

Orthogonal Subspace Projection Target Detector for Hyperspectral Anomaly Detection

Chein-I Chang, *Life Fellow, IEEE*, Hongju Cao , and Meiping Song

Abstract—Orthogonal subspace projection (OSP) is a versatile hyperspectral imaging technique which has shown great potential in dimensionality reduction, target detection, spectral unmixing, etc. However, due to its inherent requirement of prior target knowledge, OSP has not been explored in anomaly detection. This article takes advantage of an unsupervised OSP-based algorithm, automatic target generation process (ATGP), and a recently developed OSP-go decomposition (OSP-GoDec) along with data sphering (DS) to make OSP applicable to anomaly detection. Its idea is to implement ATGP on the background (BKG) and target subspaces constructed from the low-rank matrix \mathbf{L} and sparse matrix \mathbf{S} generated by OSP-GoDec to derive an OSP-based anomaly detector (OSP-AD). In particular, OSP-AD also includes DS to remove BKG interference from the target subspace so as to enhance anomaly detection. Surprisingly, operating data samples on different constructions of the BKG subspace and the target subspace yields various versions of OSP-AD. Experiments show that given an appropriate construction of the BKG subspace and the target subspace, OSP-AD can be shown to outperform existing anomaly detectors including Reed-Xiaoli anomaly detector and collaborative representation-based anomaly detector (CRD).

Index Terms—Anomaly detection (AD), automatic target generation process (ATGP), data sphering (DS), go decomposition (GoDec), low rank and sparse matrix decomposition (LRaSMD), orthogonal subspace projection (OSP), OSP-based anomaly detector (OSP-AD).

I. INTRODUCTION

DUE to its use of hundreds of contiguous spectral bands, a hyperspectral imaging sensor is capable of uncovering

Manuscript received February 6, 2021; revised March 3, 2021; accepted March 20, 2021. Date of publication March 25, 2021; date of current version May 26, 2021. 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 61971082, Grant 61890964, and Grant 3132019341. (*Corresponding author: Hongju Cao.*)

Chein-I Chang is with the Center for Hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China, with the Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD 21250 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, 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).

Meiping Song is with the Center for Hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: 49322784@qq.com).

Digital Object Identifier 10.1109/JSTARS.2021.3068983

many subtle material substances which cannot be visualized by inspection or known by prior knowledge such as anomalous targets including subpixel targets, endmembers, mixed pixel targets, etc. Accordingly, hyperspectral anomaly detection (AD) has received considerable interest in hyperspectral data exploitation [1]–[3].

Theoretically speaking, an anomaly is a target that is generally unknown and also unexpected. In most cases, its size is relatively very small so that its appearance is generally invisible. Specifically, its presence cannot be known *a priori* or by visual inspection. To address this issue, anomaly detection must be performed in a completely unknown environment. One commonly used and general approach is the one developed by Reed and Yu in [4], called RXD, which is essentially Mahalanobis distance, $(\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu})$ derived from a Gaussian distribution with the sample mean $\boldsymbol{\mu}$ and sample auto-covariance matrix \mathbf{K} . It was then modified and extended to various versions in [5], particularly the one which replaced the \mathbf{K} used in RXD with the sample auto-correlation matrix \mathbf{R} to derive a new anomaly detector

$$\mathbf{r}^T \mathbf{R}^{-1} \mathbf{r}. \quad (1)$$

Because (1) uses the sample auto-correlation matrix to perform AD, it is referred to as R-AD as opposed to RXD which is referred to as RX-AD in this article for distinction. Since then, a wide variety of anomaly detectors derived from RX/R-AD have been reported in the literature. For example, window-based local anomaly detectors, which implemented \mathbf{K} or \mathbf{R} using local windows [6], [7], sliding windows [8], dual windows [9], [10], multiple windows [11] and kernel anomaly detector [12], anomaly detection for unlabeled classification [13], real time processing of anomaly detection [14], guided filtering-based AD [15], spectral-spatial feature extraction-based AD [16], background separation-based AD [17], and sparsity score-estimation framework for AD [18]. Most recently, other approaches have been also developed such as deep learning-based anomaly detector, low rank and sparse matrix decomposition (LRaSMD) model-based anomaly detectors [19]–[24], low rank and sparse representation [25]–[27], autoencoder [28]–[30], generative adversarial network (GAN) [31], [32], game theory-based AD [33]. All of these works did not go beyond the original idea of RX/R-AD.

This article deviates the commonly used RX/R-AD-based approach mentioned above by particularly looking into AD from an orthogonal subspace projection (OSP) perspective. It is referred to as OSP target detector (OSP-TD) [1], [33], [34] to

distinguish from OSP used as a projection process in the literature. Over the past years, OSP has been widely used for various applications, for example, data dimensionality reduction and classification [34], target detection [35]–[37], spectral unmixing [38], endmember finding [39]–[41], etc. However, as OSP-TD was first developed as target detector in [34], it implemented a linear matched filter form given by

$$\mathbf{d}^T P_{\mathbf{U}}^{\perp} \mathbf{r} \quad (2)$$

where \mathbf{r} , \mathbf{d} , and \mathbf{U} are specified by the current data sample to be processed, the desired target signature, and a set of the undesired target signatures, respectively. Specifically, \mathbf{d} and \mathbf{U} must be known in advance. In order for OSP to work as an anomaly detector, the key issue is how to remove the requirement of prior knowledge of \mathbf{d} and \mathbf{U} in (2) from OSP-TD.

To resolve this issue, two approaches were investigated in the past. One is an unsupervised OSP-based approach, called automatic target generation process (ATGP) developed in [42], which implements a form of

$$\mathbf{t}^T P_{\mathbf{U}}^{\perp} \mathbf{t} \quad (3)$$

similar to OSP-TD in (2) with \mathbf{t} considered as an unsupervised target to be generated at the current ongoing process and \mathbf{U} comprising of all previously generated unsupervised targets.

A second approach, called constrained energy minimization (CEM) [43] operates a form of

$$\mathbf{d}^T \mathbf{R}^{-1} \mathbf{r} / \mathbf{d}^T \mathbf{R}^{-1} \mathbf{d} \quad (4)$$

which replaces $P_{\mathbf{U}}^{\perp}$ used in (2) with \mathbf{R}^{-1} in (1) to suppress background (BKG) instead of annihilating the undesired target signatures in \mathbf{U} by $P_{\mathbf{U}}^{\perp}$. Unfortunately, neither ATGP nor CEM can be directly applicable to AD because ATGP in (3) requires the knowledge of \mathbf{U} and CEM requires the target knowledge of \mathbf{d} .

There are several obstacles to prevent OSP-TD from being applied to AD. Most importantly, if a data sample \mathbf{r} is considered as an anomalous target sample, its spectral properties should be significantly distinct from that of its surrounding data samples. This is a crucial element in AD. OSP-TD in (2) is a pixel-based technique which does not take the spectral correlation among pixels into account. Although ATGP can generate its own \mathbf{d} and \mathbf{U} in an unsupervised pixel-by-pixel manner without prior target knowledge, unfortunately, ATGP in (3) can only find one target at a time through a series of OSPs and cannot address the issue of interpixel spectral correlation. In addition, when the image background (BKG) is complex, finding appropriate and reliable \mathbf{U} for ATGP to characterize BKG is nearly impossible. On the other hand, CEM in (4) can address the issue of interpixel spectral correlation via \mathbf{R}^{-1} , but it requires the prior target knowledge of \mathbf{d} . Interestingly, can we combine ATGP and CEM to make OSP-TD work as an anomaly detector?

Since the only differences between (1) and (3) are $P_{\mathbf{U}}^{\perp}$ and \mathbf{t} used by ATGP in (3) and \mathbf{R}^{-1} and \mathbf{r} used by R-AD in (1), we can replace the unsupervised target \mathbf{t} in ATGP with the data sample \mathbf{r} to yield a form of

$$\mathbf{r}^T P_{\mathbf{U}}^{\perp} \mathbf{r} \quad (5)$$

in which case, the requirement of knowing the target sample \mathbf{t} is removed by using the current data sample \mathbf{r} . However, we still need to address the issue of \mathbf{U} . Fortunately, a recently developed LRaSMD model in [21] and [44]–[48] can be well served for this purpose. It decomposes a data space \mathbf{X} into a three-component model given by

$$\mathbf{L} + \mathbf{S} + \mathbf{n} \quad (6)$$

with low-rank matrix, sparse matrix, and noise matrix specified by \mathbf{L} , \mathbf{S} , and \mathbf{n} , respectively. Such LRaSMD seems to suggest a possible solution to making OSP-TD an anomaly detector. Specifically, the widely used go decomposition (GoDec) in [48] can be used to produce an LRaSMD model. Most recently, GoDec was further extended to an OSP version of GoDec, referred to as OSP-go decomposition (OSP-GoDec) in [49] where the commonly used RX-AD and R-AD were extended to be implemented by operating the data sample vector \mathbf{r} and the \mathbf{K}/\mathbf{R} not in the original data space \mathbf{X} but rather in their own respective spaces, low-rank matrix \mathbf{L} , and sparse matrix \mathbf{S} .

Inspired by OSP-GoDec, the issue of \mathbf{U} in (5) can be resolved by taking advantage of the low-rank matrix \mathbf{L} and sparse matrix \mathbf{S} in (6) to replace \mathbf{U} and \mathbf{t} , respectively. The key idea is to construct the BKG subspace by either \mathbf{L} alone or $\mathbf{L}+\mathbf{S}$, denoted by $\mathbf{L}/(\mathbf{L}+\mathbf{S})$ to replace \mathbf{U} in (5), while operating \mathbf{r} in (5) in the target subspaces constructed by either \mathbf{S} alone or $\mathbf{L}+\mathbf{S}$, denoted by $\mathbf{S}/(\mathbf{L}+\mathbf{S})$. This gives rise to a new OSP-based AD, referred to as OSP-AD, which implements

$$\mathbf{r}_{\mathbf{A}}^T P_{\mathbf{B}}^{\perp} \mathbf{r}_{\mathbf{A}} \quad (7)$$

where both \mathbf{A} and \mathbf{B} can be specified by $\mathbf{S}/(\mathbf{L}+\mathbf{S})$ and $\mathbf{L}/(\mathbf{L}+\mathbf{S})$, respectively. Due to the fact that anomalies are relatively small targets and do not contribute much statistics to second order of statistics, they can be better characterized by the orders of data statistics higher than 2. To further capture such high-order statistics of anomalies, an additional data sphering (DS) process is also introduced into OSP-AD to remove first- and second-order statistics of the target subspace constructed from \mathbf{A} to produce its sphered spaces, denoted by $\hat{\mathbf{A}}$. As a result, $\mathbf{r}_{\mathbf{A}}^T P_{\mathbf{B}}^{\perp} \mathbf{r}_{\mathbf{A}}$ in (7) becomes OSPDS-AD operating on the sphered target subspace, $\hat{\mathbf{A}}$ and \mathbf{B} as $\mathbf{r}_{\hat{\mathbf{A}}}^T P_{\mathbf{B}}^{\perp} \mathbf{r}_{\hat{\mathbf{A}}}$. This DS process is included to mimic what $\mathbf{K}^{-1}/\mathbf{R}^{-1}$ accomplishes in the RX/R-AD by suppressing the second-order statistics in BKG. This DS is absent in [49] because $\mathbf{K}^{-1}/\mathbf{R}^{-1}$ used by RX/R-AD basically performs data whitening similar to DS.

One key measure to assess the effectiveness of AD is the BKG suppressibility of an anomaly detector. It has been shown in [50] that the traditional 2-D receiver operating characteristic (ROC) curve was not capable of doing so. In this case, detection measures developed in [50] are used for this purpose. Finally, two real hyperspectral image scenes are used for experiments. The experimental results demonstrate that with appropriate selection of \mathbf{A} and \mathbf{B} OSP-AD outperforms classic RX/R-AD [5] and the collaborative representation-based hyperspectral anomaly detection (CRD) [10].

The remainder of this article is organized as follows. Section II briefly reviews related works. Section III describes novelties of this article. Section IV develops OSP-AD. Section V briefly

reviews OSP-GoDec. Section VI derives OSP-AD which constructs the BKG and target subspaces from the OSP-Godec generated low rank and sparse matrices. Section VII introduces 3-D ROC analysis-derived detection measures for detection performance evaluation. Section VIII describes images to be used for experiments. Section IX presents experimental results and discussions. Section X draws conclusion and summarizes contributions.

II. RELATED WORKS

Using the low-rank and sparse matrices to perform AD is not new and has been studied quite sometime in the past [20]–[27]. Its main idea is to take advantage of robust principal component analysis (RPCA) [46] or GoDec [48] or low rank and spare representation [25]–[27] to estimate BKG for suppression and a sparse component to characterize anomalies. Since the residual resulting from the BKG suppression may be likely to be leaked into the sparse component, introducing weighting probabilities into anomalies has been proposed to avoid anomalies being overwhelmed by BKG prior to their detection. Finally, a detection measure such as Euclidean distance, Mahalanobis distance, or RXD is implemented to extract anomalies. Although many reported RPCA-based and low rank and sparse representation-based works have been reported, this article is not based on these approaches. Instead, LRSMD is the major model used in our work. In this case, some of most recent developed LRSMD-based works are briefly reviewed in the following.

In an early attempt, [20] proposed a randomized subspace learning-based anomaly detector (RSLAD) derived from RPCA for hyperspectral anomaly detection. It assumed that BKG could be characterized by the low-rank matrix \mathbf{L} and the anomaly matrix is sparse and has a small portion of nonzero columns (i.e., columnwise). It then used random sampling and the random Hadamard projections to separate anomaly columns from the coarse randomized subspace by solving a series of least squares problems. Once the BKG subspace is found as an undesired subspace \mathbf{U} , the anomalies can be correctly located by $P_{\mathbf{U}}^{\perp}$ onto the orthogonal complement subspace to the purified randomized BKG subspace.

By realizing that anomalies may be corrupted with noise in the sparse component produced by RPCA, [21] proposed an LRSMD-based Mahalanobis distance method for hyperspectral anomaly detection (LSMAD) where a data space can be decomposed into three components—low rank, sparse, and noise matrices modeled by LRSMD. In this case, BKG can be characterized by the low-rank matrix, while the sparse matrix contains the anomaly information so that BKG and anomalies can be separated from noise. Then the statistical features of BKG obtained from the BKG matrix can be further used to design a Mahalanobis distance-based anomaly detector.

Also using the LRSMD model, [22] proposed an LRSMD-based anomaly detector which assumed that BKG could be also specified by the low-rank matrix, whereas anomalies were assumed to be sparsely distributed throughout the image scene in such a way that anomalies could be directly separated from

BKG in one step. Then anomalies are detected by Euclidean distance by a manually selected threshold. The currently being used anomaly detector in the literature generally models BKG by statistical or geometric methods and then distinguishes anomalies from the BKG in a two-step process. By contrast, the anomaly detector designed in [22] separates the anomalies from BKG in one step process by solving a constrained convex optimization problem.

Although BKG and anomalies can be separated by the LRSMD model, such decomposition may not be effective on many occasions. To address this issue, [23] developed an anomaly detection technique by combining LRSMD and cluster weighting strategy to assign weights to potential anomalies. It first implemented LRSMD to produce the low-rank and sparse matrices and calculated the l_2 -norm of each pixel in the component specified by the sparse matrix to yield the initial anomaly matrix. It then used a k-means clustering to weight anomalies to produce final anomaly detection maps.

Following a similar idea to the clustering weight proposed in [23], [24] also took advantage of the LRSMD model to develop an OSP-based BKG suppression technique and further use RXD to estimate adaptive weights for anomalies. After OSP, the interference of BKG in the sparse component can be suppressed effectively. In addition, such RX-based adaptive weighting further improved the AD performance.

Most recently, [49] rederived GoDec as an OSP-GoDec which made use of OSP to reimplement GoDec. In particular, it was the first work ever reported to resolve the issues of determining the parameter r , rank of low-rank matrix and k , cardinality of sparsity, both of which have been determined empirically and manually in the past. Most importantly, [49] used the LRSMD generated by GoDec and OSP-Godec to further derive a number of RX/R-D variants which implemented the low-rank matrix and sparse matrix in various forms and showed to outperform many existing RX/R-D-type anomaly detectors including the collaborative representation-based anomaly detector (CRD) proposed in [10].

Upon using LRSMD, two matrices are produced, the low-rank matrix \mathbf{L} and the sparse matrix \mathbf{S} , each of which generates its own component or subspace. There are three issues which still remain and need to be addressed.

- 1) A crucial issue is how to determine the parameters r and k in LRSMD in [48] and p , or the number of randomly sampled pixels p and K , the number of random Hadamard projected spectral dimensions RSLAD in [20], all of which must be selected by trial and error empirically.
- 2) Another issue is how to effectively utilize the LRSMD-generated \mathbf{L} and \mathbf{S} to perform anomaly detection. A general approach is to construct the BKG subspace from \mathbf{L} for BKG suppression or removal. However, it is often the case that the residuals resulting from such BKG suppression or removal may be very likely also to be leaked into the sparse component to corrupt anomalies. To address this dilemma, several approaches introduced probabilities to weight potential anomalies in the sparse component to prevent them from being contaminated or influenced by BKG [23], [24].

- 3) Finally, a third issue is how to design detection technique to extract anomalies after LRaSMD is used, for example, Euclidean distance [22], Mahalanobis distance [21], RXD in [24].

These three issues have direct impacts on the performance of LRaSMD. The major contribution of this article is to develop OSP-AD to resolve these issues.

III. NOVELTIES OF OSP-AD

As noted in Section II, the effectiveness of LRaSMD hinges its utilization on the two matrices \mathbf{L} and \mathbf{S} . The innovative ideas derived from OSP-AD provide solutions to this issue.

- 1) The most significant novelty is the introduction of OSP-AD in (7) which combines OSP-TD in (2), ATGP in (3) and CEM in (4) to perform AD. Since a direct application of OSP-TD to AD is infeasible, many works reported in the literature used OSP as a preprocessing operator to remove unwanted information such as [20] and [24] by projecting data space onto the orthogonal complement subspace of the BKG subspace by suppressing or removing the BKG interference. The proposed OSP-AD implements OSP-TD as an anomaly detector not a BKG annihilator as originally designed in [34], but rather a BKG suppressor. Specifically, OSP-AD has great advantage of accomplishing BKG suppression and anomaly detection in one-shot operation. This is because OSP-AD implements $\mathbf{r}_A^T P_B^\perp \mathbf{r}_A$ in (7) which operates an OSP projector P_B^\perp on \mathbf{r}_A and a matched filter using \mathbf{r}_A as the matched signal to produce $\mathbf{r}_A^T P_B^\perp \mathbf{r}_A$. These two tasks are accomplished by $\mathbf{r}_A^T P_B^\perp \mathbf{r}_A$ in a one-step process. Compared to OSP-AD, [20] and [24] needed a two-stage process to perform AD by implementing OSP to annihilate BKG in the first stage and then followed by a detector such as RXD in the second stage.
- 2) Another significant novelty is to provide a five-way relationship among OSP-TD in (2), ATGP in (3), OSP-AD in (7) and CEM in (4), R-AD in (1) as shown in Fig. 1 with \mathbf{R}^{-1} being used to suppress BKG and P_B^\perp being used to annihilate the space generated by \mathbf{B} . In particular, two routes can be accomplished in Fig. 1 to arrive OSP-AD. One is OSP-TD \rightarrow ATGP \rightarrow OSP-AD via $P_U^\perp \rightarrow P_B^\perp$ and the other is OSP-TD \rightarrow CEM \rightarrow R-AD \rightarrow OSP-AD via $\mathbf{R}^{-1} \rightarrow P_B^\perp$. Specifically, as illustrated in Fig. 1, replacing \mathbf{t} and \mathbf{U} of ATGP in (3) with \mathbf{r}_A and \mathbf{B} , respectively, yields OSP-AD in (7).
- 3) In order to ensure that anomalies are not compromised by BKG, an additional whitening process must be included to calculate probabilities assigned to anomalies such as [23] and [24]. OSP-AD accomplishes the same task by including DS which removes data statistics of the first and second orders. The resulting OSP-AD is then denoted by OSPDS-AD. This is due to the fact that BKG and anomalies can be characterized by second order and high-order statistics, respectively. DS can effectively reduce BKG interference, while retaining anomalies after the LRaSMD decomposition and prior to OSP-AD. Despite that there are works reported in [24] and [49] using OSP

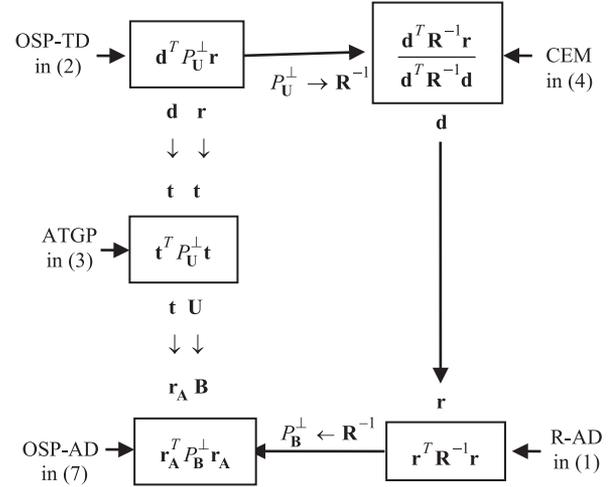


Fig. 1. Diagram of a five-way relationship among OSP, ATGP, OSP-AD, CEM, and R-AD via annihilated by P_U^\perp and suppressed by \mathbf{R}^{-1} along with their matching signatures \mathbf{d} and \mathbf{r} .

to suppress BKG, their used anomaly detectors were still RXD-type anomaly detectors. It seems that using OSP-TD as anomaly detector has not been investigated for AD.

- 4) By means of using the low-rank matrix \mathbf{L} and sparse matrix \mathbf{S} along with their corresponding sphered counterparts $\hat{\mathbf{L}}$ and $\hat{\mathbf{S}}$, we can construct the BKG subspace and target subspace for \mathbf{B} and \mathbf{A} in (7), respectively, for OSPDS-AD to yield various forms of OSPDS-AD, for example, $\mathbf{r}_S^T P_L^\perp \mathbf{r}_S$, $\mathbf{r}_S^T P_{L+S}^\perp \mathbf{r}_S$, etc. This has never been explored in the existing literature.

IV. OSP-BASED ANOMALY DETECTOR

This section further deliberates two major routes depicted in Fig. 1 which have mathematical derivations evolved from the original OSP-TD in (2), ATGP in (3), CEM in (4), anomaly detector in (1), to the development of OSP-AD in (7).

A. Orthogonal Subspace Projection

OSP was originally developed from signal-to-noise ratio (SNR) and is basically a target detector which assumes that both the target to be detected specified by as a desired target signature \mathbf{d} and the undesired targets to be eliminated, specified by $p-1$ target signatures, $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{p-1}$ are to be known *a priori*. An OSP target detector is designed by first undesired target annihilator, P_U^\perp given by

$$P_U^\perp = \mathbf{I} - \mathbf{U}\mathbf{U}^\# = \mathbf{I} - \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T \quad (8)$$

where $\mathbf{U} = [\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_{p-1}]$ is the undesired target matrix and $\mathbf{U}^\#$ is the pseudoinverse of \mathbf{U} given by $(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T$. Assume that a data sample in the original data space be noted by \mathbf{r} and its resulting \mathbf{U} -eliminated data sample be denoted by $\bar{\mathbf{r}}$, that is

$$\bar{\mathbf{r}} = P_U^\perp \mathbf{r}. \quad (9)$$

It is then followed by a matched filter specified by the desired target signature \mathbf{d} given by

$$\mathbf{d}^T \bar{\mathbf{r}} = \mathbf{d}^T P_{\mathbf{U}}^\perp \mathbf{r} \quad (10)$$

which gives rise to an OSP-TD projector developed in [34] and [35]

$$\delta^{\text{OSP-TD}}(\mathbf{r}) = \mathbf{d}^T P_{\mathbf{U}}^\perp \mathbf{r}. \quad (11)$$

Since the OSP-TD projector in (11) requires the prior knowledge of \mathbf{d} and \mathbf{U} , it cannot be directly applied to AD.

B. Automatic Target Generation Process

So, the primary issue with using (11) to perform AD is how to take care of the undesired target matrix \mathbf{U} used in (11) which is unknown. This is probably the main reason that OSP-TD in (11) has never been considered to perform AD in the past.

Fortunately, the commonly used ATGP developed in [42] seems to provide a solution. In fact, ATGP repeatedly implements (11) to find a sequence of $\{\mathbf{t}_k^{\text{ATGP}}\}_k$ which maximizes (11) by replacing \mathbf{U} with $\mathbf{U}_{k-1} = [\mathbf{t}_1^{\text{ATGP}} \mathbf{t}_2^{\text{ATGP}} \dots \mathbf{t}_{k-1}^{\text{ATGP}}]$

$$\mathbf{t}_k^{\text{ATGP}} = \arg \left\{ \max_{\mathbf{r}} \left[\mathbf{r}^T P_{[\mathbf{U}_{k-1}]}^\perp \mathbf{r} \right] \right\} \quad (12)$$

where the undesired target matrix \mathbf{U}_{k-1} in (12) is augmented by previously generated targets $\{\mathbf{t}_j^{\text{ATGP}}\}_{j=1}^{k-1}$ by (11). In this case, \mathbf{U} is generated in an unsupervised manner without its prior knowledge. Using (12) ATGP has been widely used to find endmembers [39]–[41].

Unfortunately, ATGP in (12) cannot be directly used as an anomaly detector due to the fact that it is a pixel-based target detector and does not take the spectral correlation among its detected targets into account as an anomaly detector does. In addition, since L is the number spectral bands, ATGP can find at most L targets via OSP. This is because OSP allows to find one target for one spectral dimension due to orthogonality principle.

C. Constrained Energy Minimization

Another TD which has been shown to be effective in subpixel target detection is the CEM developed in [43]. In analogy with OSP-TD, it also assumes that \mathbf{d} is also the target signature of a desired target to be detected, which is known *a priori*. But unlike OSP-TD in (11) CEM does not assume the prior knowledge of the spectral signatures of undesired targets, $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{p-1}$. Accordingly, it cannot use the $P_{\mathbf{U}}^\perp$ to annihilate undesired target signatures. CEM was developed to overcome this issue by take advantage of a well-known technique, called linearly constrained minimum variance (LCMV) approach in array processing [51]. Basically, CEM is a finite impulse response (FIR) filter which constrains the desired target signature \mathbf{d} while minimizes the variance of signal signatures present in BKG. That is, CEM solves the following constrained optimization:

$$\min_{\mathbf{w}} \mathbf{w}^T \mathbf{R} \mathbf{w} \text{ subject to } \mathbf{d}^T \mathbf{w} = 1 \quad (13)$$

where $\mathbf{R}_{L \times L} = (1/N) \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T$ is the sample auto-correlation matrix of the image data, $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$.

So, the optimal solution to (13) is given by

$$\mathbf{w}^{\text{CEM}} = \frac{\mathbf{R}^{-1} \mathbf{d}}{\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}}. \quad (14)$$

Using the optimal weight, \mathbf{w}^{CEM} as the filter coefficient vector CEM, denoted by $\delta^{\text{CEM}}(\mathbf{r})$ can be derived in [1], [43], and [52] by

$$\delta^{\text{CEM}}(\mathbf{r}) = (\mathbf{w}^{\text{CEM}})^T \mathbf{r} = (\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d})^{-1} (\mathbf{R}^{-1} \mathbf{d})^T \mathbf{r}. \quad (15)$$

By comparing (11) to (15), we immediately discover that OSP-TD and CEM have nearly identical functional forms with $P_{\mathbf{U}}^\perp$ in (11) replaced with \mathbf{R}^{-1} where $(\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d})^{-1}$ in (15) is simply a normalization constant [37], [53]. Relationship between OSP-TD and CEM are shown in Fig. 1 and were further investigated in [54] and [55].

In analogy with OSP, $\delta^{\text{CEM}}(\mathbf{r})$ in (15) also requires the prior knowledge of target signature \mathbf{d} , it cannot be applied to AD either.

D. Anomaly Detection

One important application of target detection is anomaly detection. The well-known and commonly used anomaly detector (AD) is the one developed by Reed and Yu in [4], referred to as RX-AD in this article, which is given by

$$\delta^{\text{RX-AD}}(\mathbf{r}) = (\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \quad (16)$$

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

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

$$\delta^{\text{R-AD}}(\mathbf{r}) = \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \quad (17)$$

referred to as R-AD in [1], [2], and [5] and $\mathbf{R} = (1/N) \sum_{i=1}^N \mathbf{r}_i \mathbf{r}_i^T$.

Interestingly, if we compare (15) to (17), we immediately discover that R-AD and CEM share similar function forms with the desired target signature \mathbf{d} used in (15) replaced by the data sample \mathbf{r} subject to a normalization constant $(\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d})^{-1}$ as shown in Fig. 1. This suggests that R-AD can be viewed as an AD version of CEM.

E. OSP-Anomaly Detector

Now, by comparing OSP-TD in (11) against CEM, the key difference between them is the use of the prior knowledge of the undesired target signatures, \mathbf{U} which is not required by CEM. Apparently, a similar idea of using R-AD to interpret CEM as AD is not applicable. However, according to Fig. 1 it is interesting to see that if we replace \mathbf{R}^{-1} in (17) with $P_{\mathbf{U}}^\perp$ in (11), we simply obtain an alternative OSP-based projector given by

$$\delta^{\text{OSP-AD}}(\mathbf{r}) = \mathbf{r}^T P_{\mathbf{U}}^\perp \mathbf{r} \quad (18)$$

which can be considered as OSP-anomaly detector (OSP-AD) in detecting anomalous target signature without knowing \mathbf{U} . Since

\mathbf{U} is not supposed to be known *a priori* in (18), OSP-AD cannot be implemented as an anomaly detector.

So, how to make (18) an anomaly detector presents a great challenge as shown in Fig. 1, in which case determining the \mathbf{U} in (11) and (18) is a key issue. This issue will be addressed in Section V by appealing for the LRaSMD model.

V. OSP-GO DECOMPOSITION

An earlier development of the LRaSMD model was derived from the video security surveillance system [44]. In hyperspectral data exploitation the LRaSMD model has found its potential in many applications where two models of LRaSMD are generally developed, $\mathbf{X} = \mathbf{L} + \mathbf{S}$ and $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{n}$ along with their variants. In this article, we are particularly interested in the GoDec-generated LRaSMD model.

Although GoDec has shown promising in hyperspectral AD, 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 low-rank matrix m , which represents BKG. The one is how to determine the cardinality of the sparse signals k , which is sparsity of signal sources. With the given values of m and k , GoDec iteratively finds low-rank matrix \mathbf{L} and sparsity matrix \mathbf{S} . Consequently, to effectively utilize GoDec, how to determine these two values is a key success to its use.

In order to resolve the above issues, an OSP version of GoDec, called OSP-GoDec, was developed in [49] where the values of m and k were estimated by the concept of virtual dimensionality (VD) developed in [1] and [56]–[59] is used to estimate the total number of signal sources, p which is the sum of m and j , i.e., $p = m + j$. Then the minimax-SVD (MX-SVD) used in maximum orthogonal complement algorithm (MOCA) developed by Kuybeda *et al.* in [60] with more details in [61] and [62] is used to determine m . As a result, $j = p - m$ from which the k can be determined by $k = j \times N$ [49]. More details can be found in [49].

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. Also let the total number of spectral bands be denoted by $L \leq N$. The low-rank and sparsity matrix decomposition is modeled by

$$\mathbf{X}_{N \times L} = \mathbf{L}_{N \times L} + \mathbf{S}_{N \times L} + \mathbf{n}_{N \times L} \quad (19)$$

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 “low-rank + sparse” decomposition problem in (19) can be solved by minimizing the following constrained optimization decomposition error:

$$\min_{\mathbf{L}_{N \times L}, \mathbf{S}_{N \times L}} \|\mathbf{X}_{N \times L} - \mathbf{L}_{N \times L} - \mathbf{S}_{N \times L}\|_F^2 \quad (20)$$

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 [48] was designed to solve (20) by iteratively solving the following two optimization problems:

$$\begin{aligned} \mathbf{L}_{N \times L}^{(k)} &= \arg \left\{ \min_{\text{rank}(\mathbf{L}_{N \times L}) \leq m} \|\mathbf{X}_{N \times L} - \mathbf{L}_{N \times L} - \mathbf{S}_{N \times L}^{(k-1)}\|_F^2 \right\} \\ \mathbf{S}_{N \times L}^{(k)} &= \arg \left\{ \min_{\text{card}(\mathbf{S}_{N \times L}) \leq k} \|\mathbf{X}_{N \times L} - \mathbf{L}_{N \times L}^{(k)} - \mathbf{S}_{N \times L}\|_F^2 \right\} \end{aligned} \quad (21)$$

GoDec first generates two random matrices, $\Psi_{L \times m}$ and $\Phi_{N \times m}$ to construct $\mathbf{Y}_{N \times m}$ and $\mathbf{Z}_{L \times m}$ given by

$$\mathbf{Y}_{N \times m} = \mathbf{X}_{N \times L} \Psi_{L \times m} \quad (22)$$

$$\mathbf{Z}_{L \times m} = \mathbf{X}_{N \times L}^T \Phi_{N \times m}. \quad (23)$$

Then, $\mathbf{L}_{N \times L}$ can be obtained by

$$\begin{aligned} \mathbf{L}_{N \times L} &= \mathbf{Y}_{N \times m} (\Phi_{N \times m}^T \mathbf{Y}_{N \times m})^{-1} \mathbf{Z}_{L \times m}^T \\ &= \mathbf{X}_{N \times L} \Psi_{L \times m} (\Phi_{N \times m}^T \mathbf{X}_{N \times L} \Psi_{L \times m})^{-1} (\mathbf{X}_{N \times L}^T \Phi_{N \times m})^T. \end{aligned} \quad (24)$$

In order to improve the approximation precision of $\mathbf{L}_{N \times L}$ in (24), we can use the update: $\Phi_{N \times m} = \mathbf{Y}_{N \times m}$ and $\Psi_{L \times m} = \mathbf{Z}_{L \times m}$ and rederive $\mathbf{L}_{N \times L}$ as

$$\begin{aligned} \mathbf{L}_{N \times L} &= \\ &= \mathbf{X}_{N \times L} \Psi_{L \times m} \left((\mathbf{X}_{N \times L} \Psi_{L \times m})^T \mathbf{X}_{N \times L} \Psi_{L \times m} \right)^{-1} \\ &= (\mathbf{X}_{N \times L} \Psi_{L \times m})^T \mathbf{X}_{N \times L} \end{aligned} \quad (25)$$

Now, let

$$\mathbf{U}_{N \times m} = \mathbf{X}_{N \times L} \Psi_{L \times m}. \quad (26)$$

Then (25) becomes

$$\begin{aligned} \mathbf{L}_{N \times L} &= \mathbf{U}_{N \times m} (\mathbf{U}_{N \times m}^T \mathbf{U}_{N \times m})^{-1} \mathbf{U}_{N \times m}^T \mathbf{X}_{N \times L} \\ &= P_{\mathbf{U}_{N \times m}} \mathbf{X}_{N \times L} \end{aligned} \quad (27)$$

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

$$P_{\mathbf{U}_{N \times m}} = \mathbf{U}_{N \times m} (\mathbf{U}_{N \times m}^T \mathbf{U}_{N \times m})^{-1} \mathbf{U}_{N \times m}^T. \quad (28)$$

Then

$$\mathbf{S}_{N \times L} = P_{\Omega} ((\mathbf{I} - P_{\mathbf{U}_{N \times m}}) \mathbf{X}_{N \times L}) = P_{\Omega} (P_{\mathbf{U}_{N \times m}}^{\perp} \mathbf{X}_{N \times L}) \quad (29)$$

where $P_{\Omega} (P_{\mathbf{U}_{N \times m}}^{\perp} \mathbf{X}_{N \times L})$ is the projection of the matrix $P_{\mathbf{U}_{N \times m}}^{\perp} \mathbf{X}_{N \times L}$ onto an entry set Ω which is the nonzero subset of the first k largest entries of $P_{\mathbf{U}_{N \times m}}^{\perp} \mathbf{X}_{N \times L}$. In this case, $\mathbf{L}_{N \times L}$ can be further derived as

$$\mathbf{L}_{N \times L} = P_{\mathbf{U}_{N \times m}} (\mathbf{X}_{N \times L} - \mathbf{S}_{N \times L}) = P_{\mathbf{U}_{N \times m}} \mathbf{X}_{N \times L}. \quad (30)$$

So, when the GoDec [48] is implemented as OSP-GoDec the parameter q in GoDec is set to 0 and there is no need to determine the value of q as GoDec does. A step-by-step implementation of OSP-GoDec is summarized as follows.

VI. OSP-AD IMPLEMENTED WITH LRASMD AND DS

As noted in Section III-E, the issue that OSP-AD cannot be implemented as AD is how to remove the requirement of prior knowledge of the undesired target signatures \mathbf{U} .

In the LRaSMD model generated by OSP-GoDec, the low-rank matrix \mathbf{L} is generally used to characterize BKG as opposed to the sparse matrix \mathbf{S} used to characterize targets of interest. So, we can interpret \mathbf{U} as the BKG subspace constructed from either $\mathbf{L} + \mathbf{S}$ or \mathbf{L} and anomalies as the targets subspace constructed from either $\mathbf{L} + \mathbf{S}$ or \mathbf{S} . In this case, we can use $P_{\mathbf{B}}^{\perp}$ to annihilate BKG, while a data sample \mathbf{r} can be extracted as an anomaly from the target space constructed by \mathbf{A} . With this interpretation,

Algorithm: OSP-GoDec.

1. Initial conditions:

 Prescribe an error threshold, ε
 p determined by VD.

 m estimated by MX-SVD developed in [49].

 $j = p - m$ with $k = j \times N$

 Randomly generate $\Psi_{L \times m}$

 Let $\mathbf{L}_{N \times L}^{(0)} = \mathbf{X}_{N \times L}$, $\mathbf{S}_{N \times L}^{(0)} = \mathbf{O}_{N \times L}$
2. Calculate for $i \geq 1$

$$\mathbf{U}_{N \times m}^{(i)} = (\mathbf{X}_{N \times L} - \mathbf{S}_{N \times L}^{(i-1)}) \Psi_{L \times m} \quad (31)$$

3. In order to implement two iterative loops using (21),

 we need include $\mathbf{S}_{N \times L}^{(i-1)}$ and $\mathbf{L}_{N \times L}^{(i)}$ in (21) as follows:

$$\mathbf{L}_{N \times L}^{(i)} = P_{\mathbf{U}_{N \times m}^{(i)}}(\mathbf{X}_{N \times L} - \mathbf{S}_{N \times L}^{(i-1)}) \quad (32)$$

$$\mathbf{S}_{N \times L}^{(i)} = P_{\Omega}(\mathbf{X}_{N \times L} - \mathbf{L}_{N \times L}^{(i)}) \quad (33)$$

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

 two iterative loops are carried out by beginning the initial condition $\mathbf{L}_{N \times L}^{(0)}$ and then iterating (32) and (33) as follows:

$$\mathbf{L}_{N \times L}^{(0)} \xrightarrow{(33)} \mathbf{S}_{N \times L}^{(0)} \xrightarrow{(32)} \mathbf{L}_{N \times L}^{(1)} \xrightarrow{(33)} \mathbf{S}_{N \times L}^{(1)} \cdots \xrightarrow{(32)} \mathbf{L}_{N \times L}^{(i)} \xrightarrow{(33)} \mathbf{S}_{N \times L}^{(i)} \quad (34)$$

4. Stopping rule

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

the OSP-AD in (18) can be further rederived as an OSP-based anomaly detector (OSP-AD) by

$$\delta_{\mathbf{B}}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{A}}) = \mathbf{r}_{\mathbf{A}}^T P_{\mathbf{B}}^{\perp} \mathbf{r}_{\mathbf{A}}. \quad (35)$$

A. DS and Sparsity Matrix

In step 3 of OSP-GoDec, the sparsity cardinality k is used in (33) to find the sparse matrix, \mathbf{S} . However, since an anomaly is generally characterized by high-order statistics [63], there may be very likely that BKG signal sources could be still leaked into the target subspace specified by \mathbf{S} , $\langle \mathbf{S} \rangle$. So, to further improve OSP-AD in (35) for anomaly detection, we can first sphere the \mathbf{A} in (35) to obtain its sphered dataset $\hat{\mathbf{A}}$ and then implement (35) with $\mathbf{r}_{\mathbf{A}}$ replaced by $\mathbf{r}_{\hat{\mathbf{A}}}$. In this case, target signal sources characterized by the data statistics of the first two orders in the target subspace \mathbf{A} specified by $\langle \mathbf{S} \rangle$ can be removed prior to extracting anomalies in $\langle \hat{\mathbf{S}} \rangle$. This gives rise to an OSPDS-based anomaly detector (OSPDS-AD) which operates the data sample on the sphered target subspace, $\hat{\mathbf{A}}$ as follows:

$$\delta_{\mathbf{B}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{A}}}) = \mathbf{r}_{\hat{\mathbf{A}}}^T P_{\mathbf{B}}^{\perp} \mathbf{r}_{\hat{\mathbf{A}}}. \quad (36)$$

B. DS and Low-Rank Matrix

On many occasions, anomalies may exhibit strong spectral characteristics in all orders of data statistics. Under this circumstance, such anomalies will appear in the subspace, $\langle \mathbf{L} + \mathbf{S} \rangle$ constructed by the joint low rank and sparse matrices $\mathbf{L} + \mathbf{S}$. Using these facts, OSP-AD has two ways to suppress BKG by letting $\mathbf{B} = \mathbf{L}$ or $\mathbf{L} + \mathbf{S}$ and also two ways to specify anomalies by letting $\hat{\mathbf{A}} = \hat{\mathbf{S}}$ or $\hat{\mathbf{L}} + \hat{\mathbf{S}}$. This leads to four versions of OSP-AD that can be implemented for (35), $\delta_{\mathbf{L}}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{S}})$, $\delta_{\mathbf{L} + \mathbf{S}}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{L} + \mathbf{S}})$, $\delta_{\mathbf{L} + \mathbf{S}}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{S}})$, $\delta_{\mathbf{L} + \mathbf{S}}^{\text{OSP-AD}}(\mathbf{r}_{\hat{\mathbf{L}} + \hat{\mathbf{S}}})$ along with their 4 corresponding OSPDS-AD counterparts in (36), $\delta_{\mathbf{L}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}})$, $\delta_{\mathbf{L} + \mathbf{S}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}} + \hat{\mathbf{S}}})$, and $\delta_{\mathbf{L} + \mathbf{S}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}} + \hat{\mathbf{S}}})$.

VII. 3-D ROC ANALYSIS-DERIVED QUANTITATIVE DETECTION MEASURES

3-D ROC analysis has recently developed as an effective evaluation tool for target detection [50] and classification [64]. A 3-D ROC curve of $(P_{\text{D}}, P_{\text{F}}, \tau)$ extends the traditional 2-D ROC curve of $(P_{\text{D}}, P_{\text{F}})$ by including the threshold τ to specify a third dimension. It can be used to generate three new 2-D ROC curves, 2-D ROC curve of $(P_{\text{D}}, P_{\text{F}})$, 2-D ROC curve of (P_{D}, τ) , and 2-D ROC curve of (P_{F}, τ) . The area under the curve (AUC) calculated from 2-D ROC curve of $(P_{\text{D}}, P_{\text{F}})$, 2-D ROC curve of (P_{D}, τ) , and 2-D ROC curve of (P_{F}, τ) , denoted by $\text{AUC}_{(\text{D}, \text{F})}$, $\text{AUC}_{(\text{D}, \tau)}$, $\text{AUC}_{(\text{F}, \tau)}$ respectively, can be further used to design quantitative detection measures.

A. Overall Detection Probability (ODP)

A quantitative detection measure, called overall detection, AUC_{OD} , is defined by

$$-1 \leq \text{AUC}_{\text{ODP}} = \text{AUC}_{(\text{D}, \text{F})} + \text{AUC}_{(\text{D}, \tau)} - \text{AUC}_{(\text{F}, \tau)} \leq 2. \quad (37)$$

B. Target Detectability (TD)

Another measure, called TD of a detector, AUC_{TD} is defined by

$$0 \leq \text{AUC}_{\text{TD}} = \text{AUC}_{(\text{D}, \text{F})} + \text{AUC}_{(\text{D}, \tau)} \leq 2 \quad (38)$$

to quantitatively calculate the effectiveness and TD of a detector.

C. Background Suppressibility (BS)

A detection measure factoring $\text{AUC}_{(\text{F}, \tau)}$ into a detector is called BKG suppressibility (BS) of a detector, AUC_{BS} , defined by

$$-1 \leq \text{AUC}_{\text{BS}} = \text{AUC}_{(\text{D}, \text{F})} - \text{AUC}_{(\text{F}, \tau)} \leq 1. \quad (39)$$

D. TD in BKG Suppression (TDBS)

Since P_{F} is caused by the probability of misdetecting noise as a signal, this error probability should be subtracted from P_{D} . To take care of this effect, a measure that factors P_{F} into P_{D} , called TD in BKG suppression (TDBS), can be defined as AUC_{TDBS} by

$$-1 \leq \text{AUC}_{\text{TDBS}} = \text{AUC}_{(\text{D}, \tau)} - \text{AUC}_{(\text{F}, \tau)} \leq 1. \quad (40)$$

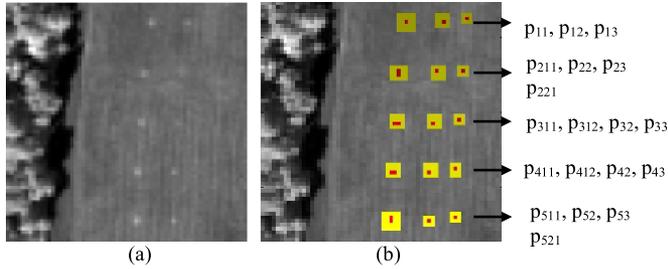


Fig. 2. (a) HYDICE panel scene which contains 15 panels. (b) Ground truth map of spatial locations of the 15 panels.

E. Signal-To-Noise Probability Ratio (SNPR)

The most effective detection measure comes from a similar idea that is widely used in communications/signal processing, SNR. It is called SNPR by

$$0 \leq \text{AUC}_{\text{SNPR}} = \frac{\text{AUC}_{(D,\tau)}}{\text{AUC}_{(F,\tau)}} \quad (41)$$

where the false alarm probability P_F can be assumed to be caused by noise.

VIII. IMAGES TO BE STUDIED FOR EXPERIMENTS

A. HYDICR 15-Panel Scene

The image scene shown in Fig. 2 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 [1] and [2]. There are 15 square panels with three different sizes— 3×3 m, 2×2 m, and 1×1 m, respectively, shown in Fig. 2(a). Due to the approximately 1.56 m ground sampling distance each of panels in the first column except the first row contains two panel pixels highlighted by red, p_{211} , p_{221} in row 2, p_{311} , p_{312} in row 3, p_{411} , p_{412} in row 4, p_{511} , p_{521} in row 5 as shown in Fig. 2(b) by ground truth. All the 11 remaining panels in Fig. 2(b) contain one single panel pixel for each panel also highlighted by red, p_{11} , p_{12} , p_{13} in row 1, p_{22} , p_{23} in row 2, p_{32} , p_{33} in row 3, p_{42} , p_{43} in row 4, p_{52} , p_{53} in row 5. Therefore, there are a total of 19 panel pixels. Fig. 2(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 dataset is another HYDICE image with pseudocolor shown in Fig. 3(a). It is an urban scene and comprised 210 spectral bands with 174 bands being used for experiments after the noise and water absorption bands had been removed. A region with a size of 80×100 pixels located at the upper right of the scene was selected as the test image shown in Fig. 3(b) along with the ground-truth map shown in Fig. 3(c) where 21 pixels were identified as anomalies, which were cars and roofs, because they had spectra that differ from BKG.

OSP-GoDec was used to decompose the HSI data \mathbf{X} into $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{n}$. The OSP-AD in (35) and OSPDS-AD in (36)



Fig. 3. HYDICE urban scene. (a) pseudo-color image of the whole scene. (b) pseudo-color image of the selected area. (c) the ground truth map.

TABLE I
M ESTIMATED FOR HYDICE DATA BY MOCA, HFC/NWHFC USING MX-SVD

Data	VD	p	m by MX-SVD
HYDICE 15-panel	MOCA	32	24
	HFC, $P_F \leq 10^{-3}$	9	5
	NWHFC $P_F \leq 10^{-3}$	13	7
HYDICE Urban	MOCA	61	35
	HFC, $P_F = 10^{-4}$	9	5
	NWHFC $P_F = 10^{-4}$	13	7

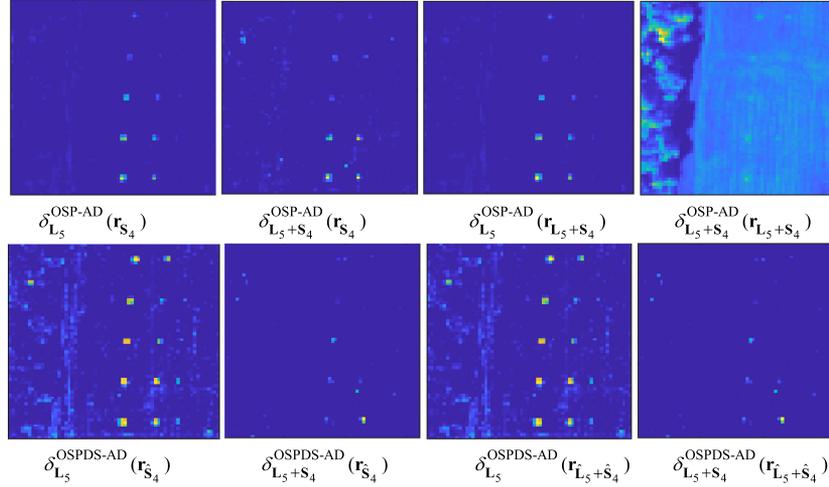
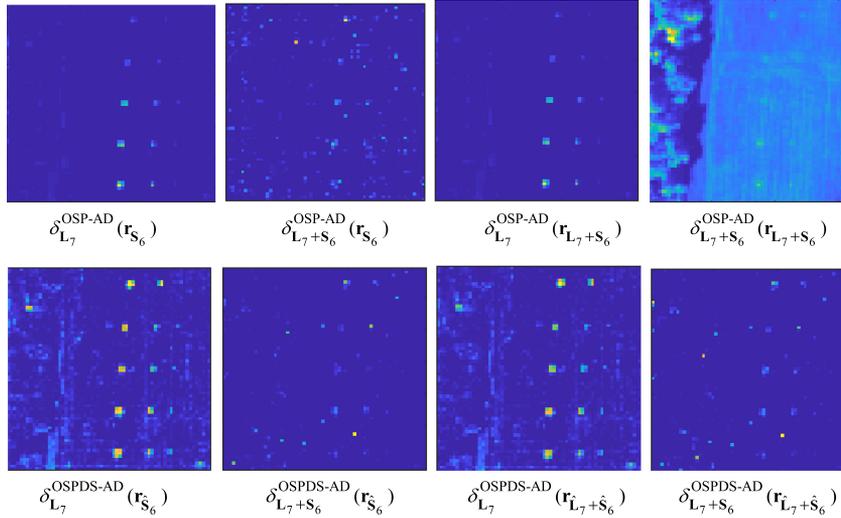
were then implemented to perform anomaly detection. The two key parameters of OSP-GoDec, p and m are estimated by HFC/NWHFC/MOCA with their values tabulated in shown in Table I.

It should be noted that j can be determined by $j = p - m$ which in turn determines k by $k = j \times N$.

IX. EXPERIMENTAL RESULTS AND DISCUSSIONS

In order to evaluate the performance of OSP-AD in (35) and OSPDS-AD in (36), a sparse representation-based model, referred to as l_2 -norm minimization and distance weighted regularization matrix and sum-to-one constraint (CRD-DW-STO), developed in [10] was used for comparison since it was shown to outperform sparse representation-based detector (SRD), RPCA-based anomaly detector, and local RX-AD. In addition, classic RX-AD developed by Reed and Yu in [4] and R-AD were also included as a benchmark comparison.

Eight versions of anomaly detectors derived from (35) and (36), $\delta_{\mathbf{L}}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{S}})$, $\delta_{\mathbf{L}}^{\text{OSPDS-AD}}(\mathbf{r}_{\mathbf{S}})$, $\delta_{\mathbf{L}+\mathbf{S}}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{L}+\mathbf{S}})$, $\delta_{\mathbf{L}}^{\text{OSPDS-AD}}(\mathbf{r}_{\mathbf{L}+\mathbf{S}})$, $\delta_{\mathbf{L}+\mathbf{S}}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{S}})$, $\delta_{\mathbf{L}+\mathbf{S}}^{\text{OSPDS-AD}}(\mathbf{r}_{\mathbf{S}})$, $\delta_{\mathbf{L}+\mathbf{S}}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{L}+\mathbf{S}})$, $\delta_{\mathbf{L}+\mathbf{S}}^{\text{OSPDS-AD}}(\mathbf{r}_{\mathbf{L}+\mathbf{S}})$ were used to conduct a comparative study and analysis where the low-rank matrix \mathbf{L} and sparse matrix \mathbf{S} constructed by the LRaSMD models generated by OSP-GoDec were also used by OSP-AD in (35)


 Fig. 4. Detection maps of HYDICE 15-panel scene by OSP-AD and OSPDS-AD with $p = 9$, $m = 5$, and $j = 4$.

 Fig. 5. Detection maps of HYDICE 15-panel scene by OSP-AD and OSPDS-AD with $p = 13$, $m = 7$, and $j = 6$.

and OSPDS-AD in (36) to specify each of the eight anomaly detectors. In addition, upon implementing OSP-GoDec, the three sets of the values estimated for the parameters p , j , and k in Table I were used for experiments. According to [49] RX/R-AD can be also implemented by taking advantage of \mathbf{A} and \mathbf{B} in a similar manner that OSP-AD is implemented in (35) as

$$\delta_{\mathbf{B}}^{\text{RX-AD}}(\mathbf{r}_{\mathbf{A}}) = (\mathbf{r}_{\mathbf{A}} - \mu_{\mathbf{A}})^T \mathbf{K}_{\mathbf{B}}^{-1} (\mathbf{r}_{\mathbf{A}} - \mu_{\mathbf{A}}) \quad (42)$$

$$\delta_{\mathbf{B}}^{\text{R-AD}}(\mathbf{r}_{\mathbf{A}}) = \mathbf{r}_{\mathbf{A}}^T \mathbf{R}_{\mathbf{B}}^{-1} \mathbf{r}_{\mathbf{A}}. \quad (43)$$

Based on the results in [49], (42) and (43) generally outperformed (16) and (17) if \mathbf{A} and \mathbf{B} were selected appropriately. In the following experiments (42) and (43) were implemented by selecting specific \mathbf{A} and \mathbf{B} to produce the best possible results.

A. HYDICE 15-Panel Scene

Figs. 4–6 show detection maps of the eight anomaly detectors using the three sets of values of p , m , j given by

Table I and the LRA-SMD models produced by OSP-GoDec where the best anomaly detectors were $\delta_{\mathbf{L}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}})$ and $\delta_{\mathbf{L}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}+\hat{\mathbf{S}}})$ by visual inspection for all the three scenarios. Fig. 7 shows the detection maps of RX/R-AD in (42) and (43) using $\mathbf{A} = \mathbf{B} = \mathbf{L}_m + \mathbf{S}_j$ where their detection results were comparable to the best results produced by $\delta_{\mathbf{L}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}})$ and $\delta_{\mathbf{L}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}+\hat{\mathbf{S}}})$ in Figs. 4–6 but had more falsely alarm detected data samples in BKG.

The visual inspection of Figs. 4–7 only provides qualitative analysis. On many occasions, a detector may detect all anomalies but may also detect many BKG pixels as falsely alarmed anomalies. In this case, it will end up with poor BKG suppression. To address this issue, Tables II–IV calculated $\text{AUC}_{(\text{D},\text{F})}$, $\text{AUC}_{(\text{D},\tau)}$, $\text{AUC}_{(\text{F},\tau)}$, AUC_{ODP} , AUC_{TD} , AUC_{BS} , AUC_{TDBS} , AUC_{SNPR} in Section VII to evaluate the effectiveness of a detector in TD and BS where the best results are boldfaced. As shown in Tables II–IV the results for three scenarios were close but overall, the results produced by $p = 13$, $m = 7$, and $j = 6$ were

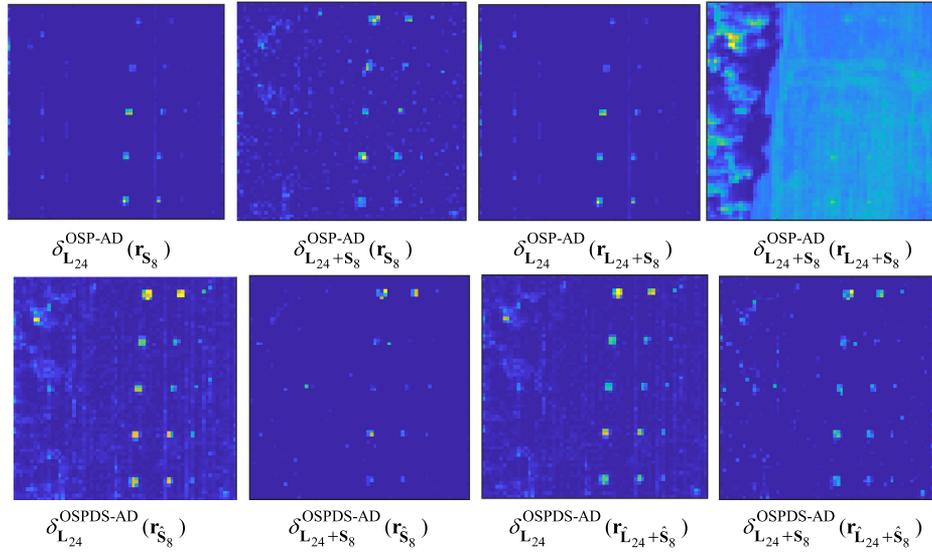


Fig. 6. Detection maps of HYDICE 15-panel scene by OSP-AD and OSPDS-AD with $p = 32$, $m = 24$, and $j = 8$.

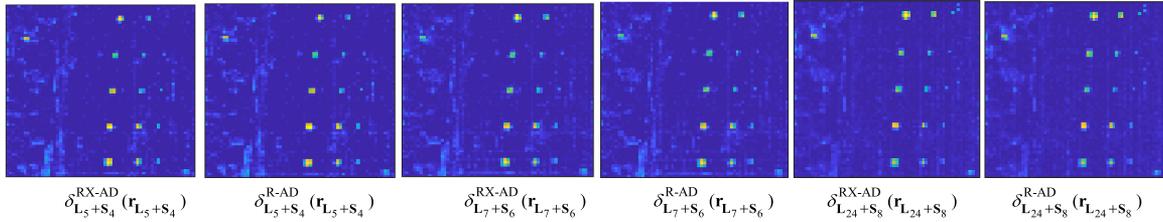


Fig. 7. Detection maps of HYDICE 15-panel scene produced by RX/R-AD in (42) and (43) using $\mathbf{A} = \mathbf{B} = \mathbf{L}_m + \mathbf{S}_j$.

TABLE II
AUC OF HYDICE 15-PANEL SCENE USING OSP-GoDEC WITH $P = 9$, $M = 5$, AND $J = 4$

DETECTOR	AUC _(D_sF)	AUC _(D_sT)	AUC _(F_sT)	AUC _{ODP}	AUC _{TD}	AUC _{BS}	AUC _{TDBS}	AUC _{SNPR}
$\delta_{L_m+S_j}^{RX-AD}(\mathbf{r}_{L_m+S_j})$	0.9926	0.6182	0.0263	1.5845	1.6108	0.9663	0.5919	23.5384
$\delta_{L_m+S_j}^{R-AD}(\mathbf{r}_{L_m+S_j})$	0.9926	0.6180	0.0261	1.5845	1.6106	0.9665	0.5919	23.6669
$\delta_{L_m}^{OSP-AD}(\mathbf{r}_{S_j})$	0.9784	0.2672	0.0048	1.2408	1.2456	0.9736	0.2624	55.5187
$\delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{S_j})$	0.9870	0.6496	0.0269	1.6097	1.6366	0.9601	0.6228	24.1891
$\delta_{L_m+S_j}^{OSP-AD}(\mathbf{r}_{S_j})$	0.9649	0.2567	0.0043	1.2173	1.2216	0.9606	0.2524	59.6192
$\delta_{L_m+S_j}^{OSPDS-AD}(\mathbf{r}_{S_j})$	0.9533	0.0393	0.0016	0.9910	0.9926	0.9517	0.0376	24.1993
$\delta_{L_m}^{OSP-AD}(\mathbf{r}_{L_m+S_j})$	0.9784	0.2672	0.0048	1.2408	1.2456	0.9736	0.2624	55.5458
$\delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{L_m+\hat{S}_j})$	0.9871	0.6398	0.0243	1.6026	1.6269	0.9628	0.6155	26.3314
$\delta_{L_m+S_j}^{OSP-AD}(\mathbf{r}_{L_m+S_j})$	0.9242	0.4016	0.2414	1.0843	1.3258	0.6827	0.1602	1.6634
$\delta_{L_m+S_j}^{OSPDS-AD}(\mathbf{r}_{L_m+\hat{S}_j})$	0.9563	0.0409	0.0017	0.9955	0.9972	0.9545	0.0392	23.4081

TABLE III
AUC OF HYDICE 15-PANEL SCENE USING OSP-GoDEC WITH $P = 13$, $M = 7$, AND $J = 6$

DETECTOR	AUC _(D,F)	AUC _(D,τ)	AUC _(F,τ)	AUC _{ODP}	AUC _{TD}	AUC _{BS}	AUC _{TDBS}	AUC _{SNPR}
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{RX-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9925	0.5801	0.0292	1.5434	1.5726	0.9633	0.5509	19.8499
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{R-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9925	0.5797	0.0291	1.5431	1.5722	0.9634	0.5506	19.9481
$\delta_{\mathbf{L}_m}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9875	0.2721	0.0046	1.2550	1.2596	0.9829	0.2675	59.0987
$\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9939	0.6669	0.0351	1.6256	1.6608	0.9588	0.6317	18.9794
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9711	0.1643	0.0078	1.1276	1.1354	0.9633	0.1565	21.0457
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{OSPDS-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9720	0.1071	0.0037	1.0754	1.0791	0.9683	0.1034	28.9533
$\delta_{\mathbf{L}_m}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9875	0.2721	0.0046	1.2550	1.2596	0.9829	0.2675	59.1018
$\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$	0.9942	0.5877	0.0286	1.5534	1.5819	0.9656	0.5592	20.5851
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9279	0.4238	0.2441	1.1075	1.3516	0.6837	0.1796	1.7356
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$	0.9726	0.1175	0.0047	1.0854	1.0901	0.9679	0.1128	24.9949

TABLE IV
AUC OF HYDICE 15-PANEL SCENE USING OSP-GoDEC WITH $P = 32$, $M = 24$, AND $J = 8$

DETECTOR	AUC _(D,F)	AUC _(D,τ)	AUC _(F,τ)	AUC _{ODP}	AUC _{TD}	AUC _{BS}	AUC _{TDBS}	AUC _{SNPR}
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{RX-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9962	0.5946	0.0310	1.5598	1.5908	0.9652	0.5636	19.2098
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{R-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9962	0.5941	0.0307	1.5595	1.5903	0.9654	0.5633	19.3206
$\delta_{\mathbf{L}_m}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9887	0.2938	0.0067	1.2759	1.2826	0.9821	0.2872	44.0699
$\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9945	0.6164	0.0396	1.5714	1.6109	0.9549	0.5768	15.5688
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9871	0.3764	0.0156	1.3479	1.3634	0.9715	0.3608	24.2022
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{OSPDS-AD}}(\mathbf{r}_{\mathbf{S}_j})$	0.9915	0.1991	0.0034	1.1872	1.1906	0.9881	0.1956	57.9215
$\delta_{\mathbf{L}_m}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9887	0.2938	0.0067	1.2759	1.2826	0.9821	0.2872	44.0699
$\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$	0.9953	0.5421	0.0285	1.5089	1.5374	0.9668	0.5135	19.0187
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{OSP-AD}}(\mathbf{r}_{\mathbf{L}_m+\mathbf{S}_j})$	0.9133	0.4566	0.2854	1.0845	1.3699	0.6279	0.1712	1.6000
$\delta_{\mathbf{L}_m+\mathbf{S}_j}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$	0.9918	0.3047	0.0096	1.2868	1.2964	0.9821	0.2951	31.6999

slightly better than the other two. It is interesting to see that by RX/R-AD produced the best or very high AUC_(D,F) values with also very high AUC_(F,τ) values and low AUC_{SNPR} values. This indicates that RX/R-AD produced very high AUC_(D,F) values at the expense of high AUC_(F,τ) values. In addition, their AUC values calculated from other detection measures were also not competitive compared to the best results obtained by OSPDS-AD. Unfortunately, many works on anomaly detection reported in the literature only used AUC_(D,F) to measure performance of

an anomaly detector and drew their conclusions solely based on AUC_(D,F) values, which is very likely to mislead incorrect conclusions. The above experiments provided such examples of how unreliable using only AUC_(D,F) for performance evaluation is.

According to the results presented in Tables II–IV, we can conclude that when OSPDS-AD in (36) is implemented as an anomaly detector, the BKG, \mathbf{B} used in (36) must be constructed from the low-rank matrix \mathbf{L}_m and the \mathbf{A} used in (36) must be constructed from either sphered $\hat{\mathbf{S}}_j$ or sphered $\hat{\mathbf{L}}_m + \hat{\mathbf{S}}_j$.

TABLE V
COMPARATIVE PERFORMANCE ANALYSIS BETWEEN $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}_j})$ AND $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$ AGAINST CRD-DW-STO FOR HYDICE 15-PANEL SCENE

	(p,m,j)	AUC _(D,F)	AUC _(D,τ)	AUC _(F,τ)	AUC _{ODP}	AUC _{TD}	AUC _{BS}	AUC _{TDBS}	AUC _{SNPR}
$\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}_j})$	(9,5,4)	0.9870	0.6496	0.0269	1.6097	1.6366	0.9601	0.6228	24.1891
	(13,7,6)	0.9939	0.6669	0.0351	1.6256	1.6608	0.9588	0.6317	18.9794
	(32,24,8)	0.9945	0.6164	0.0396	1.5714	1.6109	0.9549	0.5768	15.5688
$\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$	(9,5,4)	0.9871	0.6398	0.0243	1.6026	1.6269	0.9628	0.6155	26.3314
	(13,7,6)	0.9942	0.5877	0.0286	1.5534	1.5819	0.9656	0.5592	20.5851
	(32,24,8)	0.9953	0.5421	0.0285	1.5089	1.5374	0.9668	0.5135	19.0187
CRD-DW-STO	$(w_{\text{outer}}, w_{\text{inner}}) = (11,7)$	0.9953	0.4673	0.0274	1.4352	1.4626	0.9679	0.4399	17.0547
	RX-AD	0.9898	0.3574	0.0430	1.3042	1.3472	0.9468	0.3144	8.3116
	R-AD	0.9900	0.3584	0.0435	1.3049	1.3484	0.9465	0.3149	8.2391

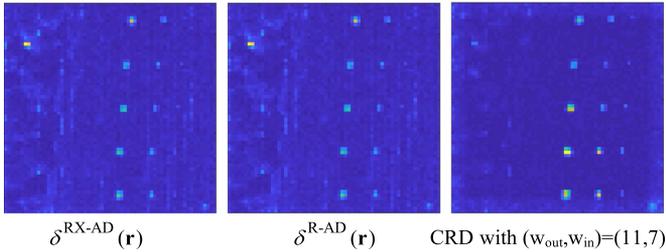


Fig. 8. Detection maps of RX/R-AD and CRD-DW-STO for HYDICE 15-panel scene.

Recall that signal sources are assumed to be embedded in either the sparse space \mathbf{S}_j or the space constructed from joint low-rank matrix \mathbf{L}_m and sparse matrix \mathbf{S}_j and also characterized by higher order statistics. In this case, \mathbf{L}_m and sparse matrix \mathbf{S}_j must be sphered to remove their second-order statistics. It explains why BKG must be estimated from the second-order data statistics and be removed to bring anomalies out of BKG. It is worth noting that there is no need of performing DS on the data space for R-AD. This is because inverting the sample autocorrelation matrix \mathbf{R} in (17) to suppress BKG has a similar effect as DS performed on the data space to remove the second-order data statistics. It also indicates that there is no need of DS implemented in RX-AD. As a result, $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}_j})$ and $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$ indeed performed very effectively. In order to see that, we also conducted experiments to compare these two OSPDS-based anomaly detectors against one of most effective and widely used anomaly detectors, which is based on a sparse representation-based model, referred to as l_2 -norm minimization and distance weighted regularization matrix and sum-to-one constraint (CRD-DW-STO) developed in [10] because it was shown to outperform SRD, RPCA-based anomaly detector, and RX/R-AD. Since all these results are available in [49], only the results of CRD-DW-STO using an outer window size $w_{\text{outer}} = 11$ and an inner window size $w_{\text{inner}} = 7$ are included for comparison. Fig. 8 shows the detection results of RX-AD, R-AD, and CRD-DW-STO.

Table V tabulates the results produced by $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}_j})$, $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$ with $(p,m,j) = (9,5,4), (13,7,6), (32,24,8)$ and CRD-DW-STO with $(w_{\text{outer}}, w_{\text{inner}}) = (11,7)$ where the best results are boldfaced. Also included at the bottom rows of

Table V are the detection maps of classic RX/R-AD in (16) and (17) for comparison.

As shown in Table V, the best results were produced by $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}_j})$, $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$ using OSP-GoDec with $(p,m,j) = (13,7,6)$ and the results of RX/R-AD were nowhere close to that produced by OSPDS-AD. In addition, the results of CRD-DW-STO were not comparable to the results of $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}_j})$ and $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$. However, if only the AUC_(D,F) value based on Table V is used to evaluate their performance, it may lead to a conclusion that CRD-DW-STO performed better than $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}_j})$, $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$ and RX/R-AD also performed very comparably to all other anomaly detectors within a negligible error threshold within 10^{-3} . In this case, it is very difficult to judge which one is better than another. But if AUC_(D, τ), AUC_(F, τ), AUC_{ODP}, AUC_{TD}, AUC_{BS}, AUC_{TDBS}, and AUC_{SNPR} are included as additional detection measures for performance evaluation, these three types of anomaly detectors behaved quite different. In particular, $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}_j})$ and $\delta_{\mathbf{L}_m}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}_m+\hat{\mathbf{S}}_j})$ using OSP-GoDec with $(p,m,j) = (13,7,6)$ outperformed CRD-DW-STO in nearly every category in Table V except AUC_(D,F) and AUC_{BS} which were only slightly better by less than 10^{-3} respectively.

Also shown in Table V, when implementing OSP in conjunction with the OSP-GoDec generated LRA-SMD model along with DS as an anomaly detector, the resultant OSP-AD can be very effective and even performs better than the existing state-of-the-art anomaly detectors.

B. HYDICE Urban Scene

Same experiments conducted for the HYDICE 15-panel scene in Fig. 2 were also performed for the HYDICE urban scene in Fig. 3. Figs. 9–11 show detection maps of the eight versions of OSP-AD using the three sets of (p,m,j) values given in Table I where the best anomaly detectors seemed to be $\delta_{\mathbf{L}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{S}}})$ and $\delta_{\mathbf{L}}^{\text{OSPDS-AD}}(\mathbf{r}_{\hat{\mathbf{L}}+\hat{\mathbf{S}}})$ by visual inspection. For comparison, $\delta_{\mathbf{B}}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{A}})$ were also implemented as their best cases by letting $\mathbf{A} = \mathbf{S}_j$ and $\mathbf{B} = \mathbf{L}_m + \mathbf{S}_j$. Their detection maps are shown in Fig. 12.

Unlike the results obtained from the HYDICE 15-panel scene, which showed better visual inspection in Figs. 4–8, it is very difficult to assess the effectiveness of various anomaly detectors solely based on visual evaluation from Figs. 9–12. In this case, we must rely on the 3-D ROC-derived quantitative

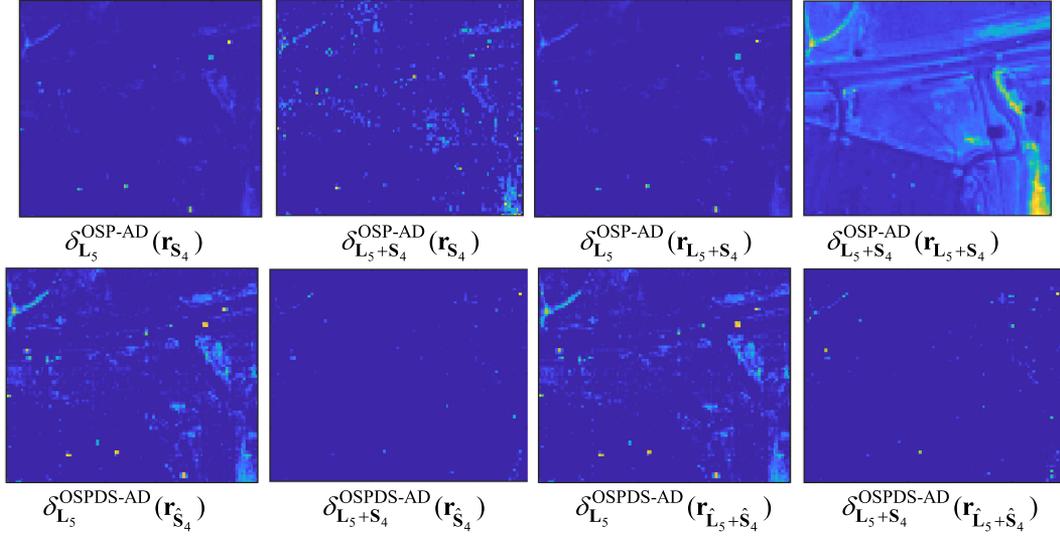


Fig. 9. Detection maps of HYDICE urban scene by OSP-AD and OSPDS-AD $p = 9$, $m = 5$, and $j = 4$.

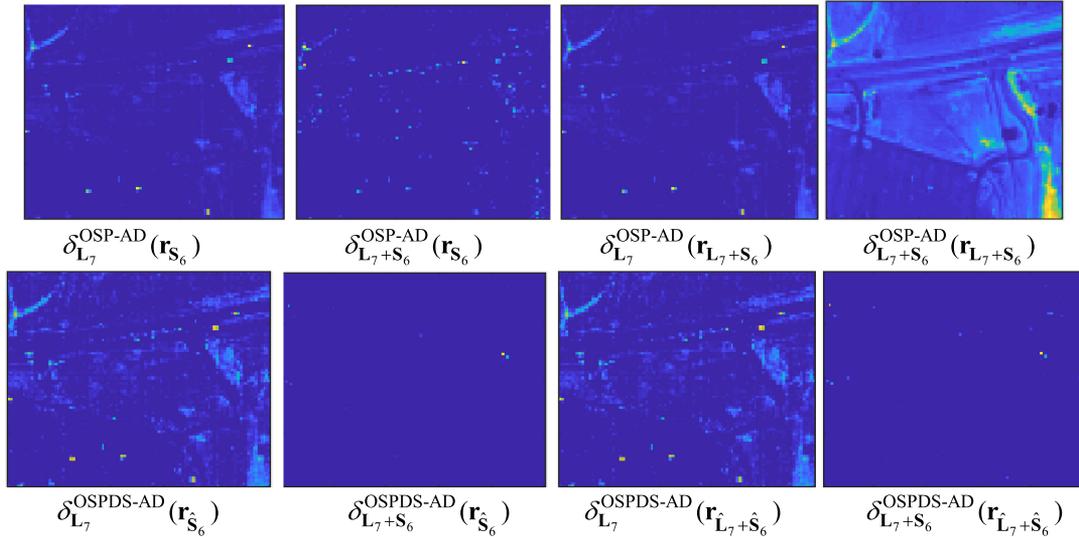


Fig. 10. Detection maps of HYDICE urban scene by OSP-AD and OSPDS-AD with $p = 13$, $m = 7$, and $j = 6$.

detection measures in Section VII to calculate $AUC_{(D,F)}$, $AUC_{(D,\tau)}$, $AUC_{(F,\tau)}$, AUC_{ODP} , AUC_{TD} , AUC_{BS} , AUC_{TDBS} , AUC_{SNPR} , which were used to evaluate the effectiveness of the test anomaly detectors in terms of their TD and BS. Tables VI–VIII tabulate their detection results with the best results boldfaced where the \mathbf{L} and \mathbf{S} generated by OSP-GoDec were used to specify \mathbf{A} and \mathbf{B} for OSP-AD in (35) and OSPDS-AD in (36) and for RX/R-AD in (42) and (43).

As shown in Tables VI–VIII, it seemed that $\delta_{L_5}^{OSPDS-AD}(r_{S_4})$ produced the overall best results for $(p,m,j) = (9,5,4)$ and performed better than $\delta_{L_m}^{OSPDS-AD}(r_{L_m+S_j})$. Like the HYDICE experiments, RX/R-AD in (42) and (43) also produced very high $AUC_{(D,F)}$ values at the expense of high $AUC_{(F,\tau)}$ values. These experiments further provided evidence that solely relying on $AUC_{(D,F)}$ values to evaluate a detector is not reliable.

For comparison, Fig. 13 shows the detection maps produced by RX/R-AD and CRD-DW-STO with $(w_{outer}, w_{winner}) = (11,9)$ where it is very difficult to see which one performed better than others.

In this case, we once again calculated $AUC_{(D,F)}$, $AUC_{(D,\tau)}$, $AUC_{(F,\tau)}$, AUC_{ODP} , AUC_{TD} , AUC_{BS} , AUC_{TDBS} , AUC_{SNPR} for performance evaluation. Table IX tabulates the detection results produced by CRD-DW-STO with $(w_{outer}, w_{winner}) = (11,9)$ where the best results are boldfaced. For comparison, also included in Table IX are the results produced by $\delta_{L_m}^{OSPDS-AD}(r_{S_j})$, $\delta_{L_m}^{OSPDS-AD}(r_{L_m+S_j})$ with $(p,m,j) = (9,5,4)$, $(13,7,6)$, $(61,35,26)$ and the detection results of classic RX/R-AD in (16) and (17). Once again, the best results were produced by $\delta_{L_m}^{OSPDS-AD}(r_{S_j})$ followed by $\delta_{L_m}^{OSPDS-AD}(r_{L_m+S_j})$ for $(p,m,j) = (9,5,4)$. CRD-DW-STO performed slightly higher

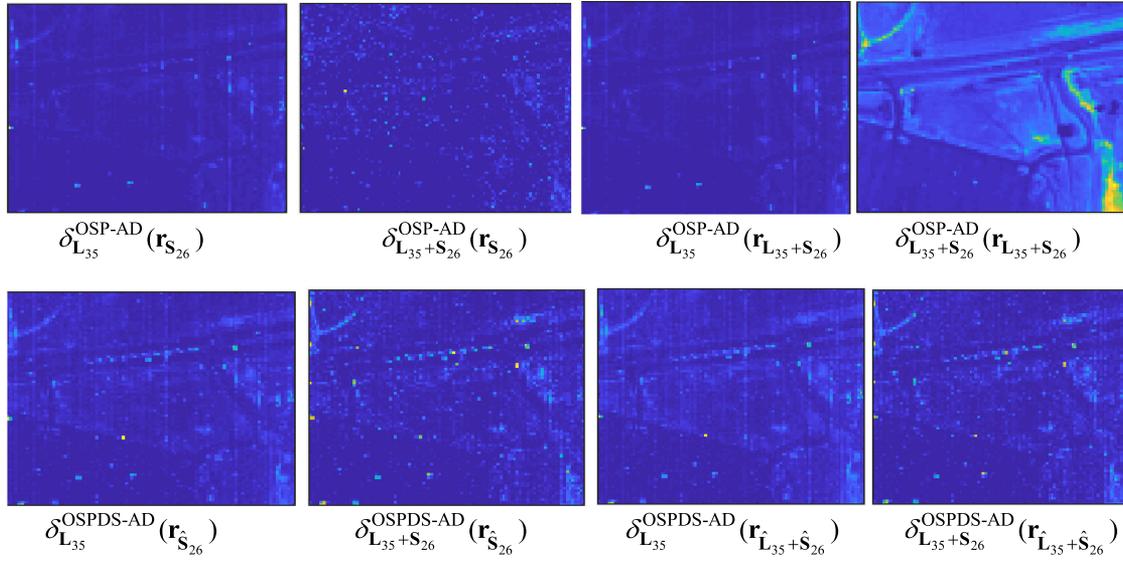


Fig. 11. Detection maps of HYDICE urban scene by OSP-AD and OSPDS-AD with $p = 61$, $m = 35$, and $j = 26$.

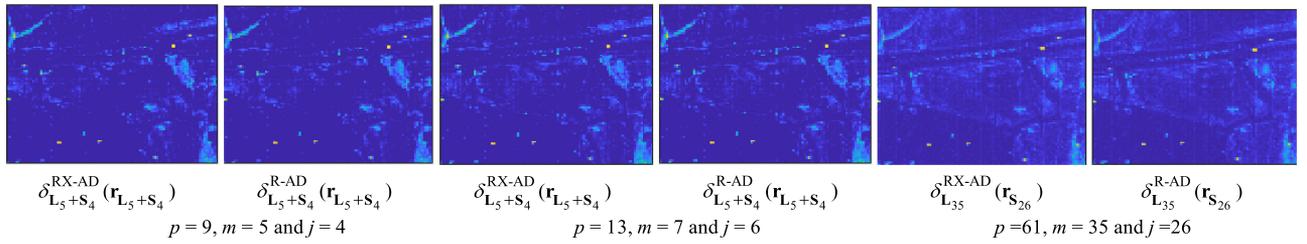


Fig. 12. Detection maps of RX/R-AD in (41) and (42) for HYDICE urban scene.

TABLE VI
AUC OF HYDICE URBAN SCENE USING OSP-GoDEC WITH $P = 9$, $M = 5$, AND $J = 4$

DETECTOR	AUC _(D,F)	AUC _(D,τ)	AUC _(F,τ)	AUC _{ODP}	AUC _{TD}	AUC _{BS}	AUC _{TDBS}	AUC _{SNPR}
$\delta_{L_m+S_j}^{RX-AD}(r_{L_m+S_j})$	0.9898	0.6172	0.0267	1.5802	1.6070	0.9631	0.5904	23.0986
$\delta_{L_m+S_j}^{R-AD}(r_{L_m+S_j})$	0.9898	0.6170	0.0266	1.5802	1.6068	0.9632	0.5904	23.1954
$\delta_{L_m}^{OSP-AD}(r_{S_j})$	0.9868	0.3203	0.0063	1.3008	1.3071	0.9805	0.3140	50.9060
$\delta_{L_m}^{OSPDS-AD}(r_{S_j})$	0.9892	0.6377	0.0248	1.6020	1.6268	0.9643	0.6128	25.6640
$\delta_{L_m+S_j}^{OSP-AD}(r_{S_j})$	0.9462	0.2043	0.0173	1.1331	1.1505	0.9289	0.1870	11.7925
$\delta_{L_m+S_j}^{OSPDS-AD}(r_{S_j})$	0.9661	0.0581	0.0012	1.0230	1.0242	0.9649	0.0569	48.3581
$\delta_{L_m}^{OSP-AD}(r_{L_m+S_j})$	0.9868	0.3203	0.0063	1.3008	1.3071	0.9805	0.3140	50.9886
$\delta_{L_m}^{OSPDS-AD}(r_{L_m+S_j})$	0.9890	0.6196	0.0241	1.5845	1.6087	0.9649	0.5955	25.6959
$\delta_{L_m+S_j}^{OSP-AD}(r_{L_m+S_j})$	0.6627	0.1933	0.1412	0.7147	0.8559	0.5214	0.0520	1.3683
$\delta_{L_m+S_j}^{OSPDS-AD}(r_{L_m+S_j})$	0.9698	0.1608	0.0022	1.1285	1.1306	0.9677	0.1586	74.1973

TABLE VII
AUC OF HYDICE URBAN SCENE USING OSP-GoDEC WITH $P = 13, M = 7,$ AND $J = 6$

DETECTOR	AUC _(D,F)	AUC _(D,τ)	AUC _(F,τ)	AUC _{ODP}	AUC _{TD}	AUC _{BS}	AUC _{TDBS}	AUC _{SNPR}
$\delta_{L_m+S_j}^{RX-AD}(\mathbf{r}_{L_m+S_j})$	0.9942	0.6382	0.0321	1.6002	1.6323	0.9620	0.6060	19.8601
$\delta_{L_m+S_j}^{R-AD}(\mathbf{r}_{L_m+S_j})$	0.9941	0.6378	0.0321	1.5999	1.6320	0.9621	0.6057	19.8782
$\delta_{L_m}^{OSP-AD}(\mathbf{r}_{S_j})$	0.9879	0.3675	0.0149	1.3405	1.3554	0.9730	0.3526	24.6735
$\delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{S_j})$	0.9943	0.5948	0.0292	1.5599	1.5891	0.9651	0.5656	20.4027
$\delta_{L_m+S_j}^{OSP-AD}(\mathbf{r}_{S_j})$	0.9795	0.1977	0.0061	1.1711	1.1772	0.9733	0.1916	32.1695
$\delta_{L_m+S_j}^{OSPDS-AD}(\mathbf{r}_{S_j})$	0.9749	0.0035	0.0004	0.9780	0.9784	0.9745	0.0031	8.9790
$\delta_{L_m}^{OSP-AD}(\mathbf{r}_{L_m+S_j})$	0.9879	0.3675	0.0149	1.3405	1.3554	0.9730	0.3526	24.6769
$\delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{L_m+\hat{S}_j})$	0.9944	0.5883	0.0285	1.5543	1.5827	0.9659	0.5599	20.6613
$\delta_{L_m+S_j}^{OSP-AD}(\mathbf{r}_{L_m+S_j})$	0.6232	0.1704	0.1435	0.6502	0.7936	0.4798	0.0269	1.1878
$\delta_{L_m+S_j}^{OSPDS-AD}(\mathbf{r}_{L_m+\hat{S}_j})$	0.9829	0.0312	0.0007	1.0134	1.0141	0.9822	0.0304	42.2041

TABLE VIII
AUC OF HYDICE URBAN SCENE USING OSP-GoDEC WITH $P = 61, M = 35,$ AND $J = 26$

DETECTOR	AUC _(D,F)	AUC _(D,τ)	AUC _(F,τ)	AUC _{ODP}	AUC _{TD}	AUC _{BS}	AUC _{TDBS}	AUC _{SNPR}
$\delta_{L_m+S_j}^{RX-AD}(\mathbf{r}_{L_m+S_j})$	0.9893	0.5737	0.0516	1.5114	1.5630	0.9377	0.5220	11.1091
$\delta_{L_m+S_j}^{R-AD}(\mathbf{r}_{L_m+S_j})$	0.9894	0.5729	0.0517	1.5106	1.5623	0.9376	0.5212	11.0799
$\delta_{L_m}^{OSP-AD}(\mathbf{r}_{S_j})$	0.9794	0.2229	0.0202	1.1821	1.2023	0.9593	0.2027	11.0496
$\delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{S_j})$	0.9841	0.3409	0.0362	1.2888	1.3250	0.9479	0.3046	9.4086
$\delta_{L_m+S_j}^{OSP-AD}(\mathbf{r}_{S_j})$	0.9494	0.1424	0.0204	1.0713	1.0918	0.9290	0.1220	6.9713
$\delta_{L_m+S_j}^{OSPDS-AD}(\mathbf{r}_{S_j})$	0.9791	0.4166	0.0349	1.3608	1.3957	0.9442	0.3818	11.9496
$\delta_{L_m}^{OSP-AD}(\mathbf{r}_{L_m+S_j})$	0.9794	0.2229	0.0202	1.1821	1.2023	0.9593	0.2027	11.0496
$\delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{L_m+\hat{S}_j})$	0.9834	0.3353	0.0462	1.2725	1.3187	0.9372	0.2892	7.2631
$\delta_{L_m+S_j}^{OSP-AD}(\mathbf{r}_{L_m+S_j})$	0.6600	0.1908	0.1409	0.7098	0.8508	0.5190	0.0498	1.3536
$\delta_{L_m+S_j}^{OSPDS-AD}(\mathbf{r}_{L_m+\hat{S}_j})$	0.9856	0.4082	0.0381	1.3557	1.3938	0.9475	0.3701	10.7036

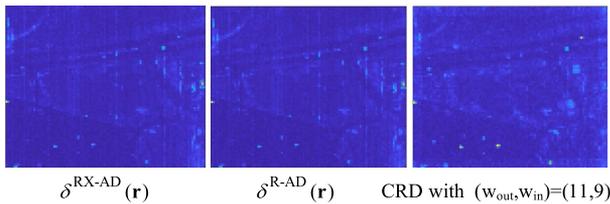


Fig. 13. Detection maps of RX/R-AD and CRD-DW-STO for HYDICE urban scene.

AUC_(D,F) values than $\delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{S_j}), \delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{L_m+\hat{S}_j})$ for $(p,m,j) = (9,5,4)$ but CRD-DW-STO performed much worse than $\delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{S_j}), \delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{L_m+\hat{S}_j})$ in all other categories of detection measures. So, if the AUC_(D,F) value was only detection measure to evaluate the anomaly detection performance, we would immediately conclude that CRD-DW-STO was a better anomaly detector than $\delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{S_j}), \delta_{L_m}^{OSPDS-AD}(\mathbf{r}_{L_m+\hat{S}_j})$. Apparently, this is not true according to Table IX. What is more, as also shown in Table IX (the

TABLE IX
COMPARATIVE PERFORMANCE ANALYSIS BETWEEN $\delta_{L_m}^{\text{OSPDS-AD}}(\mathbf{r}_{S_j})$ AND $\delta_{L_m}^{\text{OSPDS-AD}}(\mathbf{r}_{L_m+\hat{S}_j})$ AGAINST CRD-DW-STO FOR HYDICE 15-PANEL SCENE

	(p,m,j)	AUC _(D,F)	AUC _(D,ST)	AUC _(F,ST)	AUC _{ODP}	AUC _{TD}	AUC _{BS}	AUC _{TDBS}	AUC _{SNPR}
$\delta_{L_m}^{\text{OSPDS-AD}}(\mathbf{r}_{S_j})$	(9,5,4)	0.9892	0.6377	0.0248	1.6020	1.6268	0.9643	0.6128	25.6640
	(13,7,6)	0.9943	0.5948	0.0292	1.5599	1.5891	0.9651	0.5656	20.4027
	(61,35,26)	0.9855	0.3393	0.0436	1.2811	1.3248	0.9418	0.2956	7.7733
$\delta_{L_m}^{\text{OSPDS-AD}}(\mathbf{r}_{L_m+\hat{S}_j})$	(9,5,4)	0.9890	0.6196	0.0241	1.5845	1.6087	0.9649	0.5955	25.6959
	(13,7,6)	0.9944	0.5883	0.0285	1.5543	1.5827	0.9659	0.5599	20.6613
	(61,35,26)	0.9834	0.3353	0.0462	1.2725	1.3187	0.9372	0.2892	7.2631
CRD-DW-STO	$(W_{\text{outer}}, W_{\text{inner}}) = (11,9)$	0.9966	0.4469	0.0506	1.3928	1.4435	0.9460	0.3963	8.8320
RX-AD		0.9872	0.2575	0.0361	1.2087	1.2447	0.9511	0.2214	7.1330
R-AD		0.9872	0.2549	0.0359	1.2062	1.2421	0.9513	0.2190	7.1003

last two rows at the bottom), if RX/R-AD were implemented without using the LRaSMD model, their performances were far worse than OSPDS-AD in every category of detection measures. Nevertheless, their AUC_(D,F) value 0.9872 were very close to the best result 0.9966 within 10^{-2} . This indicated that RX/R-AD performed very closely to CRD-DW-STO. This was obviously incorrect. So, these experiments further demonstrated that the commonly used 2-D ROC curve as an evaluation criterion is not an effective detection indicator when detectors perform very closely. In this case, using the 3-D ROC-derived detection measures in Section VII is more appropriate and effective.

C. Discussions

The proposed OSP-AD offers a new application for OSP in OSP-AD. Technically speaking, OSP and RX/R-AD are designed from two completely different rationales since the former requires prior target knowledge, \mathbf{d} and \mathbf{U} in (10), while the latter does not. So, applying OSP to AD seems infeasible. This article shows otherwise. There are two key ideas of making OSP capable of performing AD. One is to take advantage of the sparse matrix \mathbf{S} and the low-rank matrix \mathbf{L} generated by OSP-GoDec. A second one is to appeal for DS to remove data statistics of first and second orders so that BKG can be suppressed or annihilated, while anomalies can be enhanced by retaining its high-order statistics via DS. To make it work, the prior target knowledge \mathbf{d} and \mathbf{U} used in (10) are replaced by \mathbf{A} and \mathbf{B} in (35), both of which can be specified by \mathbf{L} and \mathbf{S} . In particular, the $P_{\mathbf{U}}^{\perp}$ used in (10) is replaced by BKG annihilator specified by $P_{\mathbf{B}}^{\perp}$ in (36), while the \mathbf{d} used in (10) is replaced with $\mathbf{r}_{\hat{\mathbf{A}}}$ in (36).

On the other hand, OSP-AD is also quite different from RX/R-AD developed in [49]. There are two significant differences between $\delta_{\mathbf{B}}^{\text{RX/R-AD}}(\mathbf{r}_{\mathbf{A}})$ in (42) and (43) developed in [49] and OSP-AD in (35) and OSPDS-AD in (36) derived in this article. The first one is that OSP-AD performs $P_{\mathbf{B}}^{\perp}$ in (35) rather than $\mathbf{K}_{\mathbf{B}}^{-1}/\mathbf{R}_{\mathbf{B}}^{-1}$ performed by RX/R-AD in (42) and (43) respectively. As a consequence, OSP-AD annihilates BKG specified by \mathbf{B} via $P_{\mathbf{B}}^{\perp}$ compared to RX/R-AD which use $\mathbf{K}_{\mathbf{B}}^{-1}/\mathbf{R}_{\mathbf{B}}^{-1}$ to suppress BKG specified by \mathbf{B} . A second one is that OSP-AD makes use of DS to further clean up first- and second-order statistics from the data subspaces specified by the LRaSMD model. This DS process in (36) is not included in (42) and (43) but is very effective for OSPDS-AD to extract anomalies from the sphered target subspace specified by $\hat{\mathbf{A}}$.

X. CONCLUSION

OSP is a versatile hyperspectral imaging technique which has been used in many applications in hyperspectral data exploitation. It makes use of \mathbf{d} and \mathbf{U} in (10) to extract the desired targets specified by \mathbf{d} , while annihilating undesired targets specified by \mathbf{U} . Therefore, the knowledge of both \mathbf{d} and \mathbf{U} must be provided *a priori* before implementing OSP-TD. Unfortunately, finding the precise knowledge of \mathbf{U} is nearly impossible in practical applications. As a consequence, OSP has never been considered as an anomaly detector. Interestingly, Fig. 1 provides a clue that essentially leads to feasibility of making OSP an anomaly detector. However, a direct use of OSP is apparently not applicable. This article introduces the LRaSMD model and DS into OSPDS-AD in (36) to make it work effectively for anomaly detection.

It should be also noted that there are at least three significant differences between $\mathbf{r}_{\hat{\mathbf{A}}}^T P_{\mathbf{B}}^{\perp} \mathbf{r}_{\mathbf{A}}$ developed for OSP-AD in (35) and (36) and $(\mathbf{r}_{\mathbf{A}} - \mu_{\mathbf{B}})^T \mathbf{K}_{\mathbf{B}}^{-1} (\mathbf{r}_{\mathbf{A}} - \mu_{\mathbf{B}}) / \mathbf{r}_{\hat{\mathbf{A}}}^T \mathbf{R}_{\mathbf{B}}^{-1} \mathbf{r}_{\mathbf{A}}$ developed RX/R-AD in [49]. First of all, OSP-AD uses a target detector to perform anomaly detection compared many existing anomaly detectors using RX/R-AD types of detectors. Second, OSP-AD “*annihilates*” BKG specified by \mathbf{B} via $P_{\mathbf{B}}^{\perp}$ compared to $\mathbf{r}_{\hat{\mathbf{A}}}^T \mathbf{R}_{\mathbf{B}}^{-1} \mathbf{r}_{\mathbf{A}}$ which “*suppresses*” BKG specified by \mathbf{B} via $\mathbf{R}_{\mathbf{B}}^{-1}$. Third, OSPDS-AD in (36) makes use of DS as opposed to OSP-AD in (35) without using DS.

Several major contributions are summarized as follows.

- 1) OSP-AD provides an alternative to commonly used the sample auto-covariance matrix-based AD including RX-AD and its variants including the sample auto-correlation matrix-based R-AD. Most works reported in the AD literature still use RX/R-AD-type detectors to find anomalies, while OSP-AD utilizes OSP-TD in conjunction with DS to perform AD. It is believed that OSP-AD is the first work which implements OSP-TD as an anomaly detector without appealing for RX/R-AD.
- 2) OSP-AD takes advantage of the low-rank matrix \mathbf{L} and sparse matrix \mathbf{S} generated by OSP-GoDec to design various versions of OSP-AD.
- 3) OSP-AD separates the original data space \mathbf{X} into BKG and target subspaces. Under such circumstance, the undesired target space \mathbf{U} is represented by BKG subspace specified by $\mathbf{B} = \mathbf{L}$ or $\mathbf{L}+\mathbf{S}$ to remove BKG effects on AD via $P_{\mathbf{B}}^{\perp}$ and in the mean time, the data samples are operated

in the target subspace \mathbf{A} specified by either \mathbf{S} or $\mathbf{L}+\mathbf{S}$ to detect anomalies simultaneously. As a result, the prior target knowledge required by OSP is no longer necessary.

- 4) Most importantly, OSPDS-AD in (36) includes DS to remove BKG interference from the target subspace constructed from the OSP-GoDec generated sparse matrix \mathbf{S} for further improving AD.

REFERENCES

- [1] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, NY, USA: Kluwer/Plenum, 2003.
- [2] C.-I Chang and M. Hsueh, "Characterization of anomaly detection for hyperspectral imagery," *Sensor Rev.*, vol. 26, no. 2, pp. 137–146, 2006.
- [3] C.-I Chang, *Real-Time Progressive Hyperspectral Image Processing: End-member Finding and Anomaly Detection*. New York, NY, USA: Springer, 2016.
- [4] 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.
- [5] 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.
- [6] J. M. Moleró, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.
- [7] S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "A locally adaptive background density estimator: An evolution for RX-based anomaly detectors," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 323–327, Jan. 2014.
- [8] C.-I Chang, Y. Wang, and S. Y. Chen, "Anomaly detection using causal sliding windows," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3260–3270, Jul. 2015.
- [9] H. Kwon, S. Z. Der, and N. M. Nasrabadi, "Adaptive anomaly detection using subspace separation for hyperspectral imagery," *Opt. Eng.*, vol. 42, no. 11, pp. 3342–3351, Nov. 2003.
- [10] L. Wei and D. Qian, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [11] W. Liu and C.-I Chang, "Multiple window anomaly detection for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 664–658, Apr. 2013.
- [12] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [13] Y. Wang *et al.*, "A posteriori hyperspectral anomaly detection for unlabeled classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 3091–3106, Jun. 2018.
- [14] S. Y. Chen, Y. Wang, C. C. Wu, C. Liu, and C.-I Chang, "Real time causal processing of anomaly detection in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 1511–1534, Apr. 2014.
- [15] W. Xie, T. Jiang, Y. Li, X. Jia, and J. Lei, "Structure tensor and guided filtering-based algorithm for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4218–4230, Jul. 2019.
- [16] J. Lei, W. Xie, J. Yang, Y. Li, and C.-I Chang, "Spectral-spatial feature extraction for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8131–8143, Oct. 2019.
- [17] S. Chang, B. Du, and L. Zhang, "BASO: A background-anomaly component projection and separation optimized filter for anomaly detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3747–3761, Jul. 2018.
- [18] R. Zhao, B. Du, and L. Zhang, "Hyperspectral anomaly detection via a sparsity score estimation framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3208–3222, Jun. 2017.
- [19] 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.
- [20] 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, 2018, Art. no. 417. [Online]. Available: <https://doi.org/10.3390/rs10030417>
- [21] 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.
- [22] 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, May 2014, Art. no. 083641.
- [23] L. Zhu, G. Wen, and S. Qiu, "Low-rank and sparse matrix decomposition with cluster weighting for hyperspectral anomaly detection," *Remote Sens.*, vol. 10, no. 5, May 2018, Art. no. 707.
- [24] Y. Yang, J. Zhang, S. Song, C. Zhang, and D. Liu, "Low-rank and sparse matrix decomposition with orthogonal subspace projection-based background suppression for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1378–1382, Aug. 2020.
- [25] S. Y. Chen, S. Yang, K. Kalpakis, and C. I. Chang, "Low-rank decomposition-based anomaly detection," *Proc. SPIE*, vol. 8743, 2013, Art. no. 87430N. [Online]. Available: <https://doi.org/10.1117/12.2015652>
- [26] 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.*, 2019. [Online]. Available: <https://doi.org/10.3390/rs11131578>.
- [27] X. Ma, X. Zhang, X. Tang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection based on low-rank representation with data-driven projection and dictionary construction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 5, pp. 2226–2239, May 2020.
- [28] A. Borghesi, A. Bartolini, M. Lombardi, M. Milano, and L. Benini, "Anomaly detection using autoencoders in high performance computing systems," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 9428–9433, 2019.
- [29] W. Xie, J. Lei, B. Liu, Y. Li, and X. Jia, "Spectral constraint adversarial autoencoders approach to feature representation in hyperspectral anomaly detection," *Neural Netw.*, vol. 119, pp. 222–234, Nov. 2019.
- [30] S. Wang, X. Wang, Y. Zhong, and L. Zhang, "Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3057721.
- [31] W. Xie, B. Liu, Y. Li, J. Lei, C.-I Chang, and G. He, "Spectral adversarial feature learning for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2352–2365, Apr. 2020.
- [32] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4666–4679, Jul. 2020.
- [33] Z. Huang, X. Kang, S. Li, and Q. Hao, "Game theory-based hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2965–2976, Apr. 2020.
- [34] 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.
- [35] C.-I Chang, "Orthogonal subspace projection revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.
- [36] Q. Du and C.-I Chang, "A signal-decomposed and interference-annihilated approach to hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 892–906, Apr. 2004.
- [37] C.-I Chang, "Further results on relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 1030–1032, May 1998.
- [38] C.-I Chang, X. Zhao, M. L. G. Althouse, and J.-J. Pan, "Least squares subspace projection approach to mixed pixel classification in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.
- [39] C.-I Chang, C. Wu, W. Liu, and Y. C. Ouyang, "A growing method for simplex-based endmember extraction algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2804–2819, Oct. 2006.
- [40] C.-I Chang, S. Y. Chen, H. C. Li, and C.-H. Wen, "A comparative analysis among ATGP, VCA and SGA for finding endmembers in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4280–4306, Sep. 2016.
- [41] H.-C. Li and C.-I Chang, "Recursive orthogonal projection-based simplex growing algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 3780–3793, Jul. 2016.

- [42] 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.
- [43] J. C. Harsanyi, *Detection and Classification of Subpixel Spectral Signatures in Hyperspectral Image Sequences*. Ph.D. dissertation, Dept. Elect. Eng., Univ. Maryland, Baltimore, MD, USA, 1993.
- [44] V. Cevher, A. Sankaranarayanan, M. Duarte, D. Reddy, R. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 155–168.
- [45] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, pp. 217–288, 2011.
- [46] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 1027–1063, 2009.
- [47] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM J. Optim.*, vol. 21, pp. 572–596, 2011.
- [48] 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.
- [49] C.-I Chang, H. Cao, S. Chen, X. Shang, M. Song, and C. Yu, "Orthogonal subspace projection-based GoDec for low rank and sparsity matrix decomposition for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2403–2429, March, 2021.
- [50] C.-I Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic analysis," *IEEE Trans. Geosci. Remote Sens.*, to be published.
- [51] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- [52] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho, through the use of a constrained energy minimization technique," *Remote Sens. Environ.*, vol. 59, pp. 64–76, 1997.
- [53] C.-I Chang, "Target signature-constrained mixed pixel classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 5, pp. 1065–1081, May 2002.
- [54] H. Ren and C.-I Chang, "Target-constrained interference-minimized approach to subpixel target detection for hyperspectral imagery," *Opt. Eng.*, vol. 39, no. 12, pp. 3138–3145, Dec. 2000.
- [55] Q. Du, H. Ren, and C.-I Chang, "A comparative study for orthogonal subspace projection and constrained energy minimization" *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1525–1529, Jun. 2003.
- [56] C.-I Chang and Q. Du, "Estimation of number of spectrally distinct spectral signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
- [57] C.-I Chang, "Virtual dimensionality for hyperspectral imagery," SPIE Newsroom, Sep. 2009. [Online]. Available: <http://newsroom.spie.org/x37123.xml>
- [58] 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.
- [59] C.-I Chang, *Real-Time Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation*. New York, NY, USA: Springer, 2017.
- [60] 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.
- [61] C.-I Chang, W. Xiong, H. M. Chen, and J. W. Chai, "Maximum orthogonal subspace projection to estimating number of spectral signal sources for hyperspectral images," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 504–520, Jun. 2011.
- [62] 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.
- [63] H. Ren, Q. Du, J. Wang, C.-I Chang, and J. Jensen, "Automatic target recognition hyperspectral imagery using high order statistics," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 4, pp. 1372–1385, Oct. 2006.
- [64] M. Song, X. Shang, and C.-I Chang, "3D receiver operating characteristic analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8093–8115, Nov. 2020.



Chein-I Chang (Life Fellow, IEEE) received the B.S. degree in mathematics from Soochow University, Suzhou, China, 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. in mathematics and M.S.E.E. in electrical engineering degrees from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 1982, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 1987.

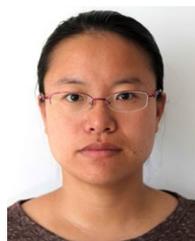
He has been with the University of Maryland, Baltimore County (UMBC), since 1987 and is currently a Professor with the Department of Computer Science and Electrical Engineering. He authored four books, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Norwell, MA, USA: Kluwer, 2003), *Hyperspectral Data Processing: Algorithm Design and Analysis* (Hoboken, NJ, USA: Wiley, 2013), *Real Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection* (New York, NY, USA: Springer, 2016), and *Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation* (New York, NY, USA: Springer, 2017). In addition, he also edited two books, *Recent Advances in Hyperspectral Signal and Image Processing, 2006 and Hyperspectral Data Exploitation: Theory and Applications* (Hoboken, NJ, USA: Wiley, 2007) and coedited the book, *High Performance Computing in Remote Sensing* (Boca Raton, FL, USA: CRC Press, 2007), with A. Plaza. He has several patents on hyperspectral image processing. His research interests include multispectral/hyperspectral image processing, automatic target recognition, and medical imaging.

Dr. Chang is a Fellow of SPIE. He is currently holding Chang Jiang Scholar Chair Professorship and the Director of Center for Hyperspectral Imaging in Remote Sensing (CHIRS) with Dalian Maritime University, Dalian, China, since 2016. In addition, he has been a Chair Professor with National Chiao Tung University, Hsinchu, Taiwan, since 2019. He was the recipient of 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.



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 currently working toward the Ph.D. degree in technology of computer applications with Dalian Maritime University, Dalian, China.

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



Meiping Song received the Ph.D. degree in computer applied technology 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 has been a Professor with the College of Information Science and Technology, Dalian Maritime University, Dalian, China, since 2020. Her research interests include remote sensing and hyperspectral image processing.