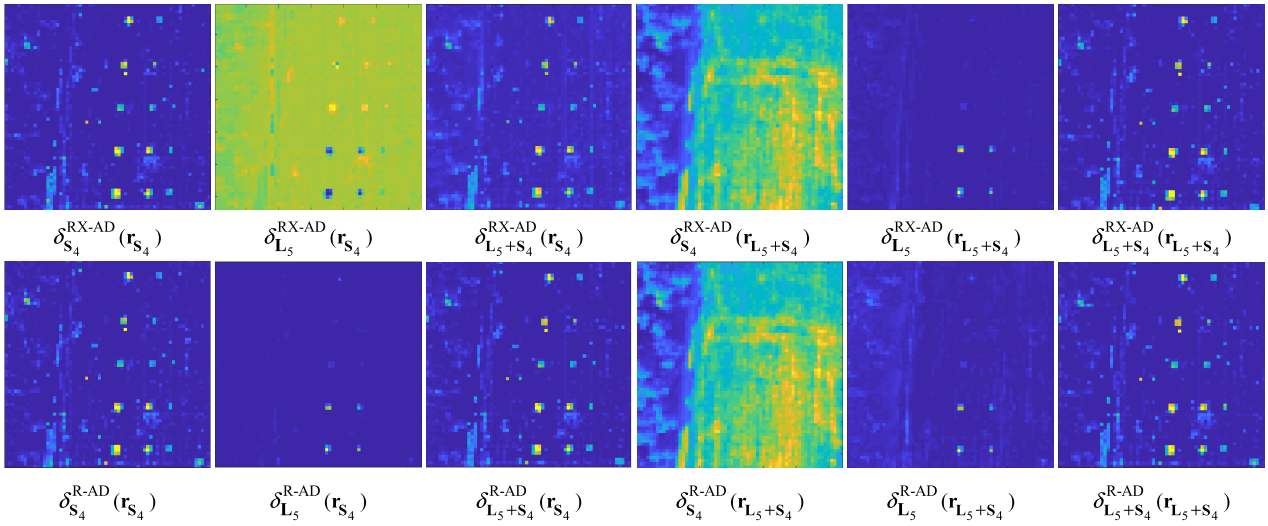


Fig. 18. HYDICE 15 panel scene: $p = 9$, $m = 5$, and $j = 4$.

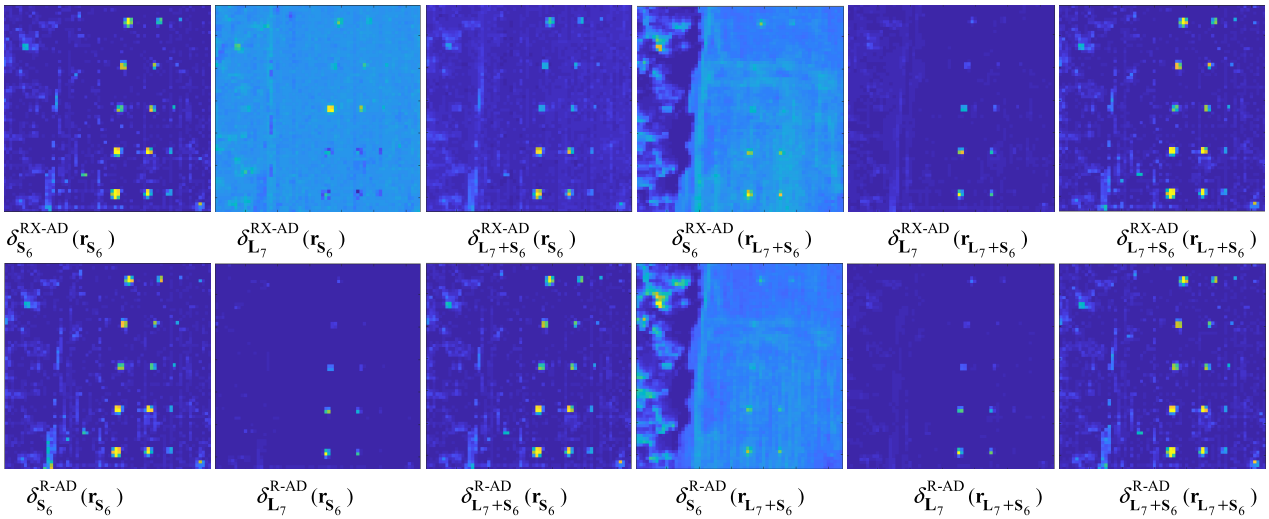


Fig. 19. HYDICE 15 panel scene: $p = 13$, $m = 7$, and $j = 6$.

TABLE VII

AUC OF RX/R-AD, RPCA, LSMAD, AND CRD-DW-STO FOR TE

			(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
RX-AD			0.9974	0.8231	0.0492	1.7713
R-AD			0.9590	0.8272	0.0944	1.6919
RPCA	$\delta^{\text{R-AD}}_{\mathbf{s}}(\mathbf{r}_{\mathbf{L}+\mathbf{S}})$		0.9974	0.9424	0.0029	**1.9369**
LSMAD	p	m	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
full EVs	6	1	0.9887	0.4092	0.0074	1.3904
m EVs	6	1	0.9974	0.9186	0.0000	1.9159
CRD-DW-STO	w_{out}	w_{in}	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
	13	7	0.9964	0.7896	0.0482	1.7378

analysis of six types of anomaly detectors derived from OSP-GoDec for the three real HSI scenes with great details. The second part is a comparative analysis of the first part results with the best results produced by the traditional RX/R-AD, RPCA, LSMAD, and CRD-DW-STO. Finally, the third part presents the discussions of the conclusions drawn from the experimental results.

A. OSP-GoDec HYDICE 15-Panel Scene

Now, we implemented the $\mathbf{L}_m + \mathbf{S}_j + \mathbf{N}$ model using OSP-GoDec. Using the results in Table VIII, Figs. 18–20 show the detection maps produced by six types of each $\delta^{\text{RX/R-AD}}(\mathbf{r})$, i.e., $\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ with $(p, m, j) = (9, 5, 4)$, $(13, 7, 6)$, and $(32, 24, 8)$. As shown in Figs. 18–20, $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$ and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ detected a partial set of panel pixels in the first two columns but missed all panel subpixels in the third column. The worst performance was produced by $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$. Interestingly, $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ could detect all 19 panel pixels including five panel subpixels in the third column at the expense of high falsely detected BKG pixels. As the value of p was increased, the false-alarm rate was significantly reduced, as shown in Figs. 18–20, but it may also sacrifice the detection of the panel subpixels, as also shown in Figs. 18–20. By the visual inspection of Figs. 18–20, the best one was produced

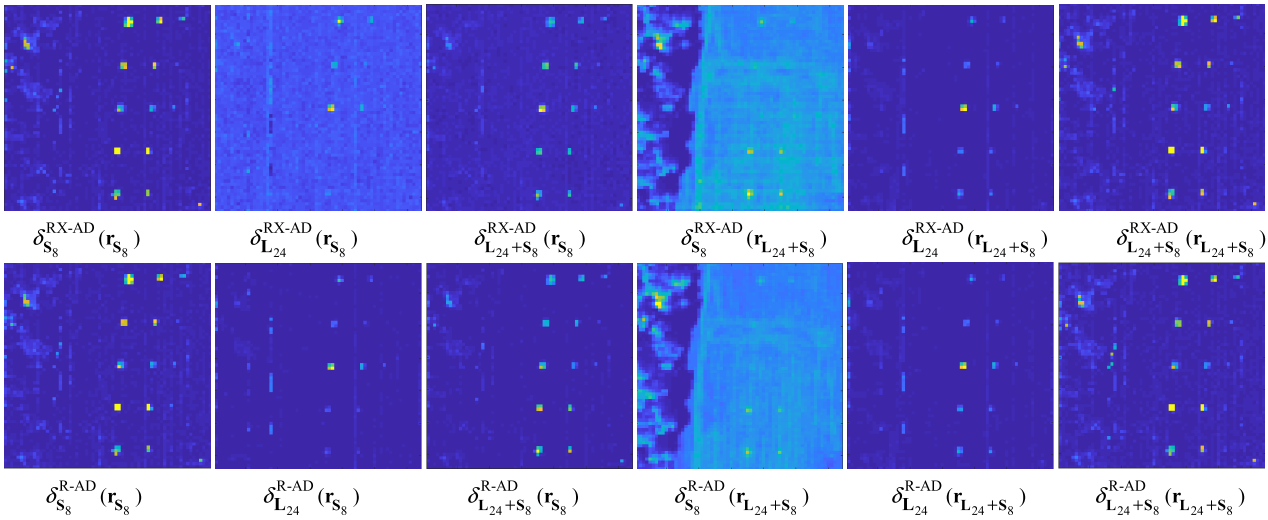
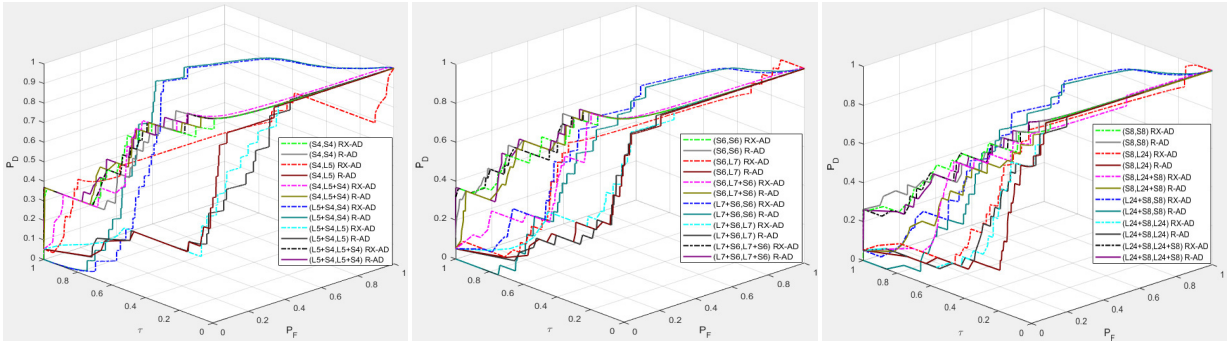


Fig. 20. HYDICE 15 panel scene: $p = 32$, $m = 24$, and $j = 8$.



p =9, m = 5 and j = 4 p =13, m = 7 and j = 6 p =32, m = 24 and j =8

Fig. 21. 3-D ROC curves of the HYDICE 15-panel scene.

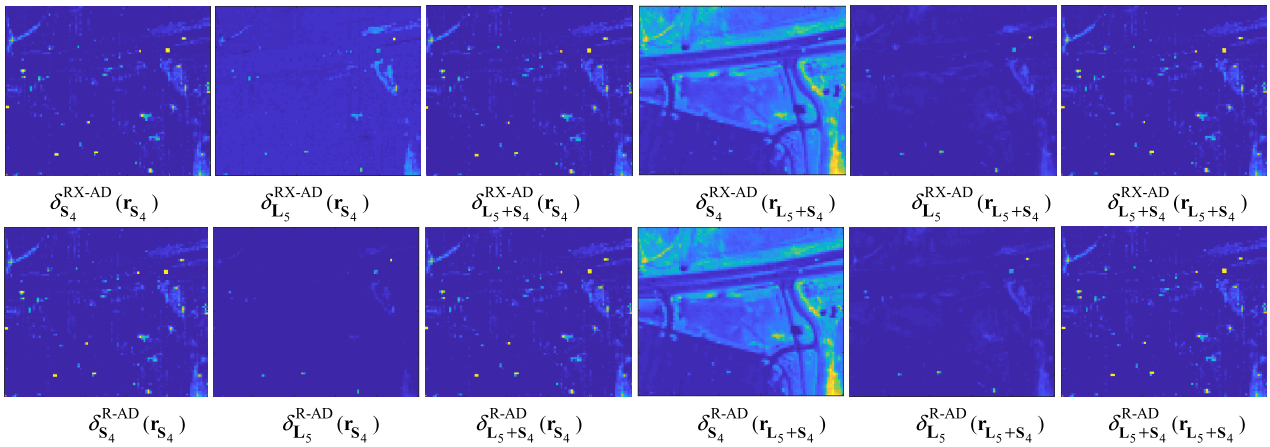


Fig. 22. HYDICE urban: $p = 9$, $m = 5$, and $j = 4$.

by $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ using $p = 13$, $m = 7$, and $j = 6$ in Fig. 19.

To perform further a quantitative analysis, the 3-D ROC analysis was conducted, where the 3-D ROC curves of Figs. 18–20 were plotted in Fig. 21 along with Table IX that tabulates the AUC values of their corresponding 2-D ROC curves [2-D ROC curves of (P_D, P_F), 2-D ROC curves of (P_D, τ), and 2-D ROC curves of (P_F, τ)] and AUC_{OD}, where the best results for each case of (p, m, j) are boldfaced. Among the best cases was $AUC_{OD} = 1.6877$ produced by $\boldsymbol{\delta}^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ using $(p, m, j) = (13, 7, 6)$.

It is interesting to note the results from Figs. 18–20 and Table IX that $\boldsymbol{\delta}^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ was the best anomaly detector to detect the anomalies, but $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ was the best detector to extract the BKG. In other words, when $\mathbf{S}$ was used to specify $\mathbf{B}$ and $\mathbf{A}$ in an anomaly detector $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{B}}(\mathbf{r}_{\mathbf{A}})$ as BKG and the targets to be detected, respectively, the detector could detect anomalies effectively. Conversely, if $\mathbf{S}$ and $\mathbf{L} + \mathbf{S}$

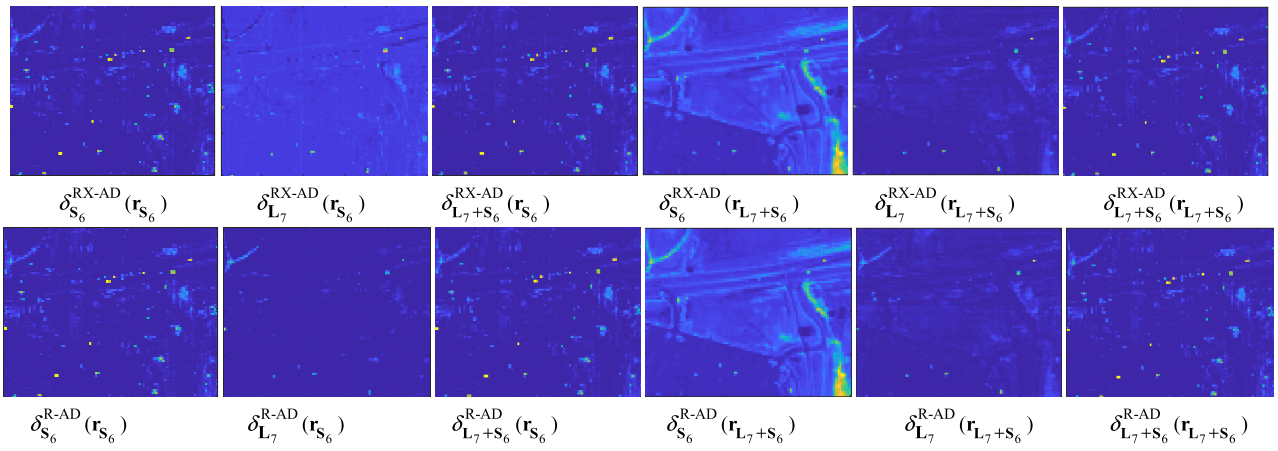


Fig. 23. HYDICE urban: $p = 13$, $m = 7$, and $j = 6$.

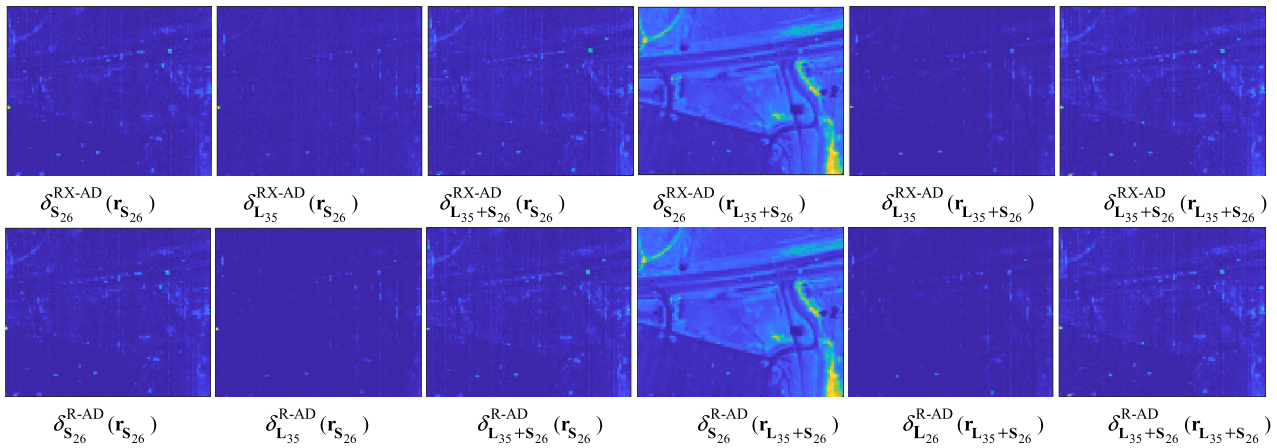
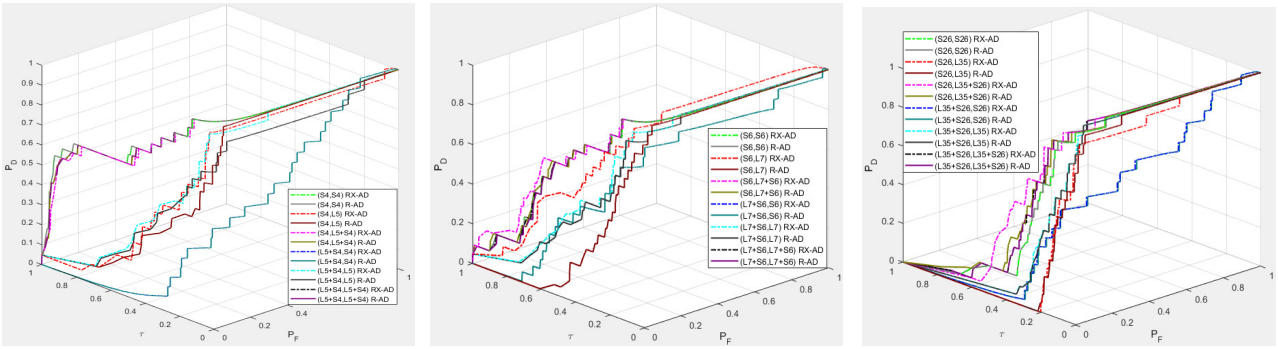


Fig. 24. HYDICE urban: $p = 61$, $m = 35$, and $j = 26$.



p =9, m = 5 and j = 4 p =13, m = 7 and j = 6 p =61, m = 35 and j =26

Fig. 25. 3-D ROC curves of the HYDICE urban scene.

were used to specify $\mathbf{B}$ in an anomaly detector $\delta^{\text{RX/R-AD}}_{\mathbf{B}}(\mathbf{r}_{\mathbf{A}})$ as BKG suppression and $\mathbf{A}$ as the targets to be detected, respectively, the detector could detect BKG as shown in Figs. 18–20.

HYDICE Urban Scene: Figs. 22–24 show the detection maps produced by six types of each $\delta^{\text{RX/R-AD}}(\mathbf{r})$, i.e., $\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ and $\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ with $(p,\ m,\ j) = (9,\ 5,\ 4)$, $(13,\ 7,\ 6)$, and $(61,\ 35,\ 26)$, where the detectors that performed well on anomaly detection should specify $\mathbf{B}$ by $\mathbf{L}_m + \mathbf{S}_j$ for the best possible BKG suppression compared with the detectors that should specify $\mathbf{A}$ by $\mathbf{L}_m + \mathbf{S}_j$ to have better BKG extraction.

For a quantitative analysis, the 3-D ROC analysis was conducted, where the 3-D ROC curves of Figs. 22–24 were plotted in Fig. 25 along with Table X that tabulates the AUC values of their corresponding 2-D ROC curves [2-D ROC curves of (P_D, P_F), 2-D ROC curves of (P_D, τ), and 2-D

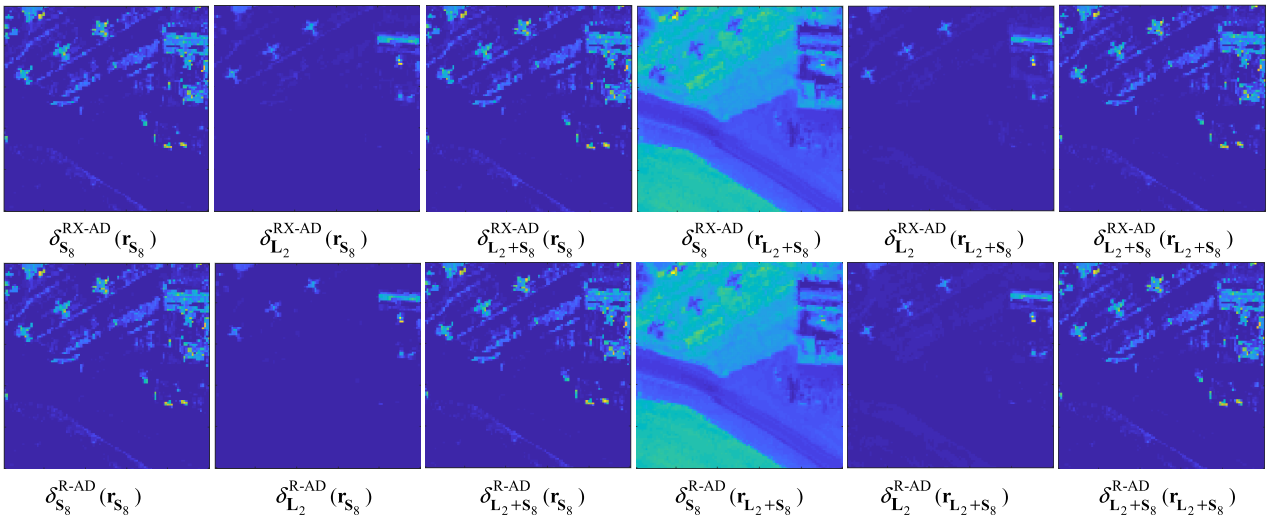


Fig. 26. San Diego Airport: $p = 10$, $m = 2$, and $j = 8$.

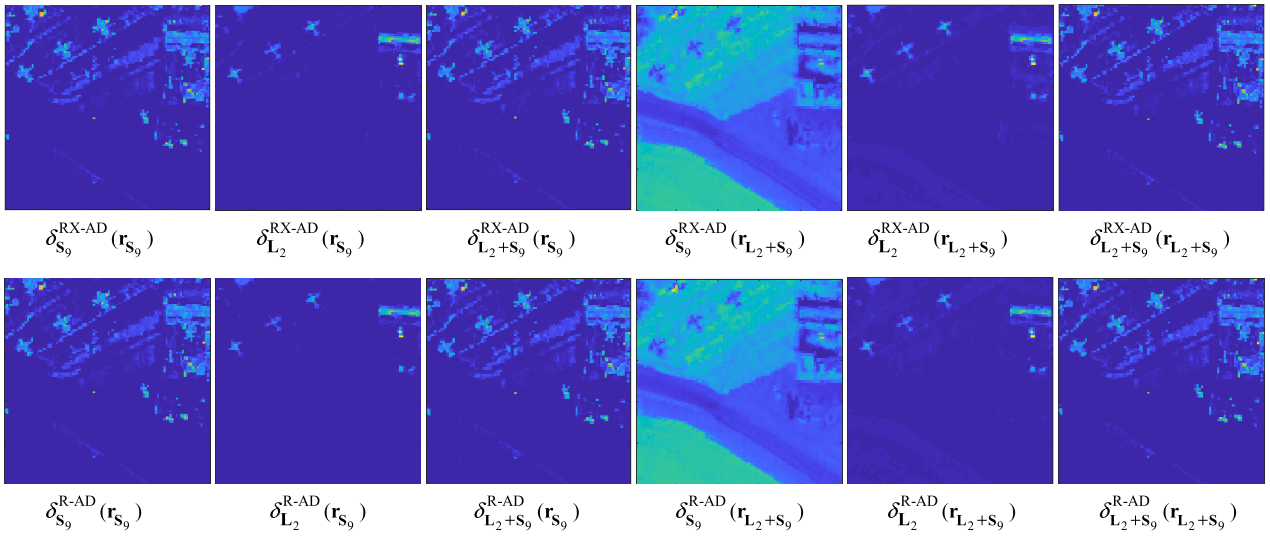


Fig. 27. San Diego Airport: $p = 11$, $m = 2$, and $j = 9$.

TABLE VIII

m ESTIMATED FOR HYDICE DATA BY MOCA AND HFC/NWHFC USING MX-SVD

Data	VD	p	m by MX-SVD
HYDICE 15-panel	MOCA	32	24
	HFC, $P_F \le 10^{-3}$	9	5
	NWHFC $P_F \le 10^{-3}$	13	7
HYDICE Urban	MOCA	61	35
	HFC, $P_F = 10^{-4}$	9	5
	NWHFC $P_F = 10^{-4}$	13	7
San Diego Airport	MOCA	29	16
	HFC, $P_F \le 10^{-3}$	10	2
	NWHFC $P_F \le 10^{-3}$	11	2

ROC curves of (P_F, τ)] and AUC_{OD}, where the best results for each case of (p, m, j) are boldfaced. Among the best cases was $AUC_{OD} = 1.7019$ produced by $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ with $(p, m, j) = (9, 5, 4)$.

Unlike the HYDICE 15-panel scene that showed that $\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ was the best detector, according to Figs. 22–24 and Table X, $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ and $\delta^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ were the best detectors to detect anomalies but $\delta^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ was the best detector to extract the BKG. In other words, when $\mathbf{L}_m + \mathbf{S}_j$ and $\mathbf{S}_j$ were used to specify $\mathbf{B}$ in an anomaly detector $\delta^{\text{RX/R-AD}}_{\mathbf{B}}(\mathbf{r}_{\mathbf{A}})$ as BKG suppression and $\mathbf{A}$ as the targets to be detected, respectively, the detector could detect the anomalies effectively. Conversely, if $\mathbf{S}_j$ and $\mathbf{L}_m + \mathbf{S}_j$ were used to specify $\mathbf{B}$ as BKG suppression and $\mathbf{A}$ as the targets to be detected, respectively, the detector could detect BKG effectively.

AVIRIS San Diego Airport Scene: Figs. 26–28 show the detection maps produced by six types of each $\delta^{\text{RX/R-AD}}(\mathbf{r})$, i.e., $\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, $\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ and $\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, $\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$, and

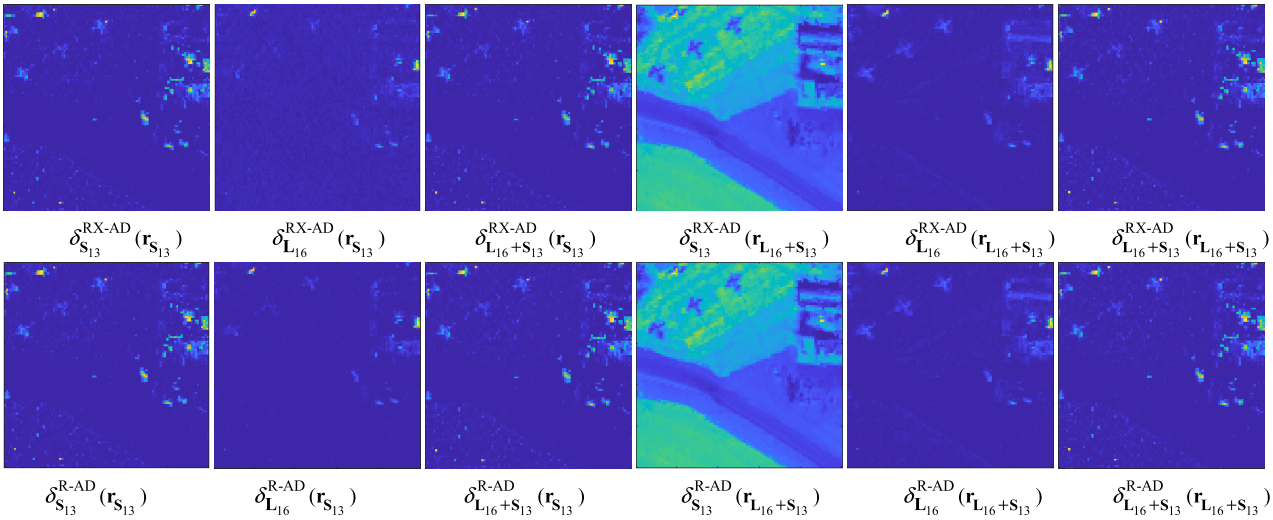


Fig. 28. San Diego Airport: $p = 29$, $m = 16$, and $j = 13$.

TABLE IX
AUC OF HYDICE 15-PANEL SCENE

	p =9, m = 5 and j = 4				p =13, m = 7 and j = 6				p =32, m = 24 and j =8			
Detector	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
$\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9906	0.6760	0.0271	1.6395	0.9952	0.6986	0.0280	1.6658	0.9870	0.6173	0.0258	**1.5785**
$\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9927	0.6366	0.0280	1.6013	0.9956	0.7221	0.0300	**1.6877**	0.9899	0.5976	0.0258	1.5617
$\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$	0.4946	0.5459	0.7006	0.3399	0.7731	0.4528	0.3297	0.8962	0.8667	0.3274	0.1551	1.0390
$\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$	0.9642	0.2025	0.0016	1.1651	0.9789	0.2293	0.0027	1.2055	0.9802	0.1900	0.0104	1.1598
$\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9934	0.6310	0.0559	1.5685	0.9952	0.4947	0.0467	1.4432	0.9646	0.4485	0.0327	1.3804
$\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9927	0.6689	0.0290	1.6326	0.9960	0.6819	0.0285	1.6494	0.9894	0.4393	0.0131	1.4156
$\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.7717	0.6472	0.4692	0.9497	0.9338	0.5126	0.2551	1.1913	0.8954	0.5227	0.2795	1.1386
$\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.8160	0.6838	0.4816	1.0182	0.9221	0.4492	0.2422	1.1291	0.9107	0.4622	0.2582	1.1147
$\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9050	0.2253	0.0116	1.1187	0.9683	0.2898	0.0150	1.2431	0.9832	0.2407	0.0134	1.2105
$\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.8570	0.2235	0.0260	1.0545	0.9791	0.2236	0.0092	1.1935	0.9755	0.2469	0.0133	1.2091
$\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9927	0.6628	0.0288	1.6267	0.9957	0.7105	0.0300	1.6762	0.9911	0.5837	0.0277	1.5471
$\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9935	0.6801	0.0283	**1.6453**	0.9960	0.7122	0.0302	1.6780	0.9921	0.5784	0.0273	1.5432

$\boldsymbol{\delta}^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ with $(p, m, j) = (10, 2, 8)$, $(11, 2, 9)$, and $(29, 16, 13)$.

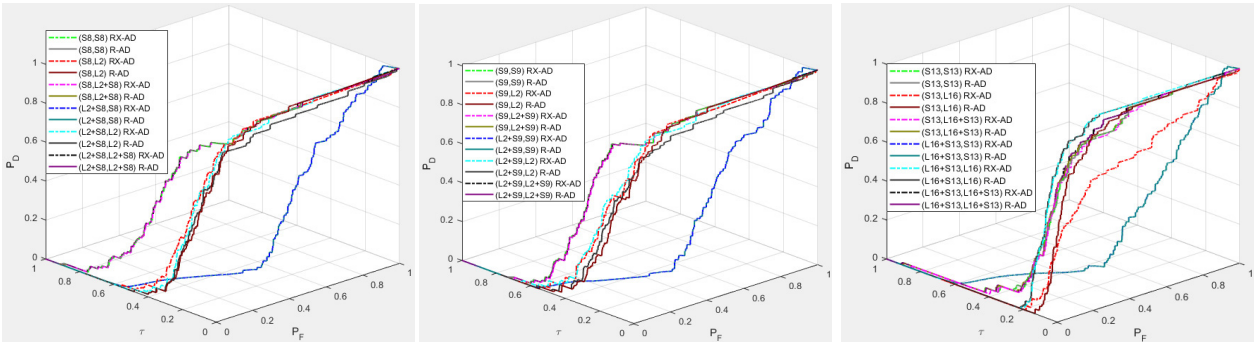
Unlike Figs. 22–24, for a detector to extract airplanes, an anomaly detector $\boldsymbol{\delta}_{\mathbf{B}}(\mathbf{r}_{\mathbf{A}})$ should specify $\mathbf{A}$ by $\mathbf{L}_m$ for the best possible BKG suppression compared with the detectors that should specify $\mathbf{A}$ by $\mathbf{L}_m + \mathbf{S}_j$ and $\mathbf{B}$ by $\mathbf{S}_j$ to have better BKG extraction, as shown in Figs. 26–28.

To perform further a quantitative analysis, the 3-D ROC analysis was conducted, where the 3-D ROC curves of Figs. 26–28 were plotted in Fig. 29 along with Table XI that tabulates the AUC values of their corresponding 2-D ROC curves [2-D ROC curves of (P_D, P_F), 2-D ROC curves of (P_D, τ), and 2-D ROC curves of (P_F, τ)] and AUC_{OD}, where the best results are boldfaced. Among the best cases was $AUC_{OD} = 1.2628$ produced by $\boldsymbol{\delta}^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ with $(p, m, j) = (11, 2, 9)$.

Similar to the experiments conducted for the HYDICE urban scene, the same interesting observations can also be made from Figs. 26–28 and Table XI, where $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ and $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ were the best anomaly detectors to detect

TABLE X
AUC OF HYDICE URBAN SCENE

	p =9, m = 5 and j = 4				p =13, m = 7 and j = 6				p =61, m = 35 and j =26			
Detector	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
$\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9958	0.7232	0.0195	1.6995	0.9954	0.5746	0.0191	1.5509	0.9843	0.2752	0.0263	1.2331
$\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9958	0.7232	0.0195	1.6994	0.9952	0.5533	0.0192	1.5293	0.9839	0.2752	0.0264	1.2328
$\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$	0.9374	0.2972	0.0538	1.1808	0.9851	0.4648	0.0922	1.3577	0.9409	0.1085	0.0217	1.0276
$\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{S}_j})$	0.9890	0.2434	0.0014	1.2310	0.9919	0.2265	0.0044	1.2141	0.9720	0.1009	0.0113	1.0617
$\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9955	0.7043	0.0200	1.6798	0.9959	0.6052	0.0205	**1.5807**	0.9888	0.3419	0.0346	**1.2961**
$\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$	0.9957	0.7126	0.0198	1.6885	0.9953	0.5576	0.0196	1.5333	0.9846	0.3170	0.0318	1.2699
$\delta^{\text{RX-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.5433	0.2002	0.1989	0.5445	0.8292	0.3591	0.1180	1.0703	0.6396	0.1949	0.1533	0.6813
$\delta^{\text{R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.5433	0.2002	0.1989	0.5445	0.8292	0.3591	0.1180	1.0703	0.6370	0.1932	0.1535	0.6768
$\delta^{\text{RX-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9794	0.3161	0.0151	1.2803	0.9881	0.3632	0.0215	1.3297	0.9878	0.1759	0.0145	1.1492
$\delta^{\text{R-AD}}_{\mathbf{L}_m}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9116	0.2944	0.0149	1.1911	0.9838	0.3432	0.0224	1.3046	0.9867	0.1726	0.0143	1.1451
$\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9960	0.7259	0.0201	**1.7019**	0.9956	0.5743	0.0198	1.5501	0.9891	0.3048	0.0346	1.2593
$\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9960	0.7257	0.0199	1.7017	0.9956	0.5707	0.0196	1.5466	0.9889	0.3040	0.0345	1.2585



p =10, m = 2 and j =8 p =11, m = 2 and j = 9 p =29, m = 16 and j =13

Fig. 29. 3-D ROC curves of the San Diego Airport scene.

the anomalies but $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{S}_j}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ was the best detector to extract the BKG. In other words, when $\mathbf{L}_m + \mathbf{S}_j$ and $\mathbf{S}_j$ were used to specify $\mathbf{B}$ in an anomaly detector $\boldsymbol{\delta}^{\text{RX/R-AD}}_{\mathbf{B}}(\mathbf{r}_{\mathbf{A}})$ as BKG suppression and $\mathbf{A}$ as the targets to be detected, respectively, the detector could detect anomalies effectively. Conversely, if $\mathbf{S}_j$ and $\mathbf{L}_m + \mathbf{S}_j$ were used to specify $\mathbf{B}$ in an anomaly detector $\boldsymbol{\delta}_{\mathbf{B}}(\mathbf{r}_{\mathbf{A}})$ as BKG suppression and $\mathbf{A}$ as the targets to be detected, respectively, the detector could detect BKG effectively.

B. Comparison With RX/R-AD, RPCA, LSMAD, and CRD-DW-STO

To conduct a comparative analysis, we performed the traditional $\boldsymbol{\delta}^{\text{RX/R-AD}}(\mathbf{r})$, RPCA, LSMAD using full EVs and m EVs with $p = 24$ and $m = 13$, and CRD-DW-STO using the size of the outer window w_{out} and the size of the inner window w_{in}, specified by (11, 13, 15) and (3, 5, 7, 9), respectively, with 12 combinations of (w_{out}, w_{in}) implemented for the HYDICE 15-panel and the urban scenes except the San Diego Airport scene for which we used the size of the outer window w_{out} and the size of the inner window w_{in} specified by (15, 17, 19) and (3, 5, 7, 9), respectively, with 12 combinations of (w_{out}, w_{in}) for the experiments. Due to the limited space, we cannot include all the results but only the best scenarios for each of test cases in this article for comparison. Figs. 30–32 show the best cases of the detection-map results for the HYDICE 15-panel and the HYDICE urban and San Diego Airport scene using $\boldsymbol{\delta}^{\text{RX/R-AD}}(\mathbf{r})$, RPCA, LSMAD using full bands and m bands, and CRD-DW-STO along with their the corresponding

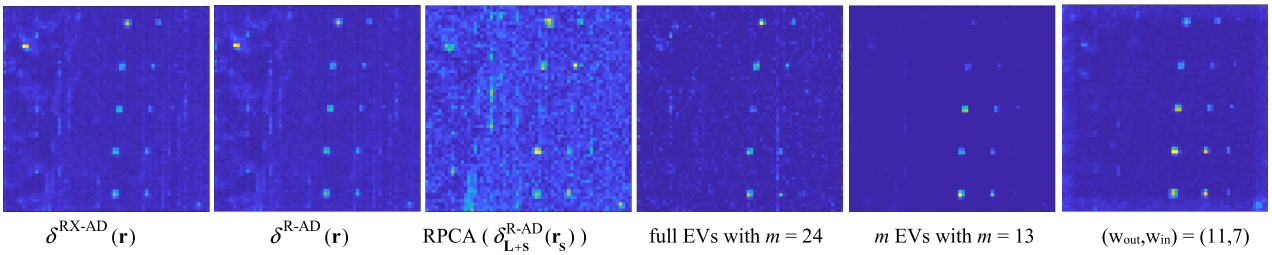


Fig. 30. Detection maps of RX/R-AD, RPCA, LSMAD, and CRD-DW-STO for the HYDICE 15-panel scene.

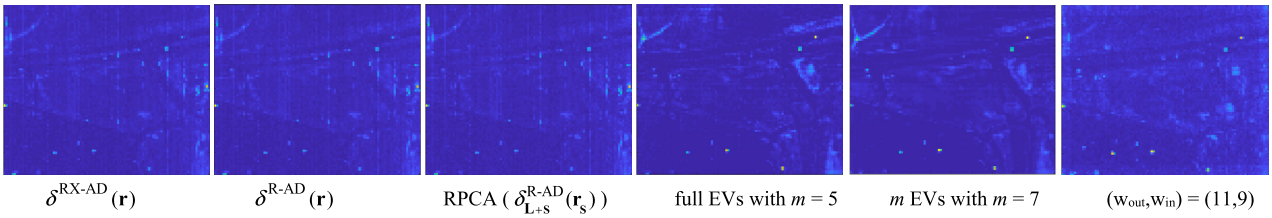


Fig. 31. Detection maps of RX/R-AD, RPCA, LSMAD, and CRD-DW-STO for the HYDICE urban scene.

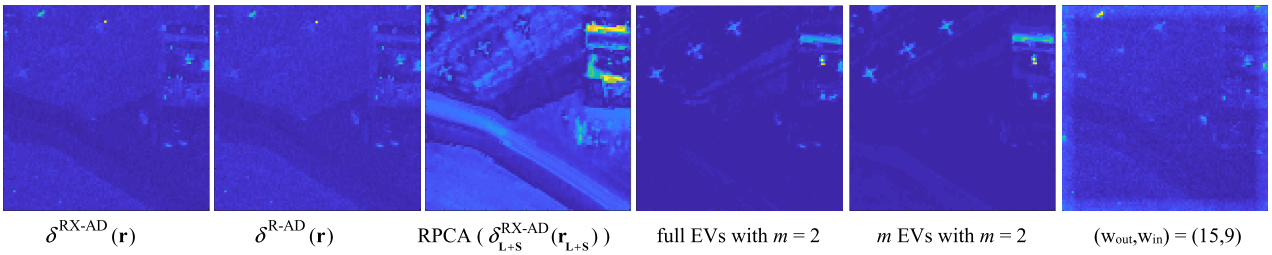


Fig. 32. Detection maps of RX/R-AD, RPCA, LSMAD, and CRD-DW-STO for the San Diego Airport scene.

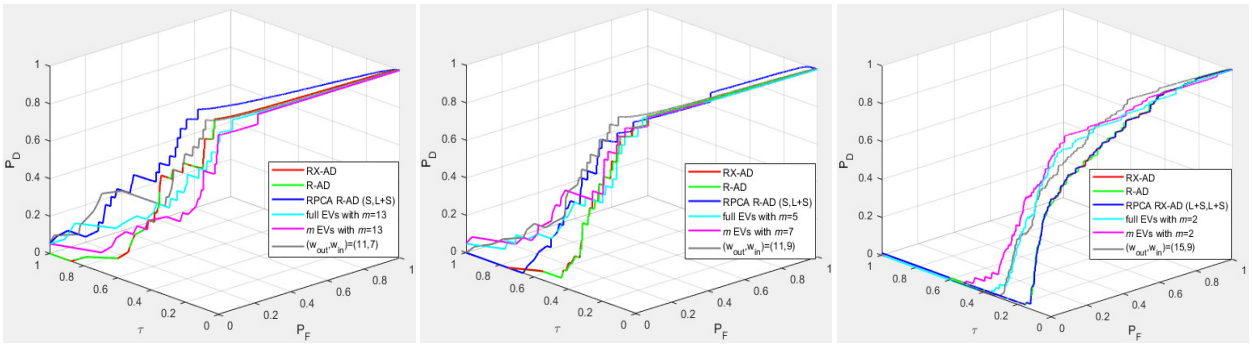


Fig. 33. 3-D ROC curves of Figs. 30–32.

3-D ROC curves shown in Fig. 33. As for HYDICE 15-panel detection in Fig. 30, the best detector was CRD-DW-STO using (w_{out}, w_{in}) = (11, 7) by visual inspection. However, comparing the results in Fig. 30 with those in Figs. 18–20, CRD-DW-STO could do better for BKG suppression than $\boldsymbol{\delta}_{\mathbf{B}}(\mathbf{r}_{\mathbf{A}})$ using $\mathbf{A}$ specified by $\mathbf{L}_m + \mathbf{S}_j$ and $\mathbf{A}$ by $\mathbf{S}_j$, respectively, but did worse in extracting the subpixel panels in the third column. As for the HYDICE urban scene, CRD-DW-STO performed nearly the same as LSMAD did in Fig. 31 by visual inspection. Comparing their results with those in Figs. 22–24, $\boldsymbol{\delta}_{\mathbf{B}}(\mathbf{r}_{\mathbf{A}})$ using $\mathbf{B}$ specified by $\mathbf{L}_m + \mathbf{S}_j$ and $\mathbf{A}$ by $\mathbf{L}_m + \mathbf{S}_j$ or $\mathbf{S}_j$ seemed to perform slightly better than CRD-DW-STO and LSMAD by visual inspection in finding anomalies. As for the San Diego Airport scene, the best detector seemed to be LSMAD according to Fig. 32 by visual inspection. However, if we compare Fig. 32 with Figs. 26–28, $\boldsymbol{\delta}^{\text{R-AD}}_{\mathbf{L}_2}(\mathbf{r}_{\mathbf{L}_2+\mathbf{S}_9})$ worked better than LSMAD by visual inspection in terms of three-airplane detection and BKG suppression, while the RPCA seemed to the best detector to extract BKG.

Obviously, the qualitative analysis using Figs. 30–32 by visual inspection may not be objective. To address this problem, a quantitative analysis using 3-D ROC analysis was also conducted, where the 3-D ROC curves of Figs. 30–32 were plotted in Fig. 33 along with Tables XII–XIV that tabulate the AUC values of their corresponding 2-D ROC curves [2-D ROC curves of (P_D, P_F), 2-D ROC curves of (P_D, τ), and 2-D ROC curves of (P_F, τ)] and their AUC_{OD} calculated by (43), where the best results are boldfaced. As shown in Tables XII–XIV, the best results are consistent with the qualitative analysis made by visual inspection from Figs. 30–32.

TABLE XI
AUC OF SAN DIEGO AIRPORT SCENE

	p =10, m = 2 and j = 8				p =11, m = 2 and j = 9				p =29, m = 16 and j =13			
Detector	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}	(P_D,P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
$\delta_{\mathbf{S}_j}^{\text{RX-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9374	0.3590	0.0347	1.2617	0.9427	0.3011	0.0312	1.2125	0.8999	0.1138	0.0232	0.9904
$\delta_{\mathbf{S}_j}^{\text{R-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9379	0.3591	0.0347	1.2623	0.9431	0.3012	0.0313	**1.2130**	0.8998	0.1138	0.0232	0.9904
$\delta_{\mathbf{L}_m}^{\text{RX-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9441	0.1736	0.0173	1.1004	0.9447	0.1863	0.0152	1.1158	0.7308	0.0480	0.0245	0.7543
$\delta_{\mathbf{L}_m}^{\text{R-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9516	0.1338	0.0058	1.0796	0.9500	0.1450	0.0057	1.0892	0.9054	0.0385	0.0115	0.9324
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{RX-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9375	0.3574	0.0348	1.2602	0.9423	0.2975	0.0311	1.2088	0.8977	0.1109	0.0227	0.9859
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{R-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9380	0.3596	0.0348	**1.2628**	0.9433	0.2995	0.0312	1.2116	0.9019	0.1164	0.0239	0.9944
$\delta_{\mathbf{S}_j}^{\text{RX-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.3145	0.1830	0.3054	0.1921	0.3175	0.1894	0.3064	0.2005	0.2646	0.1669	0.3288	0.1026
$\delta_{\mathbf{S}_j}^{\text{R-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.3144	0.1830	0.3054	0.1920	0.3174	0.1894	0.3065	0.2003	0.2646	0.1669	0.3288	0.1026
$\delta_{\mathbf{L}_m}^{\text{RX-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9453	0.1569	0.0151	1.0870	0.9444	0.1969	0.0174	1.1238	0.9537	0.0920	0.0177	**1.0279**
$\delta_{\mathbf{L}_m}^{\text{R-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9082	0.1429	0.0169	1.0341	0.9075	0.1641	0.0175	1.0541	0.9514	0.0937	0.0180	1.0271
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{RX-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9379	0.3593	0.0351	1.2621	0.9432	0.2992	0.0314	1.2109	0.9134	0.1180	0.0252	1.0062
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{R-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9362	0.3593	0.0351	1.2603	0.9427	0.2992	0.0314	1.2105	0.9120	0.1175	0.0251	1.0044

TABLE XII
AUC OF RX/R-AD, RPCA, LSMAD, AND CRD-DW-STO FOR THE HYDICE 15-PANEL SCENE

			(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
RX-AD			0.9898	0.3574	0.0430	1.3042
R-AD			0.9900	0.3584	0.0435	1.3049
RPCA	$\delta_{\mathbf{L}+\mathbf{S}}^{\text{R-AD}}(\mathbf{r}_{\mathbf{S}})$		0.9826	0.5209	0.1205	1.3829
LSMAD	p	m	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
full EVs	24	13	0.9871	0.3684	0.0140	1.3415
m EVs	24	13	0.9837	0.2720	0.0051	1.2506
CRD-DW-STO	w_{out}	w_{in}	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
	11	7	0.9953	0.4673	0.0274	**1.4352**

TABLE XIII
AUC OF RX/R-AD, RPCA, LSMAD, AND CRD-DW-STO FOR THE HYDICE URBAN SCENE

			(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
RX-AD			0.9872	0.2575	0.0361	1.2087
R-AD			0.9872	0.2549	0.0359	1.2062
RPCA	$\delta_{\mathbf{L}+\mathbf{S}}^{\text{R-AD}}(\mathbf{r}_{\mathbf{S}})$		0.9678	0.3617	0.0939	1.2356
LSMAD	p	m	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
full EVs	9	5	0.9846	0.3315	0.0246	1.2915
m EVs	13	7	0.9904	0.3808	0.0233	1.3480
CRD-DW-STO	w_{out}	w_{in}	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
	11	9	0.9966	0.4469	0.0506	**1.3928**

C. Discussion

From the experiments conducted in Sections X-A and X-B, it clearly demonstrated that anomaly detection using $\delta_{\mathbf{B}}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{A}})$ has various advantages over RX/R-AD, RPCA, LSMAD, and CRD by specifying a different combination of $\mathbf{A}$ and $\mathbf{B}$ that can be obtained by the low-rank matrix $\mathbf{L}_m$ and the sparse matrix $\mathbf{S}_j$ produced by OSP-GoDec. For example, if the detection of the anomalies is of major interest, matrix $\mathbf{A}$ should be specified by $\mathbf{A} = \mathbf{S}_j$ or $\mathbf{L}_m + \mathbf{S}_j$, but $\mathbf{B}$ must be specified by $\mathbf{L}_m + \mathbf{S}_j$. Otherwise, BKG can be detected by letting $\mathbf{A} = \mathbf{L}_m$ or $\mathbf{L}_m + \mathbf{S}_j$ and $\mathbf{B} = \mathbf{S}_j$ In addition, as also shown by the experiments, an appropriate combination of $\mathbf{A}$ and $\mathbf{B}$ always outperformed the best cases of RX/R-AD, RPCA, LSMAD, and CRD-DW-STO. Unfortunately, finding appropriate values of m, the rank of $\mathbf{L}_m$ and the rank of $\mathbf{S}_j$, j, pose a great challenging issue. To the best of our knowledge, no work of how to determine these two parameters has been reported in the literature. This article takes advantage of the HFC/NWHFC/MOCA method and the MX-SVD developed in MOCA to estimate these two parameters. As demonstrated by the experiments, the HFC/NWHFC-estimated p used to determine m and the MX-SVD-estimated j used to determine the cardinality of $\mathbf{S}_j$, $k = j \times N$ for the OSP-GoDec were effective and more appropriate to be used for finding $\mathbf{L}_m$ and $\mathbf{S}_j$. With appropriately determining m and j, $\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$ performed significantly better than RX/R-AD, LSMAD, and CRD-DW-STO in nearly all three counts, i.e., the AUC values of the 2-D ROC curves of (P_D, P_F), 2-D ROC curves of (P_D, τ), and 2-D ROC curves of (P_F, τ).

TABLE XIV
AUC OF RX/R-AD, RPCA, LSMAD, AND CRD-DW-STO FOR THE SAN DIEGO AIRPORT SCENE

			(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
RX-AD			0.8314	0.0802	0.0453	0.8663
R-AD			0.8270	0.0832	0.0501	0.8602
RPCA	$\delta^{\text{RX-AD}}_{\mathbf{L}+\mathbf{S}}(\mathbf{r}_{\mathbf{L}+\mathbf{S}})$		0.8314	0.0802	0.0453	0.8663
LSMAD	p	m	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
full EVs	11	2	0.9245	0.1416	0.0151	1.0510
m EVs	10	2	0.9405	0.1847	0.0183	**1.1069**
CRD-DW -STO	w_{out}	w_{in}	(P_D, P_F)	(P_D,τ)	(P_F,τ)	AUC_{OD}
	15	9	0.8942	0.1546	0.0589	0.9900

Since we are interested in anomaly detection, component matrix $\mathbf{A}$ used in the data sample vector, $\mathbf{r}_\mathbf{A}$, should account for the anomalies themselves by $\mathbf{S}_j$ alone or embedded in BKG by $\mathbf{L}_m+\mathbf{S}_j$. In either case, there are three scenarios, $\mathbf{S}_j$, $\mathbf{L}_m$, and $\mathbf{L}_m+\mathbf{S}_j$, which can specify $\mathbf{B}$ used in $\delta^{\text{RX/R-AD}}_{\mathbf{B}}(\mathbf{r}_\mathbf{A})$. As a result, for each anomaly detector, $\delta^{\text{RX/R-AD}}_{\mathbf{B}}(\mathbf{r}_\mathbf{A})$, six types of ADs are proposed in this article depending on how to specify matrices $\mathbf{A}$ and $\mathbf{B}$ used in $\delta^{\text{RX/R-AD}}_{\mathbf{B}}(\mathbf{r}_\mathbf{A})$ with different combinations of $\mathbf{L}_m$ and $\mathbf{S}_j$. In other words, what $\delta^{\text{RX/R-AD}}_{\mathbf{B}}(\mathbf{r}_\mathbf{A})$ does is exactly the same as $\delta^{\text{RX/R-AD}}(\mathbf{r})$, which uses the adaptive/local/sliding windows to detect the anomalies in the window analogous to matrix $\mathbf{B}$, which includes the data sample vectors to form the covariance/correlation matrix $\mathbf{K/R}$. As shown in the experiments conducted above, the role of matrix $\mathbf{B}$ played a key in anomaly detection.

According to TI, TE, and real-image data sets, the targets can be categorized into two classes: visible targets (such as panels in the first three columns in TI and TE and the panels in the first two columns in the HYDICE 15-panel scene, roofs, grass, and airplanes in the San Diego Airport scene) and invisible targets (such as ten panel subpixels in the fourth and fifth columns in TI and TE, five mixed panel pixels in the third column in the HYDICE 15-panel scene data, and 21 target pixels in the HYDICE urban scene). Of particular interest are the 2×2 mixed panels in the third column in TI and TE, which can be considered as visible targets panelwise but also as invisible targets pure signaturewise, because a pure signature of a mixed panel only occupies a single pixel. Using different matrices to specify the matrix $\mathbf{B}$ yields different types of AD, which can detect these two different types of targets, visible and invisible targets, separately or altogether. However, since the anomalies cannot be known *a priori* and cannot be visualized by inspection, technically speaking, only invisible targets can be considered as anomalies. By considering this, traditional anomaly detectors $\delta^{\text{RX/R-AD}}(\mathbf{r})$ and the most recent state-of-the-art detection techniques, RPCA, LSMAD, and CRD-DW-STO, can only detect visible targets and not anomalies. Interestingly, when $\mathbf{A}$ and $\mathbf{B}$ are specified by the same component, either $\mathbf{S}_j$ or $\mathbf{L}_m+\mathbf{S}_j$, which contains anomalies as the sparse signals of interest, then $\delta^{\text{RX/R-AD}}_{\mathbf{B}}(\mathbf{r}_\mathbf{A})$ becomes $\delta^{\text{RX/R-AD}}_{\mathbf{A}}(\mathbf{r}_\mathbf{A})$. In this case, invisible targets that are indeed the anomalies embedded in BKG could be detected by suppressing the data sample vectors using matrix $\mathbf{B}$ surrounding these anomalies, as shown in Figs. 3–5 for TI, Figs. 9–11 for TE, and Figs. 18–20 for the HYDICE 15-panel scene. On the other hand, if an AD uses $(\mathbf{A}, \mathbf{B}) = (\mathbf{S}_j/\mathbf{L}_m + \mathbf{S}_j, \mathbf{L}_m)$ or $(\mathbf{A}, \mathbf{B}) = (\mathbf{L}_m + \mathbf{S}_j, \mathbf{S}_j)$, the resulting AD could only detect visible targets and not invisible targets as anomalies shown in Figs. 3–5 for TI, Figs. 9–11 for TE, and Figs. 18–20 for the HYDICE 15-panel scene.

It is generally known that $\delta^{\text{RX-AD}}(\mathbf{r})$ and $\delta^{\text{R-AD}}(\mathbf{r})$ perform very similarly. However, there is an intriguing scenario that showed a significant difference between $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ and $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, where $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ and $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ performed significantly differently in Figs. 3–5 for TI and Figs. 9–11 for TE. Specifically, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ detected visible targets as anomalies as opposed to $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$, which detected invisible targets as anomalies. This example explained that removing the first-order statistics, i.e., sample mean, had a crucial impact on detection when both detectors used $\mathbf{L}_m + \mathbf{S}_j$ to suppress the BKG. As for the HYDICE, $\delta^{\text{RX-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ and $\delta^{\text{R-AD}}_{\mathbf{L}_m+\mathbf{S}_j}(\mathbf{r}_{\mathbf{S}_j})$ performed slightly different with a small difference in BKG suppression.

As a conclusion, the experiments presented in this section showed that using $\mathbf{L}_m$ to suppress BKG can detect visible targets as anomalies embedded in the BKG, while using $\mathbf{S}_j$ to suppress the surrounding of the anomalies can detect invisible targets as anomalies embedded in $\mathbf{S}_j$. Accordingly, it is suggested that when the LRaSMD model is used for anomaly detection, the two matrices $\mathbf{S}_j$ and $\mathbf{L}_m$ can be jointly implemented to detect the visible and invisible anomalies. This unique advantage offers rare benefits that the commonly used anomaly detectors in the literature cannot provide.

XI. NOVELTIES AND CONTRIBUTIONS OF OSP-GODEC

A number of novelties presented in this article can be summarized as follows.

1) OSP is used to reinterpret GoDec as OSP-GoDec. Since the VD in [25] and [26], MOCA and MX-SVD developed in [35] are all eigenanalysis-based techniques that also use the same concept of OSP so that they can be effectively used to estimate the parameters, the rank of the low-rank matrix $\mathbf{L}$, m, and the rank of the sparse matrix $\mathbf{S}_j$, j. The developed OSP-GoDec fully explores the use of OSP in the key concepts of GoDec. It is believed that OSP-GoDec coupled with VD/MOCA and MX-SVD is the first work ever reported in the literature to bridge OSP and GoDec.
2) Instead of directly solving the rank of the low-rank matrix, m and the rank of sparse matrix, j, VD and MOCA are used to estimate $p = m + j$ and then use MX-SVD to determine j. This is a quite different approach compared with a direct estimation of m and k.
3) The sparsity cardinality k can be found by setting $k = j \times N$. This is because each data sample has a sparsity of j, and there are N data samples. It should be noted that the cardinality of a sparsity matrix is not the rank of the sparsity matrix $\mathbf{S}$. More specifically, if the sparse matrix $\mathbf{S} = \mathbf{S}_{N\times L}$ is represented by a matrix $\mathbf{S}_{N\times L} = [\mathbf{s}_1^T \mathbf{s}_2^T \cdots \mathbf{s}_N^T]$, where N is the total number of data samples and L is the total number of spectra bands, then $\mathbf{s}_i$ is an L-dimensional signal with a sparsity order specified by j in terms of compressive sensing [75],

[76]. This gives rise to the cardinality of sparsity on all entries in the sparse matrix $\mathbf{S}$, which is $k = j \times N$.

4) The idea of using the low-rank matrix $\mathbf{L}_m$ and the sparse matrix $\mathbf{S}_j$ to implement an anomaly detector for BKG suppression and target detection is completely new and novel. It can be also applied to any variant of RX/R-AD as well as any $\mathbf{L}_m + \mathbf{S}_j$ model such as RPCA [1] and $\mathbf{L}_m + \mathbf{S}_j + \mathbf{N}$ model such as LSMAD [20].
5) The 3-D ROC analysis in [73] and [74] can be further used as an evaluation tool for the overall detection performance analysis. Specifically, a new designed metric, called AUC_{OD}, is particularly designed in (43), which can be used as a quantitative detection measure in conjunction with the qualitative real-valued detection maps by visual inspection to provide better assessment of the overall detection performance. The idea of AUC_{OD} is also believed to be the first ever proposed for this purpose and has never been explored in the literature.

XII. Conclusion

LRaSMD has received much attention lately in various applications such as data unmixing, anomaly detection, and classification. The model discussed in this article is a three-component decomposition $\mathbf{L}_m + \mathbf{S}_j + \mathbf{N}$, where the BKG subspace is assumed to be compacted into a low-rank dimensional space specified by $\mathbf{L}_m$ and the target subspace is considered as a sparse space specified by $\mathbf{S}_j$ with the noise subspace used to account for the model error. In order to find the low-rank matrix and the sparse matrix, GoDec has been widely used for this purpose, but the two parameters $m =$ rank of $\mathbf{L}_m$ and $j =$ the rank of $\mathbf{S}_j$ must be provided *a priori* or empirically in advance. This article investigates the $\mathbf{L}_m + \mathbf{S}_j + \mathbf{N}$ model and rederives a new OSP version of GoDec called OSP-GoDec to find the low-rank matrix $\mathbf{L}_m$ and the sparsity matrix $\mathbf{S}_j$, where $\mathbf{L}_m + \mathbf{S}_j$ can be considered as a target signal component embedded in the noise component $\mathbf{N}$. This is because OSP can be used as a target detector in [30]. By taking advantage of OSP-GoDec derived from OSP, this article further develops two OSP-based techniques to determine the parameters of m and j with VD and MOCA used to estimate p, the number of spectrally distinct signal sources that include the BKG sources, and MX-SVD used to determine j, the number of sparse signals. As a result, m, the number of BKG signal sources, can be obtained by $m = p - j$. The experimental results demonstrate that the new developed OSP-GoDec using m and j estimated by VD/MOCA and MX-SVD can decompose the BKG and sparse-signal matrices effectively to provide better detection performance.

Finally, synthetic and real-image data sets are used for experiments to conduct an extensive and comprehensive study and analysis for OSP-GoDec to perform the commonly used anomaly detector RX-AD [68] and R-AD [25], [69], [70] with various types of using different combinations of decomposed low-rank matrix $\mathbf{L}_m$ for BKG suppression and sparse matrix $\mathbf{S}_j$ to detect the target signals. These experiments provide many interesting and intriguing findings on anomalies such as visible mixed-pixel targets and invisible subpixel targets that may be very useful for future development of anomaly detection. In particular, the new developed OSP-GoDec using VD/MOCA and MX-SVD offers new directions of hyperspectral target detection in the future research.

References

[1] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1027–1063, 2009.
[2] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. Willsky, "Rank-sparsity incoherence for matrix decomposition," *Siam J. Optim.*, vol. 21, no. 2, pp. 572–596, 2009. [Online]. Available: http://arxiv.org/abs/0906.2220
[3] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," 2009, *arXiv:0909.4061*. [Online]. Available: http://arxiv.org/abs/0909.4061
[4] W. Wei, L. Zhang, Y. Zhang, C. Wang, and C. Tian, "Hyperspectral image denoising from an incomplete observation," in *Proc. Int. Conf. Orange Technol. (ICOT)*, Dec. 2015, pp. 177–180.
[5] S. Rambhatla, X. Li, and J. Haupt, "Target-based hyperspectral demixing via generalized robust PCA," in *Proc. 51st Asilomar Conf. Signals, Syst., Comput.*, Oct. 2017, pp. 420–424.
[6] Y. Gao, H. Yan, L. Zhang, R. Xi, Y. Zhang, and W. Wei, "Matrix decomposition based salient object detection in hyperspectral imagery," in *Proc. 13th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Jul. 2017, pp. 574–577.
[7] D. Hong and X. X. Zhu, "SULoRA: Subspace unmixing with low-rank attribute embedding for hyperspectral data analysis," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1351–1363, Dec. 2018.
[8] A. Gogna, A. Shukla, H. K. Agarwal, and A. Majumdar, "Split Bregman algorithms for sparse/joint-sparse and low-rank signal recovery: Application in compressive hyperspectral imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 1302–1306.
[9] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
[10] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5366–5380, Sep. 2017.
[11] W. Sun, L. Tian, Y. Xu, B. Du, and Q. Du, "A randomized subspace learning based anomaly detector for hyperspectral imagery," *Remote Sens.*, vol. 10, no. 3, p. 417, Mar. 2018, doi: 10.3390/rs10030417.
[12] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 713–729, Mar. 2018.
[13] K. Tan, Z. Hou, D. Ma, Y. Chen, and Q. Du, "Anomaly detection in hyperspectral imagery based on low-rank representation incorporating a spatial constraint," *Remote Sens.*, vol. 11, no. 13, p. 1578, Jul. 2019.
[14] F. Kucuk, B. Toreyin, and F. Çelebi, "Sparse and low-rank matrix decomposition-based method for hyperspectral anomaly detection," *J. Appl. Remote Sens.*, vol. 13, no. 1, Feb. 2019, Art. no. 014519.
[15] Z. Zhou, X. Li, J. Wright, E. J. Candes, and Y. Ma, "Stable principal component pursuit," in *Proc. IEEE ISIT*, Jun. 2010, pp. 1518–1522.
[16] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. 28th Int. Conf. Mach. Learn.*, Bellevue, WA, USA, 2011, pp. 33–40.
[17] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.
[18] X. Cui, Y. Tian, L. Weng, and Y. Yang, "Anomaly detection in hyperspectral imagery based on low-rank and sparse decomposition," in *Proc. 5th ICGIP*, vol. 9069, 2013, pp. 1–73.
[19] 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. 083641-1–083641-18, 2014.
[20] 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.
[21] L. Zhang and C. Zhao, "A spectral-spatial method based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection," *Int. J. Remote Sens.*, vol. 38, no. 14, pp. 4047–4068, Jul. 2017.

[22] H. Chang, T. Wang, A. Li, and H. Fang, "Local hyperspectral anomaly detection method based on low-rank and sparse matrix decomposition," *J. Appl. Remote Sens.*, vol. 13, no. 2, p. 026513, Jun. 2019.

[23] E. J. Candes and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[24] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparsity matrix decomposition in noisy case," in *Proc. 28th Int. Conf. Mach. Learn.*, Bellevue, WA, USA, 2011, pp. 33–40.

[25] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral detection and Classification*. Norwell, MA, USA: Kluwer, 2003.

[26] C.-I Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*. Hoboken, NJ, USA: Wiley, 2013.

[27] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[28] D. C. Heinz and C.-I Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[29] C.-I Chang and D. Heinz, "Constrained subpixel detection for remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1144–1159, May 2000.

[30] J. C. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[31] C.-I Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[32] C.-I Chang. (Sep. 28, 2009). *Virtual Dimensionality for Hyperspectral Imagery SPIE Newsroom*. [Online]. Available: http://newsroom.spie.org/x37123.xml, doi: 10.1117/2.1200909.1749.

[33] C.-I Chang, "A review of virtual dimensionality for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1285–1305, Apr. 2018.

[34] C.-I Chang, *Real-Time Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation*. Cham, Switzerland: Springer, 2017.

[35] O. Kuybeda, D. Malah, and M. Barzohar, "Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5579–5592, Dec. 2007.

[36] C.-I Chang, W. Xiong, H.-M. Chen, and J.-W. Chai, "Maximum orthogonal subspace projection approach to estimating the number of spectral signal sources in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 504–520, Jun. 2011.

[37] C.-I Chang, W. Xiong, and C.-H. Wen, "A theory of high-order statistics-based virtual dimensionality for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 188–208, Jan. 2014.

[38] H. Ren and C.-I Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.

[39] 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.

[40] V. Cevher, A. Sankaranarayanana, M. Duarte, D. Reddy, R. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Proc. ECCV*, 2008, pp. 155–168.

[41] E. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method of l1-regularized minimization with applications to compressed sensing," Dept. Comput. Appl. Math., Rice Univ., Houston, TX, USA, CAAM Tech. Rep. 07, 2007.

[42] X. Yuan and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," *Pacific J. Optim.*, vol. 9, no. 1, pp. 1–11, 2009.

[43] A. B. Ramirez, H. Arguello, and G. Arce, "Video anomaly recovery from compressed spectral imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 1321–1324.

[44] H. Guo and N. Vaswani, "Video denoising via online sparse and low-rank matrix decomposition," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Jun. 2016, pp. 1–5.

[45] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, "On the applications of robust PCA in image and video processing," *Proc. IEEE*, vol. 106, no. 8, pp. 1427–1457, Aug. 2018.

[46] W. Kim and C. Jung, "Illumination-invariant background subtraction: Comparative review, models, and prospects," *IEEE Access*, vol. 5, pp. 8369–8384, 2017.

[47] Y. Cheng and T. Chen, "Application of principal component pursuit to process fault detection and diagnosis," in *Proc. Amer. Control Conf.*, Jun. 2013, pp. 3535–3540.

[48] N. Wang and D.-Y. Yeung, "Bayesian robust matrix factorization for image and video processing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1785–1792.

[49] L. Xiong, X. Chen, and J. Schneider, "Direct robust matrix factorizatoin for anomaly detection," in *Proc. IEEE 11th Int. Conf. Data Mining*, Dec. 2011, pp. 844–853.

[50] D. Zonoobi and A. A. Kassim, "Low rank and sparse matrix reconstruction with partial support knowledge for surveillance video processing," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 335–339.

[51] S. Javed, A. Mahmood, S. Al-Maadeed, T. Bouwmans, and S. K. Jung, "Moving object detection in complex scene using spatiotemporal structured-sparse RPCA," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 1007–1022, Feb. 2019.

[52] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2259–2272, Nov. 2011.

[53] M. Cossalter, M. Tagliasacchi, and G. Valenzise, "Privacy-enabled object tracking in video sequences using compressive sensing," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2009, pp. 2–4.

[54] N. Vaswani, Y. Chi, and T. Bouwmans, "Rethinking PCA for modern data sets: Theory, algorithms, and applications [scanning the issue]," *Proc. IEEE*, vol. 106, no. 8, pp. 1274–1276, Aug. 2018.

[55] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, Feb. 2017.

[56] 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.

[57] K. Rong, S. Wang, X. Zhang, and B. Hou, "Low-rank and sparse matrix decomposition-based pan sharpening," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 2276–2279.

[58] K. Rong, L. Jiao, S. Wang, and F. Liu, "Pansharpening based on low-rank and sparse decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4793–4805, Dec. 2014.

[59] T. Zhou and D. Tao, "Greedy bilateral sketch, completion and smoothing," in *Proc. 16th Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2013, pp. 650–658.

[60] T. Zhou and D. Tao, "Shifted subspaces tracking on sparse outlier for motion segmentation, which reduces motion detection, tracking and segmentation to one efficient matrix factorization model," in *Proc. 23rd Int. Joint Conf. Artif. Intell. (IJCAI)*, Apr. 2013, pp. 1946–1952.

[61] C. Chen, J. Cai, W. Lin, and G. Shi, "Incremental low-rank and sparse decomposition for compressing videos captured by fixed cameras," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 338–348, Jan. 2015.

[62] M. Leadbetter, *Extremes and Related Properties of Random Sequences and Processes*. New York, NY, USA: Springer-Verlag, 1987.

[63] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[64] E. R. Malinowski, "Theory of error in factor analysis," *Anal. Chem.*, vol. 49, no. 4, pp. 606–612, Apr. 1977.

[65] E. R. Malinowski, "Determination of the number of factors and the experimental error in a data matrix," *Anal. Chem.*, vol. 49, no. 4, pp. 612–617, Apr. 1977.

[66] H. V. Poor, *An Introduction to Detection and Estimation Theory*, 2nd ed. New York, NY, USA: Springer-Verlag, 1994.

[67] J. C. Harsanyi, W. Farrand, and C.-I Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in *Proc. Annu. Meeting, Amer. Soc. Photogramm. Remote Sens.*, Reno, NV, USA, 1994, pp. 236–247.

[68] 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.

[69] C.-I Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.

[70] C.-I Chang, *Real-Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection*. New York, NY, USA: Springer, 2016.

[71] C. C. Wu, C. S. Lo, and C.-I Chang, "Improved process for use of a simplex growing algorithm for endmember extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 523–527, Jul. 2009.

[72] C.-I Chang, X. Jiao, C.-C. Wu, Y. Du, and M.-L. Chang, "A review of unsupervised spectral target analysis for hyperspectral imagery," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, Apr. 2010, Art. no. 503752.

[73] M. Song, X. Shang, and C.-I Chang, "3D receiver operating characteristic analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published.

[74] C.-I Chang, "Multiparameter receiver operating characteristic analysis for signal detection and classification," *IEEE Sensors J.*, vol. 10, no. 3, pp. 423–442, Mar. 2010.

[75] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[76] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Jun. 2007.

Chein-I Chang (Life Fellow, IEEE) received the B.S. degree in mathematics from Soochow University, Taipei, Taiwan, in 1973, the M.S. degree in mathematics from the Institute of Mathematics, National Tsing Hua University, Hsinchu, Taiwan, in 1975, the M.A. degree in mathematics from the State University of New York, Stony Brook, NY, USA, in 1977, the M.S. and M.S.E.E. degrees from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 1982, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 1987.

He has been with the University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA, since 1987, where he is a Professor with the Department of Computer Science and Electrical Engineering. He has been holding the Chang Jiang Scholar Chair Professorship and the Director of the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Dalian Maritime University, Dalian, China, since 2016. In addition, he has also been a Chair Professor of the National Chiao Tung University, Hsinchu, since 2019. He has seven patents on hyperspectral image (HSI) processing. He has authored four books, the *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Kluwer Academic Publishers, 2003), the *Hyperspectral Data Processing: Algorithm Design and Analysis* (John Wiley & Sons, 2013), the *Real-Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection* (Springer, 2016), and the *Real-Time Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation* (Springer, 2017). In addition, he also edited two books, the *Recent Advances in Hyperspectral Signal and Image Processing*, 2006 and the *Hyperspectral Data Exploitation: Theory and Applications* (John Wiley & Sons, 2007) and co-edited with A. Plaza a book on the *High Performance Computing in Remote Sensing* (CRC Press, 2007). His research interests include multispectral/HSI processing, automatic target recognition, and medical imaging.

Dr. Chang is a fellow of SPIE. He received the National Research Council Senior Research Associateship Award from 2002 to 2003 sponsored by the U.S. Army Soldier and Biological Chemical Command, Edgewood Chemical and Biological Center, Aberdeen Proving Ground, Maryland. He was a Plenary Speaker of the Society for Photo-Optical Instrumentation Engineers (SPIE) Optics + Applications, Remote Sensing Symposium in 2009. He was the Guest Editor of a special issue of the *Journal of High Speed Networks* on Telemedicine and Applications (2000) and a Co-Guest Editor of another special issue of the same journal on the Broadband Multimedia Sensor Networks in Healthcare Applications in 2007. He is also the Co-Guest Editor of special issues on High Performance Computing of Hyperspectral Imaging for the *International Journal of High Performance Computing Applications* in 2007, Signal Processing and System Design in Health Care Applications for the *EURASIP Journal on Advances in Signal Processing* in 2009, Multipsectral, Hyperspectral, and Polarimetric Imaging Technology for the *Journal of Sensors* in 2016, and Hyperspectral Imaging and Applications for the *Remote Sensing* in 2018.

Hongju Cao received the B.S. degree in applied mathematics from Qufu Normal University, Qufu, China, in 2003, and the M.S. degree in operational research from the Dalian University of Technology, Dalian, China, in 2006. He is pursuing the Ph.D. degree in technology of computer applications with Dalian Maritime University, Dalian.

He is an Associate Professor with the Dalian University of Foreign Languages, Dalian. His interests include anomaly detection and classification in hyperspectral image processing.

Shuhan Chen (Student Member, IEEE) received the B.S. degree from Ludong University, Yantai, China, in 2011, and the M.S. degree from Liaoning Technical University, Huludao, China, in 2014. She is pursuing the Ph.D. degree in control theory and control engineering with Zhejiang University, Hangzhou, China.

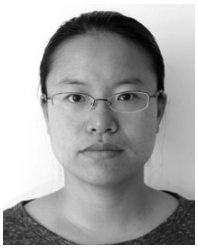
She is an Exchanging Ph.D. Student, visiting Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA. Her research interests include image registration, hyperspectral image processing, and patter recognition.

Xiaodi Shang (Student Member, IEEE) received the B.S. degree in software engineering from Qingdao University, Qingdao, China, in 2016. She is pursuing the Ph.D. degree in computer application technology with Dalian Maritime University, Dalian, China.

Her research interests include hyperspectral image classification, band selection, and applications.

Chunyan Yu (Member, IEEE) received the B.Sc. and Ph.D. degrees in environment engineering from Dalian Maritime University, Dalian, China, in 2004 and 2012, respectively.

In 2004, she joined the College of Computer Science and Technology, Dalian Maritime University. From 2013 to 2016, she was a Post-Doctoral Fellow with Information Science and Technology College, Dalian Maritime University. From 2014 to 2015, she was a Visiting Scholar with the College of Physicians and Surgeons, Columbia University, New York, NY, USA. She is an Associate Professor with Information Science and Technology College, Dalian Maritime University. Her research interests include image segmentation, hyperspectral image classification, and pattern recognition.



Meiping Song received the Ph.D. degree from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China, in 2006.

From 2013 to 2014, she was a Visiting Associate Research Scholar with Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore County, Baltimore, MD, USA. She is an Associate Professor with the College of Information Science and Technology, Dalian Maritime University, Dalian, China. Her research includes remote sensing and hyperspectral image processing.