

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

C. -I. Chang, "Target-to-Anomaly Conversion for Hyperspectral Anomaly Detection," in *IEEE Transactions on Geoscience and Remote Sensing*, 2022, doi: 10.1109/TGRS.2022.3211696

<https://doi.org/10.1109/TGRS.2022.3211696>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Target-to-Anomaly Conversion for Hyperspectral Anomaly Detection

Chein-I Chang^{ID}, *Life Fellow, IEEE*

Abstract—A known target detection assumes that the target to be detected is provided a priori, while anomaly detection is an unknown target detection without any prior knowledge. As a result, known target detection generally performs search-before-detect detection in an active mode, referred to as active target detection as opposed to anomaly detection, which performs throw-before-detect detection in a passive mode, referred to as passive target detection. Accordingly, techniques designed for these two types of detection are completely different. This article shows that there is indeed a mechanism, called target-to-anomaly conversion, which can convert hyperspectral target detection (HTD) to hyperspectral anomaly detection (HAD) via a novel idea, called dummy variable trick (DVT). By virtue of such target-to-anomaly conversion many well-known target detection techniques, such as likelihood ratio test (LRT), constrained energy minimization (CEM), and orthogonal subspace projection (OSP), the spectral angle mapper (SAM) and the adaptive cosine estimator (ACE) can be converted to their corresponding anomaly detectors, referred to as target-to-anomaly conversion-derived anomaly detectors (TAC-ADs). Since a target detector requires target knowledge while TAC-AD does not, a direct use of TAC-AD is not effective. To make TAC-AD work, a newly developed approach to effective anomaly space (EAS) is implemented in conjunction with TAC-AD so that anomalies can be retained in EAS and interference, and noise including background (BKG) can be removed from EAS. The experiments demonstrate that TAC-AD operating in EAS performs better than many existing anomaly detection approaches, including model-based methods.

Index Terms—Anomaly detector (AD), constrained energy minimization (CEM), dummy variable trick (DVT), effective anomaly space (EAS), hyperspectral anomaly detection (HAD), hyperspectral target detection (HTD), target-to-anomaly conversion-derived anomaly detector (TAC-AD).

NOMENCLATURE

| | |
|------|-----------------------------------|
| ACE | Adaptive cosine estimator. |
| AD | Anomaly detector. |
| ADBS | Anomaly detection in background. |
| ADP | Anomaly detection probability. |
| BDP | Background detection probability. |
| BKG | Background. |
| BS | Background suppressibility. |
| CDA | Component decomposition analysis. |

| | |
|----------|--|
| CEM | Constrained energy minimization. |
| CRD | Collaborative representative detector. |
| DL | Deep learning. |
| DS | Data sphering. |
| DVT | Dummy variable trick. |
| EAS | Effective anomaly space. |
| GLRT | Generalized likelihood ratio test. |
| GoDec | Go decomposition. |
| GSNR | Generalized signal-to-noise ratio. |
| IC | Independent component. |
| ICA | Independent component analysis. |
| LRR | Low-rank representation. |
| LRT | Likelihood ratio test. |
| LRaSMD | Low-rank and sparse matrix decomposition. |
| LSDM-MoG | Low-rank and sparse decomposition model with a mixture of Gaussians. |
| MD | Mahalanobis distance. |
| MX-SVD | Minimax singular value decomposition. |
| OADP | Overall anomaly detection probability. |
| OSP | Orthogonal subspace projection. |
| P_D | Detection probability. |
| P_F | False alarm probability. |
| PTA | Prior-based tensor approximation. |
| RGAE | Robust graph autoencoder. |
| ROC | Receiver operating characteristic. |
| SA | Spectral angle. |
| SAM | Spectral angle mapper. |
| SBR | Signal-to-background ratio. |
| SC | Sparsity cardinality. |
| SNR | Signal-to-noise ratio. |
| SNPR | Signal-to-noise probability ratio. |
| TAC-AD | Target-to-anomaly conversion-derived anomaly detector. |
| TD | Target detectability. |
| UN | Unnormalized. |
| UST | Unknown specified target. |
| VD | Virtual dimensionality. |

Manuscript received 29 July 2022; revised 11 September 2022; accepted 28 September 2022. Date of publication 3 October 2022; date of current version 24 October 2022.

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

Digital Object Identifier 10.1109/TGRS.2022.3211696

1558-0644 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

I. INTRODUCTION

TARGET detection is a fundamental task in statistical signal processing, which is generally formulated as a binary hypothesis testing problem with null and alternative hypotheses, H_0 and H_1 used to represent the absence and the presence of a signal, respectively, where the signal to

be detected is assumed to be known [1]. Such a known signal detection in the additive noise model has been widely used in communications so that the optimal detector can be obtained by LRT. In communications, anomaly detection is not of major concern because anomalies are supposed to be unknown. Without knowing signals, we do not know what is supposed to be transmitted for communication. By contrast, since hyperspectral imaging (HSI) is capable of uncovering and revealing unknown subtle material substances, which cannot be identified by prior knowledge or visual inspection, finding these targets is of particular interest to HSI. However, it is apparent that a direct application of the binary hypothesis testing theory to the detection of such unknown objects is not working. To resolve this dilemma, the first obstacle is to remove the known target requirement. However, complete removal of the known target is not possible. As a compromise, it assumes that the target to be detected has a known shape with an unknown amplitude. Under this circumstance, binary hypothesis testing can be still applied, but its simple hypotheses must be replaced by composite hypotheses with the signal amplitude specified by unknown parameters. As a second requirement, a probability distribution under each hypothesis must be known to calculate LRT, so a secondary Gaussian dataset is assumed to be available, which can be used to estimate the signal parameters to yield Gaussian maximum likelihood estimation. Such resulting LRT with parameters to be estimated is called GLRT. By virtue of GLRT, an AD was derived by Reed and Yu [2], referred to as RX-AD with more details in [3], [4], [5], and [6].

In the early development of hyperspectral anomaly detection (HAD), the GLRT-based RX-AD was one of the most widely used methods and has been extended to many variants, such as local RX-AD [7], [8], [9], support vector data description (SVDD) [8], weighted-RX-AD and linear filter-based RX-AD in [10], a sliding window [11], a dual-window-based eigen separation transform (DWEST) that was developed in [12] followed by collaborative representation-based AD (CRD), which also used dual windows in [13], and multiple-window AD (MWAD) in [14] by making use of multiple windows to perform anomaly detection adaptively, guided filtering-based AD [15], spectral-spatial feature extraction-based AD [16], and a nonlinear AD, kernel-based AD (KRX-AD) [17]. Specifically, in a recent comprehensive study [18], Chang derived many known target detectors from three perspectives, hypothesis testing, SNR, and SA theories, including well-known target detectors, adaptive matched filter (AMF) in [19], ACE in [20], and adaptive subspace detector (ASD) in [21].

Since target detection is a well-established theory in communications and signal processing, the motivation of this article is “can we take advantage of what we have done for target detection in [18] and transplant it to anomaly detection?” In signal communications, we generally need to know what signals are being used for data transmission. Thus, anomaly detection has little interest in statistical communications. However, this is completely opposite to HSI, which considers anomaly detection as one of its major tasks, because of its ability in extracting unknown crucial and vital substances for

image analysis. The difference between target detection and anomaly detection has been discussed in great detail in [22, Ch. 5] where the former is considered active target detection due to its required prior knowledge of targets to be detected compared to the latter, which is passive target detection with no target knowledge required. This taxonomy was further studied in [23], which shows that anomaly detection is actually a dual theory of target detection. However, there are no discussions on how to make a target detector an AD in [23]. This is easily said than done.

An earlier attempt was made in [24], which introduced a novel idea, called DVT to convert a well-known subtarget detector, CEM [25], [26], [27] to a CEM-based AD (CEM-AD). As elegant as DVT is, this article further takes advantage of DVT to develop an approach, called target-to-anomaly conversion-derived AD (TAC-AD), which converts many target detectors in [23] to their corresponding ADs, particularly, OSP-based target detectors [27], [28], SAM in [27], and ACE in [20].

One serious issue arising in TAC-AD is BKG suppressibility (BS), which is generally not of major concern in target detection. This is because target detection is performed in an active mode by assuming targets of interest to be known. As a consequence, its poor BS can be offset and compensated by enhancing the detection of known targets via a matched filter. Thus, such active target detection can be generally considered as a search-before-detect technique. Unfortunately, with the unavailability of targets to be detected, the concept of “search-before-detect” is not applicable because we do not know what targets are needed to be searched. Under this circumstance, we must rely on BS to bring unknown interesting targets, such as anomalies to light. This leads to passive target detection. One effective means for passive target detection is the so-called “throw-before-detect” technique, which separates, estimates, or reconstructs BKG for subsequent removal or suppression before detection. In other words, the concept of “throw-before-detect” can be regarded as a technique, which throws away uninteresting or unwanted targets, such as BKG prior to anomaly detection. In recent years, this approach has been a major trend in anomaly detection. Of particular interest is the LRaSMD model that has been shown in [29], [30], [31], [32], [33], and [34] to take care of BS and noise effectively.

In addition to BKG, there is also a severe issue in anomaly detection. That is, anomalies are also contaminated by noise, particularly, non-Gaussian noise, even though the second order of statistics Gaussian noise can be removed by DS. As a result, anomalies are indeed sandwiched between BKG and noise. To cope with these two issues, Chang et al. [29] investigated the LRaSMD model generated by GoDec proposed in [30], which decomposes a data space \mathbf{X} into a linear mixture model of three component subspaces, \mathbf{L} , \mathbf{S} , and \mathbf{N} as $\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$, where \mathbf{L} is a low-rank subspace to represent BKG, \mathbf{S} is a sparse subspace to represent anomaly space, and \mathbf{N} is a noise subspace. By taking advantage of the LRaSMD model, Chang et al. [31] further developed an approach to implement an AD, such as RX-AD in two different subspaces by suppressing BKG via \mathbf{L} , while extracting anomalies from \mathbf{S} via a matched filter. The same approach was further explored to derive

an OSP-based AD (OSP-AD) in [32]. Recently, a similar approach was developed by Chen et al. [33], which replaces \mathbf{L} and \mathbf{S} in the LRaSMD-model with principal component (PC) and IC subspaces, respectively, to derive a new CDA. All these approaches culminate in a new concept of EAS, which was coined in [34] to effectively address both BKG and noise issues simultaneously altogether. By operating TAC-AD in EAS, the BKG and noise issues can be effectively resolved so that its detection performance can be significantly improved.

Another important difference between target detection and AD is that the standard ROC curve widely used to evaluate target detection cannot be effectively to evaluate BS, as shown in [35]. An ROC curve is a plot of P_D versus P_F , and its area under the ROC curve (AUC), denoted by $AUC_{(D,F)}$, is generally calculated as a quantitative measure to assess the effectiveness of a detector. Since both P_D and P_F are calculated by the same threshold τ used by the detector, when both P_D and P_F are very high, its calculated $AUC_{(D,F)}$ values are also very high. Similarly, when both P_D and P_F are very low, its calculated AUCs are also very low. In other words, P_D and P_F are tied together simultaneously and cannot work independently. Consequently, P_D and P_F cannot individually measure TD and BS, respectively. To address this issue, Chang [35] developed an effective 3D ROC analysis-based evaluation tool, which extends the traditional 2D ROC analysis by including the threshold τ as an additional parameter to represent a 3D ROC curve as a function of three parameters, P_D , P_F , and τ . As a result, a 3D ROC curve can be generated by a three-parameter vector specified by $(P_D, \text{ev}P_F, \tau)$ from which three 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) can be also generated along with their respective AUC values, denoted by $AUC_{(D,F)}$, $AUC_{(D,\tau)}$, and $AUC_{(F,\tau)}$, where $AUC_{(D,\tau)}$ and $AUC_{(F,\tau)}$ can be used to evaluate TD and BS, respectively.

To demonstrate the performance of TAC-AD and the utility of 3D ROC analysis, experiments are conducted based on ADs converted from several well-known target detectors, LRT, GLRT, OSP, CEM, and two SA-based matched filters, SAM and ACE derived in [18], and their experimental results and discussions are also provided for comparative analysis.

The main contributions of this article are summarized in the following.

- 1) Since a target detector requires prior target knowledge, while an AD does not, direct conversion is infeasible. Thus, the first and foremost contribution is the novelty of developing TAC-AD, which uses the novel idea proposed by Chang [24], and DVT to connect the conversion between with/without prior knowledge via an introduction of a UST as an auxiliary variable.
- 2) The three types of target detectors derived in [18], LRT-, SNR-, and SA-based target detectors, can be converted to their corresponding TAC-ADs. Interestingly, it also shows that most matched filter-based target detectors, such as ASD, ACE, and AMF, end up with the same AD. This is mainly because the known target signals \mathbf{d} used by their matched filters have been all reset by the same unknown specified target, \mathbf{d}^{UST} without normalization. Specifically, RX-AD developed in [2] is actually a TAC-

AD, which can be derived from DVT with \mathbf{d}^{UST} used as a dummy variable to be maximized by GLRT.

- 3) Four new TAC-ADs, (RX-AD)², (CEM-AD)², SAM-based AD, and ACE-AD = (SAM-AD)², are derived, which have not been explored in the literature. As will be shown in the conducted experiments, these TAC-ADs work very well in EAS and indeed perform better than many current existing ADs.
- 4) One key difference between target detection and anomaly detection is BKG and noise issues. Since target detection assumes that the used target knowledge is known, it can be performed by standard signal detection in noise where BKG is not of main concern and noise is generally assumed to be Gaussian. This is not true for anomaly detection because target knowledge is unknown and noise is also non-Gaussian. As expected, such TAC-AD will generally not perform effectively. To address these issues, the concept of EAS derived in [34] provides a solution that can be implemented in conjunction with TAC-AD operating EAS to improve its performance in anomaly detection.
- 5) One of the most important contributions of this article is the extensive and comprehensive experiments performed along with discussions in Section VII, which shows that TAC-AD working in the EAS space indeed performs better than many existing anomaly detection methods.

The remainder of this article is organized as follows. Section II reviews three types of target detectors in [18], LRT, SNR, and OSP. Section III provides detailed mathematical derivations of TAC-AD from the target detector in Section II. Section IV discusses the crucial role that EAS plays in TAC-AD. Section V presents 3D ROC-derived detection measures for anomaly detection. Section VI describes hyperspectral real images to be used for experiments. Section VII performs comprehensive experiments and conducted experimental studies and discussions. Section VIII concludes the main results of this article related to previous works.

II. TARGET DETECTION

There are many target detectors have been designed and developed for HSI. Of particular interest are three types of pixel-based target detectors: binary hypothesis testing-derived LRT-based pure pixel target detectors, OSP-based mixed-pixel target detector [27], [28], and CEM-based subpixel target detector [25], [26], [27].

A. Hypothesis Testing Theory-Derived Pure Pixel-Based Target Detectors

A standard target detection problem for hyperspectral target detection (HTD) can be formulated by a binary simple hypothesis testing problem as follows:

$$\begin{aligned} H_0 : \mathbf{r} &= \mathbf{n} \\ \text{versus} \\ H_1 : \mathbf{r} &= \mathbf{d} + \mathbf{n} \end{aligned} \quad (1)$$

where \mathbf{r} and \mathbf{d} are L -dimensional vectors with L being the number of spectral bands and \mathbf{n} is noise. In (1), the

target signal \mathbf{d} is assumed to be known and deterministic. This type of target detection is referred to as a priori target detection.

The hypotheses H_0 and H_1 are called the “null hypothesis” that represents the signal absence and the “alternative hypothesis” that represents the signal presence, respectively, each of which is governed by their probability density functions, $p_0(\mathbf{r})$ and $p_1(\mathbf{r})$. Assume that the prior probabilities of H_0 and H_1 are given by $P(H_0)$ and $P(H_1)$, respectively. Also, let c_{ij} be the cost incurred by a decision saying H_i when H_j is true.

The solution to (1) can be found in [1]. It is the Bayes detector $\delta^{\text{Bayes}}(\mathbf{r})$, which is an LRT, $\Lambda(\mathbf{r})$ given by

$$\delta^{\text{LRT}}(\mathbf{r}) = \begin{cases} 1, & \text{if } \Lambda(\mathbf{r}) \geq \tau \\ 0, & \text{if } \Lambda(\mathbf{r}) < \tau \end{cases} \quad (2)$$

where $\Lambda(\mathbf{r}) = p_0(\mathbf{r})/p_1(\mathbf{r})$ and the threshold τ is given by $\tau = ((c_{10} - c_{00})P(H_0))/((c_{01} - c_{11})P(H_1))$. Using (3), P_D and P_F can be calculated by

$$P_D = \int_{\Lambda(\mathbf{r}) \geq \tau} p_1(\mathbf{r}) d\mathbf{r}, P_F = \int_{\Lambda(\mathbf{r}) \geq \tau} p_0(\mathbf{r}) d\mathbf{r}. \quad (3)$$

If $p_0(\mathbf{n})$ of \mathbf{n} in (2) is assumed to be Gaussian distributed with the mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{K} , then (3) can be derived by

$$\begin{aligned} \log \Lambda(\mathbf{r}) &= -(1/2)[(\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu}) \\ &\quad - (\mathbf{r} - \boldsymbol{\mu} - \mathbf{d})^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu} - \mathbf{d})] \\ &= \mathbf{d}^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu}) \geq \log \tau + (1/2) \mathbf{d}^T \mathbf{K}^{-1} \mathbf{d} = \tau'. \end{aligned} \quad (4)$$

Then, (2) becomes

$$\delta_1^{\text{LRT}}(\mathbf{r}) = \begin{cases} 1, & \text{if } \mathbf{d}^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu}) \geq \tau' \\ 0, & \text{if } \mathbf{d}^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu}) < \tau'. \end{cases} \quad (5)$$

Alternatively, we can also further simply (4) as follows:

$$\begin{aligned} \log \Lambda(\mathbf{r}) &= -(1/2)[(\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu}) \\ &\quad - (\mathbf{r} - \boldsymbol{\mu} - \mathbf{d})^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu} - \mathbf{d})] \\ &= \mathbf{d}^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu}) - (1/2) \mathbf{d}^T \mathbf{K}^{-1} \mathbf{d} \geq \log \tau \\ &= \mathbf{d}^T \mathbf{K}^{-1} \mathbf{r} \geq \mathbf{d}^T \mathbf{K}^{-1} \boldsymbol{\mu} + (1/2) \mathbf{d}^T \mathbf{K}^{-1} \mathbf{d} \\ &\quad + \log \tau = \tau''. \end{aligned} \quad (6)$$

Then, (2) also becomes

$$\delta_2^{\text{LRT}}(\mathbf{r}) = \begin{cases} 1, & \text{if } \mathbf{d}^T \mathbf{K}^{-1} \mathbf{r} \geq \tau'' \\ 0, & \text{if } \mathbf{d}^T \mathbf{K}^{-1} \mathbf{r} < \tau''. \end{cases} \quad (7)$$

The detectors $\delta_1^{\text{LRT}}(\mathbf{r})$ in (5) and $\delta_2^{\text{LRT}}(\mathbf{r})$ in (7) are basically the same detectors using different thresholds, τ' and τ'' , respectively. Both use a linear matched filter which uses the known target signal \mathbf{d} as the matching signal.

It should be noted that, according to [1, pp. 56 and 57], SNR can be also used as a criterion for (1) to derive $\mathbf{d}^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu})$ in (5) and $\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r}$ in (7) with

$$\text{SNR}^{\text{LRT}} = \text{var}[\delta^{\text{LRT}}(\mathbf{r})|H_j] = \mathbf{d}^T \mathbf{K}^{-1} \mathbf{d} \quad (8)$$

which is independent of hypotheses.

Since τ' in (5) and τ'' in (7) are determined by the prior cost c_{ij} and the prior knowledge of probabilities, $P(H_0)$ and $P(H_1)$, and $p_0(\mathbf{r})$ and $p_1(\mathbf{r})$, they are not generally available in practical applications. To resolve this issue, the Neyman–Pearson (NP) detection theory is widely used to replace τ and τ' with a constraint β imposed on P_F in (5) or (7) and then maximize the P_D in (5) or (7) as follows:

$$\max_{\delta} \{P_D(\delta(\mathbf{r}))\}, \quad \text{s.t. } P_F(\delta(\mathbf{r})) \leq \beta. \quad (9)$$

The resulting detector is called an NP detector, which is shown in [1] to be the same LRT given by

$$\delta^{\text{NP}}(\mathbf{r}) = \begin{cases} 1, & \text{if } \Lambda(\mathbf{r}) \geq \tau \\ 0, & \text{if } \Lambda(\mathbf{r}) < \tau \end{cases} \quad (10)$$

with τ constrained by $P_F(\delta(\mathbf{r})) \leq \beta$ in (8) where $\Lambda(\mathbf{r})$ can be either specified by $\mathbf{d}^T \mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu})$ in (5) or $\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r}$ in (7), and the threshold τ is specifically determined by

$$P_F(\delta(\mathbf{r})) = \int_{\Lambda(\delta(\mathbf{r})) \geq \tau} p_0(\mathbf{r}) d\mathbf{r} = \beta \Leftrightarrow \tau = P_F^{-1}(\beta). \quad (11)$$

The major advantage of using the NP detector in (10) over (5) or (7) is that no prior knowledge is required. Instead, the NP detector is determined by its detection level of significance β , which can be specified by real-world applications.

B. Mixed Pixel-Based OSP Target Detector

Now, we assume that there are the p target signatures, and $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_p$ present in the data and are used to unmix data samples via a linear mixture model given by

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n} \quad (12)$$

where \mathbf{r} , \mathbf{M} , $\boldsymbol{\alpha}$, and \mathbf{n} are the data sample vector, the spectral signature matrix, and the abundance fractional vector associated with \mathbf{r} and noise vector, respectively. The presence of \mathbf{n} in (12) can be interpreted as noise or measurement error or model error. Let $\mathbf{m}_p = \mathbf{d}$ be the desired target signature to be detected without loss of generality. Since $\mathbf{M}\boldsymbol{\alpha}$ is a mixed signal vector, and we are only interested in detecting the desired spectral signature \mathbf{d} , a direct standard detection technique cannot be applied to (12).

To resolve this dilemma, we divide the set of the p target signatures, $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_p$, into two classes of signatures. One class is made up of the single desired signature, $\mathbf{m}_p = \mathbf{d}$ with its proportional abundance fraction α_p , and the other class consists of the remaining signatures, $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{p-1}$, as undesired spectral signatures to form an undesired spectral signature matrix denoted by $\mathbf{U}_{p-1} = [\mathbf{m}_1 \mathbf{m}_2 \dots \mathbf{m}_{p-1}]$ with $\boldsymbol{\gamma} = (\alpha_1, \alpha_2, \dots, \alpha_{p-1})^T$ being its corresponding abundance fraction vector. By virtue of such p -signatures decomposition (i.e., desired signature specified by \mathbf{d} and undesired signatures specified by \mathbf{U}_{p-1}), we can further reexpress the mixed signal detection model in (11) as a $(\mathbf{d}, \mathbf{U}_{p-1})$ -signal decomposition detection model represented by

$$\mathbf{r} = \mathbf{d}\alpha_p + \mathbf{U}_{p-1}\boldsymbol{\gamma} + \mathbf{n}. \quad (13)$$

Once the undesired signatures $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{p-1}$ in \mathbf{U} are separated from the desired signature \mathbf{d} in (13), we can eliminate their interfering effects before the detection of \mathbf{d} takes

place, so as to enhance the detectability of \mathbf{d} . In doing so, we design an orthogonal subspace projector

$$P_{U_{p-1}}^\perp = \mathbf{I} - U_{p-1} U_{p-1}^\# \quad (14)$$

to annihilate \mathbf{U} from the pixel vector \mathbf{r} prior to detection of \mathbf{d} , where $U_{p-1}^\# = (U_{p-1}^T U_{p-1})^{-1} U_{p-1}^T$ is the pseudoinverse of \mathbf{U} . The notation $\frac{1}{U_{p-1}}$ in $P_{U_{p-1}}^\perp$ indicates that the projector $P_{U_{p-1}}^\perp$ maps the observed pixel vector \mathbf{r} into the orthogonal complement of $\langle U_{p-1} \rangle$, denoted by $\langle U_{p-1} \rangle^\perp$. Applying $P_{U_{p-1}}^\perp$ in (14) to (13) results in a $P_{U_{p-1}}^\perp$ -transformed data space, that is, for each data sample vector \mathbf{r}

$$P_{U_{p-1}}^\perp \mathbf{r} = P_{U_{p-1}}^\perp \mathbf{d} \alpha_p + P_{U_{p-1}}^\perp \mathbf{n} \quad (15)$$

where the original data sample vector \mathbf{r} and the noise \mathbf{n} become $P_{U_{p-1}}^\perp \mathbf{r}$ and $P_{U_{p-1}}^\perp \mathbf{n}$, and the effects of all the undesired signatures $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{p-1}$ on \mathbf{d} and \mathbf{n} have been eliminated by $P_{U_{p-1}}^\perp$. By further applying a matched filter using \mathbf{d} as the matching signal, we can obtain a mixed-pixel-based OSP detector, $\delta_{\mathbf{d}}^{\text{OSP}}(\mathbf{r})$, derived by Harsanyi and Chang [28]

$$\delta_{\mathbf{d}}^{\text{OSP}}(\mathbf{r}) = \mathbf{d}^T P_{U_{p-1}}^\perp \mathbf{r} = \hat{\alpha}_p(\mathbf{r}) = \hat{\alpha}_{\mathbf{d}}(\mathbf{r}) \quad (16)$$

where the subscript “ \mathbf{d} ” is included to emphasize the desired signature required by the OSP target detector. It should be noted that $\delta_{\mathbf{d}}^{\text{OSP}}(\mathbf{r})$ in (16) is a target detector and is not designed to estimate the abundance fraction of \mathbf{d} present in the data sample \mathbf{r} . To make $\delta_{\mathbf{d}}^{\text{OSP}}(\mathbf{r})$ an abundance estimator, the least squares OSP (LSOSP) was developed in [36] and derived as

$$\delta_{\mathbf{d}}^{\text{LSOSP}}(\mathbf{r}) = \frac{\mathbf{d}^T P_{U_{p-1}}^\perp \mathbf{r}}{\mathbf{d}^T P_{U_{p-1}}^\perp \mathbf{d}} = \hat{\alpha}_p^{\text{LSOSP}}(\mathbf{r}) = \hat{\alpha}_{\mathbf{d}}^{\text{LSOSP}}(\mathbf{r}) \quad (17)$$

which includes a constant $\mathbf{d}^T P_{U_{p-1}}^\perp \mathbf{d}$ to normalize $\delta_{\mathbf{d}}^{\text{OSP}}(\mathbf{r})$ in (16) to make the $\delta_{\mathbf{d}}^{\text{OSP}}(\mathbf{r})$ target detector an abundance estimator that was shown to be the same as the unconstrained least squares (ULS) linear unmixer in [37] and [38]. Most interestingly, it is further shown in [27] that this normalized constant turns out to be the maximum SNR of \mathbf{r}

$$\max_{\mathbf{r}} \text{SNR}(\mathbf{r}) = \mathbf{d}^T P_{U_{p-1}}^\perp \mathbf{d}. \quad (18)$$

Interestingly, since the OSP in (15) can be also derived from SNR [27], (16) should include a constant κ resulting from Schwarz’s inequality, i.e.,

$$\delta_{\mathbf{d}}^{\text{UN-OSP}}(\mathbf{r}) = \kappa \mathbf{d}^T P_{U_{p-1}}^\perp \mathbf{r} \quad (19)$$

which can be considered as the UN version of OSP (UN-OSP). Thus, if OSP in (19) is used to estimate the true abundance fraction, then $\delta_{\mathbf{d}}^{\text{OSP}}(\mathbf{d}) = 1$, which gives

$$\delta_{\mathbf{d}}^{\text{UN-OSP}}(\mathbf{d}) = \kappa \mathbf{d}^T P_{U_{p-1}}^\perp \mathbf{d} = \hat{\alpha}_p(\mathbf{d}) = 1 \Rightarrow \kappa = \left(\mathbf{d}^T P_{U_{p-1}}^\perp \mathbf{d} \right)^{-1}. \quad (20)$$

Substituting $\kappa = \left(\mathbf{d}^T P_{U_{p-1}}^\perp \mathbf{d} \right)^{-1}$ in (20) into (18) yields exactly the same as (17).

C. SBR-Derived Subpixel-Based CEM Target Detector

Suppose that a hyperspectral image is represented by a collection of image pixel vectors, denoted by $\{\mathbf{r}_i\}_{i=1}^N$ with $\mathbf{r}_i = (r_{i1}, r_{i2}, \dots, r_{iL})^T$ being an L -dimensional pixel vector and L being the total number of spectral bands. More specifically, let the original HSI sample space be Ω_{data} , where $\Omega = \{\mathbf{r}_n\}_{n=1}^N$ and $\mathbf{r}_n \in \mathbb{R}^{L \times 1}$ are an n th data sample with L -bands. Let $\delta(\mathbf{r}_n)$ be a linear filter specified by a weighting vector $\mathbf{w} = (w_1, w_2, \dots, w_L)^T$ as

$$\delta(\mathbf{r}_n) = \mathbf{w}^T \mathbf{r}_n. \quad (21)$$

Then, its averaged output energy over the entire image scene is represented by

$$\frac{1}{N} \sum_{n=1}^N \delta^2(\mathbf{r}_n) = \mathbf{w}^T \mathbf{R} \mathbf{w} \quad (22)$$

where $\mathbf{R} = (1/N) \sum_{n=1}^N \mathbf{r}_n \mathbf{r}_n^T$ is the sample autocorrelation matrix, N is the total number of data samples, and L is the number of spectral bands. Now, the following constrained optimization problem can be formulated as follows:

$$\min_{\mathbf{w}} \{ \mathbf{w}^T \mathbf{R} \mathbf{w} \}, \quad \text{s.t. } \mathbf{d}^T \mathbf{w} = 1. \quad (23)$$

The solution to (23) is given by

$$\mathbf{w}^{\text{CEM}} = \mathbf{R}^{-1} \mathbf{d} (\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d})^{-1}. \quad (24)$$

Substituting (23) into (20) also yields the CEM detector, $\delta^{\text{CEM}}(\mathbf{r})$ developed in [25], and given by

$$\delta_{\mathbf{d}}^{\text{CEM}}(\mathbf{r}) = (\mathbf{w}^{\text{CEM}})^T \mathbf{r} = \frac{\mathbf{d}^T \mathbf{R}^{-1} \mathbf{r}}{\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}} \quad (25)$$

where the subscript “ \mathbf{d} ” is included to emphasize the use of the desired target signature \mathbf{d} . The minimum variance resulting from (25) can be obtained as

$$(\mathbf{w}^{\text{CEM}})^T \mathbf{R} \mathbf{w}^{\text{CEM}} = (\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d})^{-1}. \quad (26)$$

Now, if we let Δ be all data samples specified by the target signature \mathbf{d} , then we can define a new correlation matrix $\mathbf{R}_{\Omega-\Delta}$ by removing all data samples in Δ from the entire data sample space as

$$\mathbf{R}_{\Omega-\Delta} = \frac{1}{|\Omega - \Delta|} \sum_{\mathbf{r}_n \in \Omega - \Delta} \mathbf{r}_n \mathbf{r}_n^T \in \mathbb{R}^{L \times L} \quad (27)$$

where $|\Omega - \Delta|$ is the size of $\Omega - \Delta$. In this case, $\mathbf{R}_{\Omega-\Delta}$ is formed by all data samples not specified by \mathbf{d} but considered as BKG samples

$$\begin{aligned} \text{SBR}(\delta_{\mathbf{d}}^{\text{CEM}}) &= \mathbf{d}^T \mathbf{R}_{\Omega-\Delta}^{-1} \mathbf{d} \\ &\Rightarrow (\mathbf{w}^{\text{CEM}})^T \mathbf{R}_{\Omega-\Delta}^{-1} \mathbf{w}^{\text{CEM}} = (\mathbf{d}^T \mathbf{R}_{\Omega-\Delta}^{-1} \mathbf{d})^{-1} \end{aligned} \quad (28)$$

can be interpreted as signal \mathbf{d} -to-BKG (Δ) ratio (SBR) of $\delta_{\mathbf{d}}^{\text{CEM}}(\mathbf{r})$. Replacing \mathbf{R}^{-1} in (25) with $\mathbf{R}_{\Omega-\Delta}^{-1}$ gives rise to

$$\delta_{\mathbf{d}}^{\text{SBR-CEM}}(\mathbf{r}) = \mathbf{d}^T \mathbf{R}_{\Omega-\Delta}^{-1} \mathbf{r} \approx \delta_{\mathbf{d}}^{\text{SBR-CEM}}(\mathbf{r}) = \mathbf{d}^T \mathbf{R}^{-1} \mathbf{r} \quad (29)$$

which implies that the minimum variance of (29) is indeed the same as the reciprocal of the SBR in (28). This is also confirmed by [1, eq. (50)], which is SNR derived by the R-SNR approach and also defined in [18].

There is an interesting finding between (25) and (17). In (25), $\delta_{\mathbf{d}}^{\text{CEM}}(\mathbf{r})$ includes $\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}$ as a normalization constant, which gives the maximum SBR of the data sample \mathbf{r} produced by CEM. Correspondingly, $\delta_{\mathbf{d}}^{\text{LSOSP}}(\mathbf{r})$ in (17) uses $\mathbf{d}^T P_{\mathbf{U}_{p-1}}^{\perp} \mathbf{d}$ as a normalization constant to produce the maximum SNR of $\delta_{\mathbf{d}}^{\text{OSP}}(\mathbf{r})$ in (18).

Since inverting the correlation matrix \mathbf{R} is simply equivalent to suppressing the interfering effects of data samples that form \mathbf{R} on detection of \mathbf{d} , technically speaking, CEM should use (27) instead of (23) to have better target detection performance. However, due to the fact that the size of Δ is generally very small, the difference between \mathbf{R} in (22) and $\mathbf{R}_{\Omega-\Delta}$ in (27) is not much appreciable. This is why CEM has been using \mathbf{R}^{-1} rather than $\mathbf{R}_{\Omega-\Delta}^{-1}$. In this case

$$\text{SBR}(\delta_{\mathbf{d}}^{\text{CEM}}) = \mathbf{d}^T \mathbf{R}_{\Omega-\Delta}^{-1} \mathbf{d} \approx \mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}^T \quad (30)$$

which further shows that the minimum variance in (26) is actually SBR in (29).

A crucial finding is that the maximum SNR of LRT in (8), the maximum SNR of OSP in (18), and the maximum SNR of CEM in (30) appear in the same functional forms with the only difference in using different matrices, \mathbf{K}^{-1} by LRT, $P_{\mathbf{U}_{p-1}}^{\perp}$ by OSP, and \mathbf{R}^{-1} by CEM, respectively, to suppress BKG. This indicates that this maximum SNR is the key idea used to derive TAC-AD in Section III.

III. TAC-AD: CONVERSION OF A TARGET DETECTOR TO AN ANOMALY DETECTOR

In [18], three approaches were derived to design target detectors: hypothesis testing-based LRT, SNR-based detectors, and SA-based detectors. Unfortunately, for a target detector to perform anomaly detection, we must take care of the target knowledge used by a target detector without using any prior knowledge. As easily as it is said, such work is not straightforward. Chang [24] coined a novel idea, called DVT, which introduced a dummy variable as a hidden unspecified target signature as a liaison to bridge the gap between a target detector and an AD.

A. Dummy Variable Trick [24]

For a given problem A, which cannot be solved directly, DVT creates an auxiliary variable ξ to reformulate a new solvable problem B by using ξ as a liaison to bridge problem A to be solved and a solvable equivalent problem B. Thus, after problem B is solved, the auxiliary variable x is vanished. Because of its hidden role, x is considered a dummy variable. In other words, problem A indicates a given problem to be solved. Since there is no direct approach that can be used to solve problem A, problem A is then translated to its equivalent problem, problem B, which can be solved by introducing ξ as a hidden variable.

In what follows, DVT will be used to convert the hypothesis testing-based LRT, SNR-based detectors, and SA-based detectors in [18] to their corresponding ADs.

B. LRT-AD

According to (5) and (7), there are two types of LRT-based detectors that can be derived. One is (5) that includes the global mean μ and gives rise to the Mahalanobis distance (MD)-AD, which turns out to be RX-AD. The other is (7) that does not mean μ and leads to LRT-AD.

1) RX-AD:

Theorem 1:

$$\delta_{\kappa}^{\text{MD-AD}}(\mathbf{r}) = \kappa(\mathbf{r} - \mu)^T \mathbf{K}^{-1}(\mathbf{r} - \mu), \quad \text{for some constant } \kappa \quad (31)$$

with

$$\delta_{\kappa=1}^{\text{RX-AD}}(\mathbf{r}) = \delta_1^{\text{MD-AD}}(\mathbf{r}) = (\mathbf{r} - \mu)^T \mathbf{K}^{-1}(\mathbf{r} - \mu). \quad (32)$$

Proof: To make LRT an AD, we take advantage of the DVT developed in [24] by replacing the known desired target signature \mathbf{d} by an unknown but specified target (UST) signature \mathbf{d}^{UST} to convert $\mathbf{d}^T \mathbf{K}^{-1}(\mathbf{r} - \mu)$ in (5) to $(\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1}(\mathbf{r} - \mu)$. Then, an LRT-based AD (LRT-AD) should be the one that maximizes

$$\max_{\mathbf{d}^{\text{UST}}} \left\{ (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1}(\mathbf{r} - \mu) \right\}. \quad (33)$$

To solve (33), we can apply Schwarz's inequality to (33) to yield

$$\begin{aligned} (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1}(\mathbf{r} - \mu) &= \left\| (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1/2} \mathbf{K}^{-1/2}(\mathbf{r} - \mu) \right\| \\ &\leq \left\| (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1/2} \right\| \left\| \mathbf{K}^{-1/2}(\mathbf{r} - \mu) \right\| \\ &= \left\| \mathbf{K}^{-1/2} \mathbf{d}^{\text{UST}} \right\| \left\| \mathbf{K}^{-1/2}(\mathbf{r} - \mu) \right\| \end{aligned} \quad (34)$$

where $\left\| (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1/2} \right\| = \left\| \mathbf{K}^{-1/2} \mathbf{d}^{\text{UST}} \right\|$ and the equality holds if and only if

$$\mathbf{K}^{-1/2} \mathbf{d}^{\text{UST}} = \kappa \mathbf{K}^{-1/2}(\mathbf{r} - \mu) \Rightarrow \mathbf{d}^{\text{UST}} = \kappa(\mathbf{r} - \mu) \quad (35)$$

for some constant κ . Substituting (35) into (33), we obtain

$$\max_{\mathbf{d}^{\text{UST}}} (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1}(\mathbf{r} - \mu) = \kappa(\mathbf{r} - \mu)^T \mathbf{K}^{-1}(\mathbf{r} - \mu) \quad (36)$$

where \mathbf{d}^{UST} is introduced as a dummy parameter variable and is vanished from (36). Since $(\mathbf{r} - \mu)^T \mathbf{K}^{-1}(\mathbf{r} - \mu)$ in (35) is exactly MD between \mathbf{r} and μ , (36) is also referred to as MD-AD specified by (31). Setting $\kappa = 1$ in (31) yields exactly the AD derived by Reed and Xiaoli, RX-AD in (32). ■

2) LRT-AD: Using a similar treatment derived for Theorem 1, we can convert (7) to another LRT-AD in the following theorem.

Theorem 2:

$$\delta_{\kappa}^{\text{LRT-AD}}(\mathbf{r}) = \kappa \mathbf{r}^T \mathbf{K}^{-1} \mathbf{r}, \quad \text{for some constant } \kappa. \quad (37)$$

Proof: By virtue of DVT, we can use \mathbf{d}^{UST} to convert $\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r}$ in (7) to

$$(\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1} \mathbf{r}. \quad (38)$$

Then, an LRT-AD should be the one that maximizes the following SNR:

$$\max_{\mathbf{d}^{\text{UST}}} \left\{ (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1} \mathbf{r} \right\}. \quad (39)$$

To solve (39), we can apply Schwarz's inequality to (39) to yield

$$\begin{aligned} (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1} \mathbf{r} &= \|(\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1/2} \mathbf{K}^{-1/2} \mathbf{r}\| \\ &\leq \|\mathbf{K}^{-1/2} \mathbf{d}^{\text{UST}}\| \|\mathbf{K}^{-1/2} \mathbf{r}\| \end{aligned} \quad (40)$$

where $\|(\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1/2}\| = \|\mathbf{K}^{-1/2} \mathbf{d}^{\text{UST}}\|$ and the equality holds if and only if

$$\mathbf{K}^{-1/2} \mathbf{d}^{\text{UST}} = \kappa \mathbf{K}^{-1/2} \mathbf{r} \Rightarrow \mathbf{d}^{\text{UST}} = \kappa \mathbf{r} \quad (41)$$

for some constant κ . Substituting (41) into (39), we obtain

$$\max_{\mathbf{d}^{\text{UST}}} (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1} \mathbf{r} = \kappa \mathbf{r}^T \mathbf{K}^{-1} \mathbf{r} \quad (42)$$

where \mathbf{d}^{UST} is introduced as a dummy parameter variable and is vanished from (42). ■

Since AD $\delta_{\kappa}^{\text{LRT-AD}}(\mathbf{r})$ in Theorem 2 uses the covariance matrix \mathbf{K} to suppress BKG, the LRT-AD in (42) is also called K-AD, and $\delta_{\kappa}^{\text{K-AD}}(\mathbf{r})$ in [23] to emphasize the role of \mathbf{K} plays in the LRT-AD.

C. OSP-AD

Similarly, we can also use the DVT proposed in [24] to show how to convert the OSP target detector in (16) to an AD in the following theorem.

Theorem 3:

$$\delta_{\kappa}^{\text{OSP-AD}}(\mathbf{r}) = \kappa \mathbf{r}^T \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r}, \quad \text{for some constant } \kappa. \quad (43)$$

Proof: To make OSP an AD, we take advantage of the DVT developed in [24] by replacing the known desired target signature \mathbf{d} by a UST signature \mathbf{d}^{UST} to convert OSP in (15) to

$$(\mathbf{d}^{\text{UST}})^T \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r}. \quad (44)$$

Then, an OSP-AD should be the one that maximizes

$$\max_{\mathbf{d}^{\text{UST}}} \left\{ (\mathbf{d}^{\text{UST}})^T \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r} \right\}. \quad (45)$$

To solve (45), we can apply Schwarz's inequality to (45) to yield

$$\begin{aligned} (\mathbf{d}^{\text{UST}})^T \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r} &= \|(\mathbf{d}^{\text{UST}})^T \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r}\| \\ &\leq \|(\mathbf{d}^{\text{UST}})^T \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp}\| \|\mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r}\| \end{aligned} \quad (46)$$

where the equality holds if and only if

$$(\mathbf{d}^{\text{UST}})^T \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} = \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r} \Rightarrow \kappa \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r} \Rightarrow \mathbf{d}^{\text{UST}} = \kappa \mathbf{r} \quad (47)$$

for some constant κ . Substituting (47) into (45), we obtain an OSP-AD as

$$\max_{\mathbf{d}^{\text{UST}}} (\mathbf{d}^{\text{UST}})^T \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r} = \kappa \mathbf{r}^T \mathbf{P}_{\mathbf{U}_{p-1}}^{\perp} \mathbf{r} \quad (48)$$

where \mathbf{d}^{UST} is introduced as a dummy parameter variable and is vanished from (48). The OSP-AD derived in (48) turns out to be exactly the same as OSP-AD derived in [31]. ■

D. CEM-AD

Since AD is not supposed to know the targets that it detects, there is no specific \mathbf{d} available to be used for detection. Thus, direct use of CEM in AD is NA. If we directly replace \mathbf{d} in (25) by \mathbf{d}^{UST} , using the unknown constant in the detector design as the constraint $(\mathbf{d}^{\text{UST}})^T \mathbf{w} = 1$ does not make sense. In this case, we can use (229) as the SBR constraint, i.e., $(\mathbf{d}^{\text{UST}})^T \mathbf{w} = \text{SBR} = (\mathbf{d}^{\text{UST}})^T \mathbf{R}^{-1} \mathbf{d}^{\text{UST}}$ to replace $\mathbf{d}^T \mathbf{w} = 1$ used in (23) to derive a new constrained optimization problem is then given by

$$\begin{aligned} \min_{\mathbf{w}} \{ \mathbf{w}^T \mathbf{R} \mathbf{w} \} \\ \text{s.t. } (\mathbf{d}^{\text{UST}})^T \mathbf{w} = \text{SBR} = (\mathbf{d}^{\text{UST}})^T \mathbf{R}^{-1} \mathbf{d}^{\text{UST}}. \end{aligned} \quad (49)$$

The solution to (49) is proven in [24] and given by

$$\delta_{\mathbf{d}^{\text{UST}}}^{\text{SBR-CEM}}(\mathbf{r}) = (\mathbf{w}^{\text{SBR-CEM}})^T \mathbf{r} = (\mathbf{d}^{\text{UST}})^T \mathbf{R}^{-1} \mathbf{r} \quad (50)$$

which is exactly the same as the UN version of CEM in (25), and UN-CEM is given by

$$\delta_{\mathbf{d}}^{\text{UN-CEM}}(\mathbf{r}) = \mathbf{d}^T \mathbf{R}^{-1} \mathbf{r}. \quad (51)$$

Theorem 4:

$$\delta_{\kappa}^{\text{CEM-AD}}(\mathbf{r}) = \kappa \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r}, \quad \text{for some constant } \kappa. \quad (52)$$

Proof: Using DVT and applying Schwarz's inequality to (51) result in

$$\begin{aligned} (\mathbf{d}^{\text{UST}})^T \mathbf{R}^{-1} \mathbf{r} &= \|(\mathbf{d}^{\text{UST}})^T \mathbf{R}^{-1/2} \mathbf{R}^{-1/2} \mathbf{r}\| \\ &\leq \|(\mathbf{d}^{\text{UST}})^T \mathbf{R}^{-1/2}\| \|\mathbf{R}^{-1/2} \mathbf{r}\| = \|\mathbf{R}^{-1/2} \mathbf{r}\|^2 \end{aligned} \quad (53)$$

where the equality holds if and only if $\mathbf{t}^{\text{UST}} = \kappa \mathbf{r}$ for some constant κ . By (53), we obtain

$$\max_{\mathbf{d}^{\text{UST}}} (\mathbf{d}^{\text{UST}})^T \mathbf{R}^{-1} \mathbf{r} = \kappa \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \quad (54)$$

which turns out to be the R-SNR detector derived by SNR in [1, eq. (48)]. Setting $\kappa = 1$ in (54) yields

$$\delta_{\kappa=1}^{\text{CEM-AD}}(\mathbf{r}) = \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \quad (55)$$

which is exactly R-AD derived in [33]. ■

E. Matched Filter-Based AD

There are also several matched filter-based target detectors derived in [18], which can be converted to AD, for example, normalized LRT in [18], AMF, normalized AMD (NAMD) in [41], and ASD in [20] defined in the following:

$$\delta^{\text{NLRT}}(\mathbf{r}) = \frac{\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r}}{\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d}} \quad (56)$$

which is the normalization of LRT (7)

$$\delta^{\text{AMF}}(\mathbf{r}) = \delta^{\text{GLRT}}(\mathbf{r}) = \frac{(\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2}{\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d}} \quad (57)$$

$$\delta^{\text{NAMD}}(\mathbf{r}) = \frac{(\mathbf{d} - \mu)^T \mathbf{K}^{-1} (\mathbf{r} - \mu)}{(\mathbf{d} - \mu)^T \mathbf{K}^{-1} (\mathbf{d} - \mu)} \quad (58)$$

which is the normalization of LRT (5)

$$\delta^{\text{ASD}}(\mathbf{r}) = \left(\frac{\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r}}{\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d}} \right)^2. \quad (59)$$

However, there is an issue arising in deriving TAC-AD for the above target detectors specified by (56)–(59). This is mainly because the normalization constants, $\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d}$ and $(\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{d} - \boldsymbol{\mu})$ used in these detectors, are independent of the data sample \mathbf{r} . In other words, the target detection maps produced by target detectors specified by (56)–(59) can be scaled by any constant without affecting detection performance, e.g., the ROC analysis [1]. The use of these normalization constants is to guarantee that the detected target abundance fraction of \mathbf{d} present in each data sample vector \mathbf{r} reflects the true abundance of \mathbf{d} . When these target detectors are converted to their corresponding ADs, these constants must be removed due to the fact that, without knowing the target signature \mathbf{d} to be detected, the target-converted ADs are immune from \mathbf{d} . With no prior knowledge of the target signature \mathbf{d} , the detectors in (56)–(59) without normalization constants will still produce the detection maps to show true detected target abundance fraction of the data sample vector \mathbf{r} . Interestingly, when we apply DVT to these normalized target detectors as we did for Theorems 1–4, we immediately find the maximum value of these target detectors is always one. This makes sense because, when \mathbf{r} is replaced by \mathbf{d} , the target abundance fraction of \mathbf{d} present in \mathbf{r} detected by the normalized target detectors should be 100% of purity, which is exactly one. This also explains why LSOSP, $\delta_{\mathbf{d}}^{\text{LSOSP}}(\mathbf{r})$ in (17), uses $\mathbf{d}^T \mathbf{P}_{\mathbf{U}_{p-1}}^\perp \mathbf{d}$ for normalization, and $\delta_{\mathbf{d}}^{\text{CEM}}(\mathbf{r})$ in (25) uses $\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}$ for normalization to detect the true target abundance fraction of the \mathbf{d} present in each data sample vector \mathbf{r} . When they are applied to anomaly detection, $\max_{\mathbf{d}} \delta_{\mathbf{d}}^{\text{LSOSP}}(\mathbf{r}) = 1$ (i.e., maximum SNR = 1) and $\max_{\mathbf{d}} \delta_{\mathbf{d}}^{\text{CEM}}(\mathbf{r}) = 1$ (i.e., maximum SNR = 1) for all data sample vectors \mathbf{r} . This provides the evidence that Theorem 3 uses UN-OSP in (19) for $\delta_{\mathbf{d}}^{\text{OSP}}(\mathbf{r})$ without normalization to derive OSP-AD in (43), and Theorem 4 uses UN-CEM in (51) for $\delta_{\mathbf{d}}^{\text{SBR-CEM}}(\mathbf{r})$ to derive CEM-AD in (52).

As noted above, to obtain TAC-AD for the target detectors specified by (55)–(58), we need to apply DVT to their UN counterparts to yield the maximum SNR as follows:

$$\begin{aligned} \delta^{\text{LRT-AD}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}}} (\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1} \mathbf{r} \\ &= \kappa \mathbf{r}^T \mathbf{K}^{-1} \mathbf{r} = \delta^{\text{K-AD}}(\mathbf{r}) \end{aligned} \quad (60)$$

$$\begin{aligned} \delta^{\text{AMF-AD}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}}} \left((\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1} \mathbf{r} \right)^2 \\ &= \kappa (\mathbf{r}^T \mathbf{K}^{-1} \mathbf{r})^2 = \delta^{\text{ACE-AD}}(\mathbf{r}) \end{aligned} \quad (61)$$

$$\begin{aligned} \delta^{\text{AMD-AD}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}}} (\mathbf{d}^{\text{UST}} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \\ &= \kappa (\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}) = \delta^{\text{RX-AD}}(\mathbf{r}) \end{aligned} \quad (62)$$

$$\begin{aligned} \delta^{\text{ASD-AD}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}}} \left((\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1} \mathbf{r} \right)^2 = (\mathbf{r}^T \mathbf{K}^{-1} \mathbf{r})^2 \\ &= \delta^{\text{AMF-AD}}(\mathbf{r}) = \delta^{\text{ACE-AD}}(\mathbf{r}) = (\delta^{\text{K-AD}}(\mathbf{r}))^2 \end{aligned} \quad (63)$$

all of which end up with $\delta^{\text{K-AD}}(\mathbf{r})$ derived in Theorem 2, $(\delta^{\text{K-AD}}(\mathbf{r}))^2$ in Theorem 6, and $\delta^{\text{RX-AD}}(\mathbf{r})$ in Theorem 1.

IV. GENERALIZED SNR-BASED TAC-AD

It has been shown in [18] that using SNR as a criterion [42] to derive TD is exactly the same as the hypothesis testing to derive LRT-TD (5) and (7), while the minimum variance-based CEM in (25) can be also derived by the SNR-based CEM. However, this relationship is only true when the prior target knowledge \mathbf{d} is given. If these TDs are converted to their corresponding ADs, \mathbf{d} is unknown and must be estimated. It is obvious that a direct estimation is not feasible due to a lack of prior knowledge. The concept of DVT proposed in [24] plays a key role of liaison to bridge a pathway between a target detector and AD with \mathbf{d} replaced by \mathbf{d}^{UST} as a dummy variable. Then, MD is used to estimate \mathbf{d}^{UST} . This estimated \mathbf{d}^{UST} is then used as if it were the true \mathbf{d} to derive the SNR-TD in (8). Such SNR-derived TD is referred to as a GSNR-TD. In what follows, we used DVT to derive the three SNR-derived TDs in [18] to their corresponding TAC-ADs by operating SNR in three different whitened spaces.

A. Generalized Data Sphered-SNR (GDS-SNR) Anomaly Detector

In this section, we operate SNR in data sphered (DS) space that removed data statistics of the first two orders (i.e., mean and variances/covariances) to give rise to a GDS-SNR target detector given by

$$\begin{aligned} \delta_{\mathbf{d}}^{\text{GDS-SNR}}(\hat{\mathbf{r}}) &= \kappa \frac{(\hat{\mathbf{d}}^T \hat{\mathbf{r}})^2}{\|\hat{\mathbf{d}}\|^2} \Rightarrow \delta_{\mathbf{d}}^{\text{GDS-SNR}}(\mathbf{r}) \\ &= \kappa \frac{((\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}))^2}{(\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{d} - \boldsymbol{\mu})} \end{aligned} \quad (64)$$

where

$$\hat{\mathbf{d}} = \mathbf{K}^{-1/2} (\mathbf{d} - \boldsymbol{\mu}) \quad \text{and} \quad \hat{\mathbf{r}} = \mathbf{K}^{-1/2} (\mathbf{r} - \boldsymbol{\mu}) \quad (65)$$

are defined in the DS space, denoted by $\hat{\mathbf{X}}$. As noted in Section III-D, to convert (64) to its TAC-AD, the normalization constant $(\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{d} - \boldsymbol{\mu})$ in (64) is independent of the data sample vector \mathbf{r} and should be removed to yield a UN-GDS-SNR detector

$$\delta_{\mathbf{d}}^{\text{UN-GDS-SNR}}(\mathbf{r}) = ((\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}))^2. \quad (66)$$

Then, DVT can be used to derive the TAC-AD of $\delta_{\mathbf{d}}^{\text{GDS-SNR-AD}}(\mathbf{r})$ in the following theorem.

Theorem 5:

$$\begin{aligned} \delta_{\mathbf{d}}^{\text{GDS-SNR-AD}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}}} \delta_{\mathbf{d}^{\text{UST}}}^{\text{UN-GDS-SNR}}(\mathbf{r}) \\ &= \kappa ((\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}))^2 = \kappa (\delta^{\text{RX-AD}}(\mathbf{r}))^2 \\ &\quad \text{for some constant } \kappa. \end{aligned} \quad (67)$$

Proof: Using (66) and DVT

$$\begin{aligned} \delta_{\mathbf{d}}^{\text{GDS-SNR-AD}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}}} \left((\mathbf{d}^{\text{UST}} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \right)^2 \\ &= \max_{\mathbf{d}^{\text{UST}}} \left[(\mathbf{K}^{-1/2} (\mathbf{d}^{\text{UST}} - \boldsymbol{\mu}))^T (\mathbf{K}^{-1/2} (\mathbf{r} - \boldsymbol{\mu})) \right]^2. \end{aligned} \quad (68)$$

By Schwarz's inequality

$$(\mathbf{K}^{-1/2}(\mathbf{d}^{\text{UST}} - \boldsymbol{\mu}))^T (\mathbf{K}^{-1/2}(\mathbf{r} - \boldsymbol{\mu})) \leq \|\hat{\mathbf{d}}^{\text{UST}}\| \|\hat{\mathbf{r}}\|$$

with equality hold if and only if $\hat{\mathbf{d}}^{\text{UST}} = \kappa \hat{\mathbf{r}}$ (69)

which yields (67), which turns out to be a square of RX-AD in (32). ■

B. Generalized K-Whitened SNR (GK-SNR) Anomaly Detector

Similar to GDS-SNR, we can also operate SNR in K-whitened space, which whitens the second-order data statistics (i.e., variances/covariances) characterized by the covariance matrix \mathbf{K} . The resulting target detector is a GK-SNR target detector given by

$$\delta_{\hat{\mathbf{d}}}^{\text{GK-SNR}}(\hat{\mathbf{r}}) = \frac{(\hat{\mathbf{d}}^T \hat{\mathbf{r}})^2}{\|\hat{\mathbf{d}}\|^2} \Rightarrow \delta_{\mathbf{d}}^{\text{GK-SNR}}(\mathbf{r}) = \kappa \frac{(\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2}{\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d}} \quad (70)$$

where

$$\hat{\mathbf{d}} = \mathbf{K}^{-1/2} \mathbf{d} \quad \text{and} \quad \hat{\mathbf{r}} = \mathbf{K}^{-1/2} \mathbf{r} \quad (71)$$

are defined in the K-whitened space, denoted by $\hat{\mathbf{X}}$. Since the normalization constant $\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d}$ has nothing to do with the data sample vector \mathbf{r} , a UN-GK-SNR detector can be defined by

$$\delta_{\mathbf{d}}^{\text{UN-GK-SNR}}(\mathbf{r}) = (\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2. \quad (72)$$

Using exactly the same treatment carried out in Theorem 5, we can derive the following theorem.

Theorem 6:

$$\begin{aligned} \delta_{\mathbf{d}^{\text{UST}} \in \mathbf{X}}^{\text{GK-SNR-AD}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}} \in \mathbf{X}} \left((\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1} \mathbf{r} \right)^2 \\ &= \kappa (\mathbf{r}^T \mathbf{K}^{-1} \mathbf{r})^2 = (\delta^{\text{K-AD}}(\mathbf{r}))^2 = (\delta^{\text{ASD-AD}}(\mathbf{r}))^2 \end{aligned}$$

for some constant κ (73)

which turns out to be ASD-AD in (63).

C. Generalized R-Whitened SNR (GR-SNR) Anomaly Detector

Once again, we can also operate SNR in R-whitened space, which whitens the data statistics characterized by the covariance matrix \mathbf{R} to yield a GR-SNR target detector given by

$$\delta_{\hat{\mathbf{d}}}^{\text{GR-SNR}}(\hat{\mathbf{r}}) = \kappa \frac{(\hat{\mathbf{d}}^T \hat{\mathbf{r}})^2}{\|\hat{\mathbf{d}}\|^2} \Rightarrow \delta_{\mathbf{d}}^{\text{GR-SNR}}(\mathbf{r}) = \kappa \frac{(\mathbf{d}^T \mathbf{R}^{-1} \mathbf{r})^2}{\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}} \quad (74)$$

where $\hat{\mathbf{r}}$ is a data sample vector in the R-whitened space, denoted by $\hat{\mathbf{X}}$, given by

$$\hat{\mathbf{r}} = \mathbf{R}^{-1/2} \mathbf{r}. \quad (75)$$

Since $\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}$ in (74) is a normalization constant and is immune from the data sample \mathbf{r} , a UN-GR-SNR target detector can be defined by

$$\delta_{\mathbf{d}}^{\text{UN-GR-SNR}}(\mathbf{r}) = (\mathbf{d}^T \mathbf{R}^{-1} \mathbf{r})^2. \quad (76)$$

Using (76), we can prove the following theorem.

Theorem 7:

$$\delta^{\text{GR-SNR-AD}}(\mathbf{r}) = \kappa (\mathbf{r}^T \mathbf{R}^{-1} \mathbf{r})^2 = \kappa (\delta^{\text{CEM-AD}}(\mathbf{r}))^2$$

for some constant κ . (77)

Proof: By virtue of DVT, the TAC-AD converted from (76) is to maximize GR-SNR as follows:

$$\begin{aligned} \delta^{\text{GDS-SNR-AD}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}}} \left((\mathbf{d}^{\text{UST}})^T \mathbf{R}^{-1} \mathbf{r} \right)^2 \\ &= \max_{\mathbf{d}^{\text{UST}}} \left[(\mathbf{R}^{-1/2} \mathbf{d}^{\text{UST}})^T (\mathbf{R}^{-1/2} \mathbf{r}) \right]^2. \end{aligned} \quad (78)$$

By Schwarz's inequality

$$(\mathbf{R}^{-1/2} \mathbf{d}^{\text{UST}})^T (\mathbf{R}^{-1/2} \mathbf{r}) \leq \|\bar{\mathbf{d}}^{\text{UST}}\| \|\bar{\mathbf{r}}\|$$

with equality hold if and only if $\bar{\mathbf{d}}^{\text{UST}} = \kappa \bar{\mathbf{r}}$. (79)

Substituting (79) into (78) yields (77). ■

V. SPECTRAL ANGLE-BASED TAC-AD

In addition to LRT-AD and SNR-AD, there is a third type of target detectors derived in [18], which are generally referred to as matched filter-based detectors in [20], [21], and [40, Ch. 16] but actually perform SA discrimination. Two widely used SA measures are discussed as follows.

A. SAM-AD

The simplest SA is SAM in [27], which measures the angle between two spectral vectors defined by

$$\text{SAM}(\mathbf{d}, \mathbf{r}) = \cos \theta = \frac{\langle \mathbf{d}, \mathbf{r} \rangle}{\|\mathbf{d}\| \|\mathbf{r}\|} = \frac{\mathbf{d}^T \mathbf{r}}{\|\mathbf{d}\| \|\mathbf{r}\|}. \quad (80)$$

As noted in (80), SAM is a normalized measure, which removes the vector lengths of \mathbf{d} and \mathbf{r} to get rid of the dependence on \mathbf{d} and \mathbf{r} , and only retains their angle, $\cos \theta$. However, for SAM to be converted to AD, SAM must include the data sample \mathbf{r} in its form. To address this issue, SAM should not be normalization. In this case, we define unnormalized SAM (UN-SAM) as

$$\text{UN-SAM}(\mathbf{d}, \mathbf{r}) = \langle \mathbf{d}, \mathbf{r} \rangle = \mathbf{d}^T \mathbf{r}. \quad (81)$$

Theorem 8:

$$\delta^{\text{SAM-AD}}(\mathbf{r}) = \kappa \|\mathbf{r}\|^2, \quad \text{for some constant } \kappa. \quad (82)$$

Proof: Using (64) and DVT, we have

$$\delta^{\text{SAM-AD}}(\mathbf{r}) = \max_{\mathbf{d}^{\text{UST}}} \text{UN-SAM}(\mathbf{d}^{\text{UST}}, \mathbf{r}). \quad (83)$$

By Schwarz's inequality, we obtain

$$\begin{aligned} \text{UN-SAM}(\mathbf{d}^{\text{UST}}, \mathbf{r}) &= (\mathbf{d}^{\text{UST}})^T \mathbf{r} \leq \|\mathbf{d}^{\text{UST}}\| \|\mathbf{r}\| \\ &\text{with equality hold if and only if } \mathbf{d}^{\text{UST}} = \kappa \mathbf{r}. \end{aligned} \quad (84)$$

Substituting $\kappa \mathbf{r}$ for \mathbf{d}^{UST} in (84) yields (83), which is the vector length of the data sample vector \mathbf{r} .

Interestingly, if we sphere data by removing first-order statistics, mean $\boldsymbol{\mu}$, second statistics, and covariance matrix \mathbf{K} to produce a sphered data space, \mathbf{X}^{DS} , then we can operate $\delta^{\text{SAM-AD}}(\hat{\mathbf{r}})$ on $\hat{\mathbf{X}} = \mathbf{X}^{\text{DS}}$ with $\hat{\mathbf{r}} = \mathbf{K}^{-1/2}(\mathbf{r} - \boldsymbol{\mu}_b) \hat{\mathbf{d}}^{\text{UST}} = \mathbf{K}^{-1/2}(\mathbf{d}^{\text{UST}} - \boldsymbol{\mu}_b)$ and $\hat{\mathbf{r}} = \mathbf{K}^{-1/2}(\mathbf{r} - \boldsymbol{\mu}_b)$ to yield a DS version

of $\delta^{\text{SAM-AD}}(\mathbf{r})$ and $\hat{\delta}^{\text{SAM-AD}}(\hat{\mathbf{r}})$ on $\hat{\mathbf{X}}$. According to Theorem 8, we can obtain

$$\hat{\delta}^{\text{SAM-AD}}(\hat{\mathbf{r}}) = \kappa \|\hat{\mathbf{r}}\|^2, \quad \text{for some constant } \kappa \quad (85)$$

which includes RX-AD as a special case because of $\hat{\mathbf{r}} = \mathbf{K}^{-1/2}(\mathbf{r} - \boldsymbol{\mu})$. That is,

$$\begin{aligned} \delta^{\text{SAM-AD}}(\mathbf{r}) &= \kappa \|\mathbf{r}\|^2 \rightarrow [\hat{\mathbf{r}} = \mathbf{K}^{-1/2}(\mathbf{r} - \boldsymbol{\mu})] \rightarrow \delta^{\text{SAM-AD}}(\hat{\mathbf{r}}) \\ &= \kappa \|\hat{\mathbf{r}}\|^2. \end{aligned} \quad (86)$$

B. ACE-Based AD

Another popular and widely used SA is ACE originally developed in [20] and defined as

$$\text{ACE}(\mathbf{d}, \mathbf{r}) = \frac{(\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2}{(\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d})(\mathbf{r}^T \mathbf{K}^{-1} \mathbf{r})}. \quad (87)$$

Since $\mathbf{K}^{-1/2}$ is a whitening matrix, we can use $\mathbf{K}^{-1/2}$ to whiten the original data \mathbf{X} to yield $\mathbf{K}^{-1/2}$ -whitened data space $\hat{\mathbf{X}}$ as we did for GK-SNR-AD in Section II-B. Then, (87) can be rederived as

$$\begin{aligned} \text{ACE}(\mathbf{d}, \mathbf{r}) &= \left[\frac{(\mathbf{K}^{-1/2} \mathbf{d})^T (\mathbf{K}^{-1/2} \mathbf{r})}{\|\mathbf{K}^{-1/2} \mathbf{d}\| \|\mathbf{K}^{-1/2} \mathbf{r}\|} \right]^2 = \left(\frac{\hat{\mathbf{d}}^T \hat{\mathbf{r}}}{\|\hat{\mathbf{d}}\| \|\hat{\mathbf{r}}\|} \right)^2 \\ &= (\cos \hat{\theta})^2. \end{aligned} \quad (88)$$

By virtue of (88), when ACE is operated on the data space $\hat{\mathbf{X}}$, it becomes a square of SAM given by

$$\text{ACE}(\mathbf{d}, \mathbf{r}) = (\text{SAM}(\hat{\mathbf{d}}, \hat{\mathbf{r}}))^2. \quad (89)$$

Thus, similar to SAM, we need to define the UN version of ACE (UN-ACE) as

$$\text{UN-ACE}(\mathbf{d}, \mathbf{r}) = (\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2. \quad (90)$$

Theorem 9:

$$\begin{aligned} \delta^{\text{ACE-AD}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}}} \text{UACE}(\mathbf{d}^{\text{UST}}, \mathbf{r}) \\ &= \kappa (\mathbf{r}^T \mathbf{K}^{-1} \mathbf{r})^2 = \kappa (\delta^{\text{K-AD}}(\mathbf{r}))^2 \\ &\quad \text{for some constant } \kappa. \end{aligned} \quad (91)$$

Proof: Implementing DVT in (90) and using Schwarz's inequality, we obtain

$$\begin{aligned} \delta^{\text{ACE}}(\mathbf{r}) &= \max_{\mathbf{d}^{\text{UST}}} \text{UACE}(\mathbf{d}^{\text{UST}}, \mathbf{r}) = \max_{\mathbf{d}^{\text{UST}}} ((\mathbf{d}^{\text{UST}})^T \mathbf{K}^{-1} \mathbf{r})^2 \\ &= \left[(\mathbf{K}^{-1/2} \mathbf{d}^{\text{UST}})^T (\mathbf{K}^{-1/2} \mathbf{r}) \right]^2 \\ &\leq (\|\mathbf{K}^{-1/2} \mathbf{d}^{\text{UST}}\| \|\mathbf{K}^{-1/2} \mathbf{r}\|)^2 \end{aligned} \quad (92)$$

with equality hold if and only if

$$\mathbf{K}^{-1/2} \mathbf{d}^{\text{UST}} = \kappa \mathbf{K}^{-1/2} \mathbf{r} \Rightarrow \mathbf{d}^{\text{UST}} = \kappa \mathbf{r} \quad (93)$$

which implies

$$\delta^{\text{ACE-AD}}(\mathbf{r}) = \max_{\mathbf{d}^{\text{UST}}} \text{ACE}(\mathbf{d}^{\text{UST}}, \mathbf{r}) = \kappa (\mathbf{r}^T \mathbf{K}^{-1/2} \mathbf{r})^2 \quad (94)$$

which turns out to be a square of K-AD in (42). Setting $\kappa = 1$ yields ASD-AD in (63).

Furthermore, using the \mathbf{K} -whitened data sample \mathbf{r} , $\hat{\mathbf{r}} = \mathbf{K}^{-1/2} \mathbf{r}$, (92) can be reexpressed as

$$\delta^{\text{ACE-AD}}(\mathbf{r}) = \kappa (\mathbf{r}^T \mathbf{K}^{-1/2} \mathbf{r})^2 = (\hat{\mathbf{r}}^T \hat{\mathbf{r}})^2 = \left(\hat{\delta}^{\text{SAM-AD}}(\hat{\mathbf{r}}) \right)^2 \quad (95)$$

where

$$\hat{\delta}^{\text{SAM-AD}}(\hat{\mathbf{r}}) = \|\hat{\mathbf{r}}\|^2 \quad (96)$$

is exactly the SAM defined on $\hat{\mathbf{X}}$ according to Theorem 8. ■

Table I summarizes target detectors and their corresponding ADs.

It should be noted in Table I that there are four TAC-ADs, $\delta^{\text{SAM-AD}}(\mathbf{r}) = \|\mathbf{r}\|^2$, $(\delta^{\text{RX-AD}}(\mathbf{r}))^2$, $(\delta^{\text{CEM-AD}}(\mathbf{r}))^2$, and $\delta^{\text{ASD/ACE-AD}}(\mathbf{r}) = (\delta^{\text{K-AD}}(\mathbf{r}))^2$, which are new and have never been derived in the past.

As a concluding remark, with using specific target knowledge, target detectors can appear in different forms. However, when they are converted to their corresponding ADs, it turns out that many of them are reduced and become the same ADs, as shown in Table I. In particular, the TAC-ADs derived from ASD, ACE, and AMF, $\delta^{\text{ASD-AD}}(\mathbf{r})$, $\delta^{\text{ACE-AD}}(\mathbf{r})$, and $\delta^{\text{AMF-AD}}(\mathbf{r})$, become the same AD, $(\delta^{\text{K-AD}}(\mathbf{r}))^2 = (\delta^{\text{SAM-AD}}(\mathbf{r}))^2$. Specifically, AMD-AD is shown to be RX-AD, while RX-AD and ACE-AD can be also considered as special cases of SAM by operating the data sample vectors in DS and \mathbf{K} -whitened spaces, respectively, according to (82) and (92). The above evidence shows the crucial difference between target detectors and their corresponding TAC-ADs. Using various matched signatures provided by the target knowledge can be used to design different target detectors via normalizing matched signatures. This is why a target detector can be considered a search-before-detect detector since it makes use of specific matched signatures as a search target signature. However, when target knowledge is not available, the only available matched signature that can be used to design ADs is the data sample vector \mathbf{r} itself without normalization. As a result, AD must first clean up BKG or interferences prior to its detection, which leads an AD to a throw-before-detect detector. The “throw” operation can be generally accomplished by one of three preprocessing operations, DS, covariance matrix \mathbf{K} -whitening, and correlation matrix \mathbf{R} -whitening. Consequently, such a preprocessing gives rise to much simpler TAC-ADs, which can be characterized by covariance matrix \mathbf{K} -based ADs, such as DS-based AD, RX-AD, and \mathbf{K} -whitened-based AD, ACE-AD, and correlation matrix \mathbf{R} -based ADs, such as CEM-AD, OSP-ADs, OSP-AD, and SA-based ADs. Most interestingly, SA-based TAC-ADs become generalized versions of ADs converted from many matched filter-based target detectors, such as SAM-AD, ACE-AD, ASD-AD, and AMF-AD. It should be noted that such DS and data whitening processing is similar to but more effective than the BKG suppressed or removed by many AD methods that use training samples or Gaussian mixtures for the same “throw” purpose, as will be shown in Section VII.

TABLE I
TARGET DETECTORS AND THEIR CORRESPONDING TAC-ADS

| types | TD | AD | TD/AD forms | equations |
|-----------------------|---------|---------------|---|---------------|
| Pure pixel-based LRT | LRT | | $\delta^{\text{LRT}}(\mathbf{r}) = \mathbf{d}^T \mathbf{K}^{-1} \mathbf{r}$ | (7) |
| | | LRT-AD = K-AD | $\delta_{\kappa}^{\text{LRT-AD}}(\mathbf{r}) = \delta_{\kappa}^{\text{K-AD}}(\mathbf{r}) = \delta_{\kappa}^{\text{SAM-AD}}(\hat{\mathbf{r}}) = \kappa \mathbf{r}^T \mathbf{K}^{-1} \mathbf{r}$ | (42) and (96) |
| | GLRT | | $\delta^{\text{GLRT}}(\mathbf{r}) = (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu})$ | (5) |
| | | GLRT-AD | $\delta_{\kappa}^{\text{GLRT-AD}}(\mathbf{r}) = \kappa (\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu})$ | (31) |
| | | | $\delta^{\text{RX-AD}}(\mathbf{r}) = \delta_{\kappa=1}^{\text{GLRT-AD}}(\mathbf{r}) = (\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu})$ | (32) |
| Mixed-pixel based SNR | OSP | | $\delta^{\text{OSP}}(\mathbf{r}) = \mathbf{d}^T P_{U_{p-1}}^{\perp} \mathbf{r}$ | (16) |
| | | OSP-AD | $\delta^{\text{OSP-AD}}(\mathbf{r}) = \kappa \mathbf{r}^T P_{U_{p-1}}^{\perp} \mathbf{r}$ | (43) |
| Subpixel-based SBR | CEM | | $\delta^{\text{CEM}}(\mathbf{r}) = (\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d})^{-1} \mathbf{d}^T \mathbf{R}^{-1} \mathbf{r}$ | (25) |
| | | CEM-AD | $\delta_{\kappa}^{\text{CEM-AD}}(\mathbf{r}) = \kappa \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r}$ | (52) |
| Matched filters | AMF | | $\delta^{\text{AMF}}(\mathbf{r}) = (\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d})^{-1} (\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2$ | (57) |
| | | AMF-AD | $\delta_{\kappa}^{\text{AMF-AD}}(\mathbf{r}) = (\delta_{\kappa}^{\text{K-AD}}(\mathbf{r}))^2$ | (61) |
| | AMD | | $\delta^{\text{AMD}}(\mathbf{r}) = (\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu})$ | (58) |
| | | AMD-AD | $\delta_{\kappa}^{\text{AMD-AD}}(\mathbf{r}) = \delta_{\kappa}^{\text{GLRT-AD}}(\mathbf{r})$ | (62) |
| | ASD | | $\delta^{\text{ASD}}(\mathbf{r}) = (\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d})^{-2} (\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2$ | (59) |
| | | ASD-AD | $\delta_{\kappa}^{\text{ASD-AD}}(\mathbf{r}) = \delta_{\kappa}^{\text{ACE-AD}}(\mathbf{r}) = \delta_{\kappa}^{\text{AMF-AD}}(\mathbf{r}) = (\delta_{\kappa}^{\text{K-AD}}(\mathbf{r}))^2$ | (63) |
| GSR | GDS-SNR | | $\delta^{\text{GDS-SNR}}(\mathbf{r}) = \frac{((\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}))^2}{(\mathbf{d} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{d} - \boldsymbol{\mu})}$ | (64) |
| | | GDS-SNR-AD | $\delta_{\kappa}^{\text{GDS-SNR-AD}} = (\delta_{\kappa}^{\text{GLRT-AD}}(\mathbf{r}))^2$ | (67) |
| | GK-SNR | | $\delta^{\text{GK-SNR}}(\mathbf{r}) = \frac{(\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2}{\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d}}$ | (70) |
| | | GK-SNR-AD | $\delta_{\kappa}^{\text{GK-SNR-AD}}(\mathbf{r}) = (\delta_{\kappa}^{\text{K-AD}}(\mathbf{r}))^2$ | (73) |
| | GR-SNR | | $\delta^{\text{GR-SNR}}(\mathbf{r}) = \frac{(\mathbf{d}^T \mathbf{R}^{-1} \mathbf{r})^2}{\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}}$ | (74) |
| | | GR-SNR-AD | $\delta_{\kappa}^{\text{GR-SNR-AD}} = (\delta_{\kappa}^{\text{CEM-AD}}(\mathbf{r}))^2$ | (77) |
| | SAM | | $\delta^{\text{SAM}}(\mathbf{r}) = \frac{\mathbf{d}^T \mathbf{r}}{\ \mathbf{d}\ \ \mathbf{r}\ }$ | (80) |
| | | SAM-AD | $\delta_{\kappa}^{\text{SAM-AD}}(\mathbf{r}) = \kappa \ \mathbf{r}\ ^2$ | (82) |
| SA | | | $\delta_{\kappa}^{\text{SAM-AD}}(\hat{\mathbf{r}}) = \kappa \ \hat{\mathbf{r}}\ ^2 = \kappa \delta^{\text{RX-AD}}(\mathbf{r})$ | (85) |
| | ACE | | $\delta^{\text{ACE}}(\mathbf{r}) = \frac{(\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2}{(\mathbf{d}^T \mathbf{K}^{-1} \mathbf{d})(\mathbf{r}^T \mathbf{K}^{-1} \mathbf{r})}$ | (87) |
| | | ACE-AD | $\delta_{\kappa}^{\text{ACE-AD}}(\mathbf{r}) = \kappa (\mathbf{d}^T \mathbf{K}^{-1} \mathbf{r})^2 = (\delta_{\kappa}^{\text{K-AD}}(\mathbf{r}))^2$ | (91) |
| | | | $\delta_{\kappa}^{\text{ACE-AD}}(\mathbf{r}) = (\delta_{\kappa}^{\text{SAM-AD}}(\hat{\mathbf{r}}))^2$ | (96) |

VI. EFFECTIVE ANOMALY SPACE

Since TAC-AD is directly obtained from a target detector, it also inherits two issues arising in a target detector, which are the BKG suppression (BS) issue and a constraint issue on the known target knowledge. This is because target detection assumes targets to be known in advance, and the detection can be carried out by enhancing the known TD via a matched filter, while BS is generally considered as noise and is not of major concern from a signal detection point of view. However, this is not true for anomaly detection. Without knowing known target knowledge, anomalies face two critical issues: 1) unknowingly embedded in BKG and 2) contaminated by noise. More specifically, it says that anomalies are, in fact, sandwiched between BKG and noise. Thus, when a target detector works well, it does not necessarily imply that the same success can be repeated for its derived TAC-AD. A good example is ACE that was shown to perform well in target detection but indeed performed very poorly in BS [18]. This indicates that TAC-AD also suffers from the same issue. To improve anomaly detectability and BS (ADBS) of TAC-AD, a concept of EAS recently developed in [34] provides a feasible solution. The motivation of EAS was inspired by a need of addressing both BKG and noise issues.

As for BKG issue, several approaches have been the main streams in anomaly detection over the past years. One is referred to BKG separation-before-anomaly detection, which separates BKG before anomaly detection takes place [43], [44]. Another is BKG removal-before-anomaly detection, which removes estimated BKG [45], [46] or reconstructed BKG via autoencoder (AE) or generative adversarial network (GAN) [47], [48], [49] prior to anomaly detection. Recently, a third approach is to use graph theory to better characterize anomalies from BKG in a rather robust manner [50], [51]. In contrast to the graph-based approach, a fourth approach is the LRR developed by Xu et al. [52], which formulates anomaly detection as a constrained optimization problem solving $\min_{\mathbf{S}, \mathbf{E}} \{\text{rank}(\mathbf{S}) + \lambda \|\mathbf{E}\|_{2,1}\}$ subject to $\mathbf{X} = \mathbf{D}\mathbf{S} + \mathbf{E}$, where \mathbf{D} is the BKG dictionary with \mathbf{S} being representation coefficients, \mathbf{E} is its residual, and $\|\cdot\|_{2,1}$ is the $l_{2,1}$ norm defined as the sum of l_2 norm of the column of a matrix. The parameter λ is used to regularize the sparse representation \mathbf{E} . In other words, $\mathbf{D}\mathbf{S}$ and \mathbf{E} represent BKG and anomalies, respectively. Many works based on LRR have been reported [53], [54]. As an alternative to LRR, a model-based approach, called robust PC analysis (RPCA) developed by Candès et al. [55] for blind source separation technique, decomposed the data matrix \mathbf{X} into $\mathbf{X} = \mathbf{L} + \mathbf{S}$ in a matrix decomposition. Analogous to RPCA, Vaswani et al. [56], [57] developed an approach, called robust subspace learning (RSL), which treated an outlier as an additive sparse corruption via sparse+low-rank ($\mathbf{S}+\mathbf{L}\mathbf{R}$) matrix decomposition. Using this interpretation, RPCA can be considered as a variant of RSL via $\mathbf{S}+\mathbf{L}\mathbf{R}$ formulation [58], [59]. In a recent report, Li et al. [60] took advantage of tensor Tucker decomposition to develop a PTA, which decomposed a hyperspectral imagery into a BKG tensor and an anomaly tensor corresponding to BKG and anomaly spaces, respectively.

Unfortunately, all the abovementioned approaches do not separate noise from either low-rank space \mathbf{L} or sparse space \mathbf{S} to address the issue of noise-corrupted anomalies. Thus, Li et al. [61] used the low-rank and sparse decomposition model (LSDM) to represent the sparse space as a mixture of Gaussian (MoG), referred to as LSDM-MoG, so that noise can be separated from anomalies. This LSDM-MoG was later extended by Feng et al. [62] using a density peak-guided collaborative representation to further improve anomaly detection performance. However, such an MoG approach may not be effective since noise is not necessarily Gaussian. This is particularly true for hyperspectral imagery, which generally contains non-Gaussian noise. To resolve this dilemma, two more attractive approaches were recently developed in the literature: LRSMD developed by Zhou and Tao [30], and CDA developed by Chen et al. [33], both of which address BKG issue in the low-rank subspace or principal component (PC) space while retaining anomalies in sparse subspaces by excluding noise in a separate component space.

Most recently, Chang and Chen [32], [63] developed an approach that used DS to remove the first- and second-order data statistics to produce sphered data, \mathbf{X}^{DS} . It assumes that global BKG can be characterized by the first two orders of data statistics (2OS), mean $\boldsymbol{\mu}$, and covariance matrix \mathbf{K} . By virtue of DS, MoG proposed in [61] and [62] can be completely annihilated in \mathbf{X}^{DS} before MoG leaked into the sparse space. The idea of DS is not new and has been widely used in communications and signal processing [1]. It maps the original data sample vector \mathbf{r} to its sphered data sample vector $\hat{\mathbf{r}}$ in the sphered data space $\hat{\mathbf{X}}$ by $\hat{\mathbf{r}} = \mathbf{K}^{-1/2}(\mathbf{r} - \boldsymbol{\mu})$, where $\mathbf{K}^{-1/2}$ is referred to as the whitening matrix in [1]. Details of DS can be found in [40, Sec. 6.3.1] along with its code [40, pp. 1000 and 1001].

Although DS can remove noises characterized by second-order statistics, including Gaussian noise and Gaussian BKG from the original data space, non-Gaussian characterized noise still exists in \mathbf{X}^{DS} . EAS adopts a rather different approach by taking advantage of the blind source separability resulting from ICA [64] to extract anomalies from ICs [65]. Unfortunately, such ICA-generated ICs cannot separate non-Gaussian noises from anomalies since non-Gaussian noises are characterized by high-order statistics and likely mixed with anomalies in ICs. EAS resolves this issue by using VD in [27], [66], [67], and [68] to estimate the total number of spectrally distinct target signal sources and MX-SVD [69] to estimate the desired number of prioritized ICs, j , needed to form an IC subspace, \mathbf{X}^{IC^j} , which can be also considered as anomaly space, \mathbf{X}^{AS} . In order to further clean up the leakage from non-Gaussian noises into \mathbf{IC}^j , the idea of SC proposed in [30] is then used for this purpose due to the fact that noise cannot be sparse and can be eliminated by SC. Including SC into \mathbf{IC}^j yields EAS, $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{SC}}^{\text{IC}^j}$.

In what follows, we describe a stage-by-stage process of constructing EAS.

A. VD to Estimate p

To construct EAS, the first task is to estimate the order of the low-rank space \mathbf{L} , m , and the order of the sparse

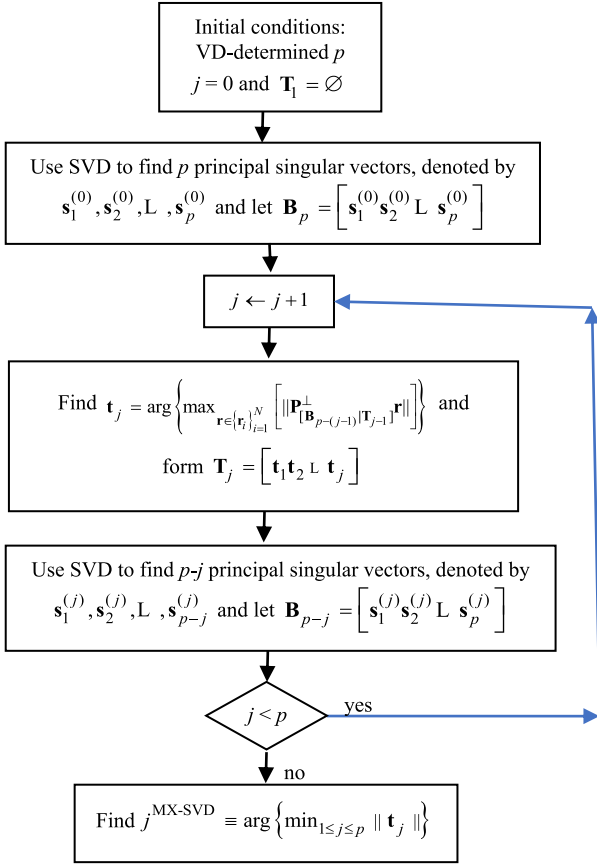


Fig. 1. Flowchart of MX-SVD.

space \mathbf{S} , j , respectively. Instead of directly estimating m , EAS takes advantage of the concept of VD, which is defined as the number of spectrally distinct signatures, p , present in a hyperspectral image. If we further assume that m and j can be used to specify the number of BKG signatures and sparse signal sources, respectively, then $p = m + j$. In other words, m can be estimated by VD via $m = p - j$. Since VD has been widely used in HSI and its code is available in [40, pp. 997–1000], we refer its details to [40].

B. MX-SVD to Estimate j

Once p is found by VD, the follow-up issue is to estimate j so that m can be, therefore, determined by $m = p - j$. Since anomalous signal sources can be considered rare signal sources, MX-SVD developed in [69] originally developed to estimate the rank of rare signal space can be used for this purpose. Its step-by-step implementation is described in the following along with its flowchart depicted in Fig. 1.

C. Construction of EAS

After m and j are estimated by VD and MX-SVD, an anomaly space, \mathbf{X}^{AS} , can be obtained by \mathbf{X}^{IC^j} , i.e., $\mathbf{X}^{\text{AS}} = \mathbf{X}^{\text{IC}^j}$. Although \mathbf{X}^{DS} already removes Gaussian noise via DS, \mathbf{X}^{AS} may still contain non-Gaussian noises. To further resolve this issue, the concept of SC developed in [30] is imposed on \mathbf{X}^{AS} to construct $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{SC}}^{\text{IC}^j}$. This is because noise cannot

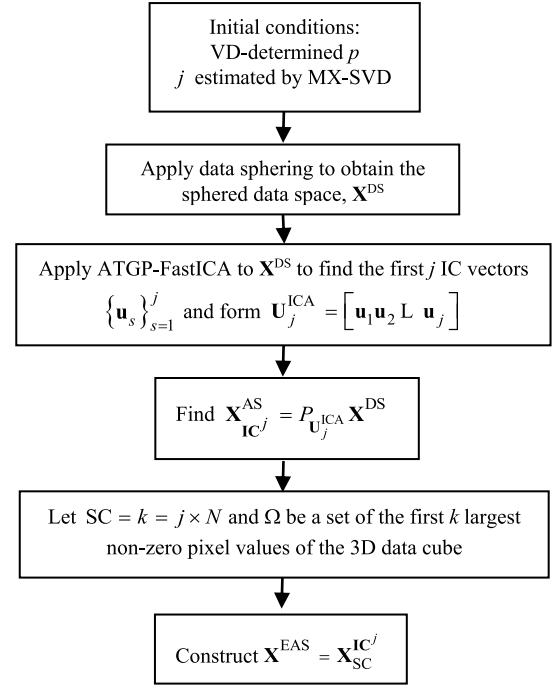


Fig. 2. Flowchart of constructing EAS.

MX-SVD

1. Initialization: Let p be estimated by VD and $j = 1$, i.e., $\mathbf{T}_1 = \emptyset$.
2. Use SVD to find the first p principal left singular vectors of the data matrix $\mathbf{X} = [\mathbf{r}_1 \mathbf{r}_2 \dots \mathbf{r}_N]$, denoted by $\mathbf{s}_1^{(0)}, \mathbf{s}_2^{(0)}, \dots, \mathbf{s}_p^{(0)}$ and let $\mathbf{B}_p = [\mathbf{s}_1^{(0)} \mathbf{s}_2^{(0)} \dots \mathbf{s}_p^{(0)}]$.
3. Let $j \leftarrow j + 1$. Find $\mathbf{t}_j = \arg\left\{\max_{\mathbf{r} \in \{\mathbf{r}_i\}_{i=1}^N} \left[\|\mathbf{P}_{[\mathbf{B}_{p-(j-1)} \mathbf{T}_{j-1}]}^\perp \mathbf{r}\|\right]\right\}$ and form $\mathbf{T}_j = [\mathbf{t}_1 \mathbf{t}_2 \dots \mathbf{t}_j]$.
4. Use SVD to find the first $p - j$ principal left singular vectors, denoted by $\mathbf{s}_1^{(j)}, \mathbf{s}_2^{(j)}, \dots, \mathbf{s}_{p-j}^{(j)}$. Let $\mathbf{B}_{p-j} = [\mathbf{s}_1^{(j)} \mathbf{s}_2^{(j)} \dots \mathbf{s}_{p-j}^{(j)}]$.
5. Repeat steps 3–4 until $j = p$. Continue.
6. At this stage, calculate $j^{\text{MX-SVD}}$ by

$$j^{\text{MX-SVD}} \equiv \arg\left\{\min_{1 \leq j \leq p} \|\mathbf{t}_j\|\right\}$$

be sparse, and SC allows \mathbf{X}^{AS} to retain only the first k largest nonzero samples in \mathbf{X}^{AS} by setting the rest of the data samples to zeros, where k can be obtained by setting $\text{SC} = k = j \times N$ according to Zhang et al. [70]. Details of exploring the SC concept can be found in [34]. A step-by-step construction of EAS is described in Fig. 2.

VII. DETECTION MEASURES FOR AD

As noted in Section IV, BKG is a major issue in anomaly detection. It cannot be simply considered as noise in communications but rather is an unwanted signal source. In this case, the binary hypothesis testing (1) can be interpreted as

EAS

1. Initial conditions: j is given.
2. Apply ATGP-FastICA to \mathbf{X}^{DS} to find the first j IC vectors $\{\mathbf{u}_s\}_{s=1}^j$ and form $\mathbf{U}_j^{\text{ICA}} = [\mathbf{u}_1 \mathbf{u}_2 \cdots \mathbf{u}_j]$.
3. Find $\mathbf{X}_{\text{IC}^j}^{\text{AS}} = P_{\text{U}_j^{\text{ICA}}} \mathbf{X}^{\text{DS}}$ where

$$P_{\text{U}_j^{\text{ICA}}} = \mathbf{U}_j^{\text{ICA}} \left((\mathbf{U}_j^{\text{ICA}})^T \mathbf{U}_j^{\text{ICA}} \right)^{-1} (\mathbf{U}_j^{\text{ICA}})^T. \quad (21)$$

4. Let $\text{SC} = k = j \times N$ and Ω be a set of the first k largest non-zero pixel values of the 3D data cube, $\mathbf{X}_{\text{IC}^j}^{\text{AS}}$.
5. Construct an SC-constrained IC sparse space of $\mathbf{X}_{\text{IC}^j}^{\text{AS}}$, denoted by $\mathbf{X}_{\text{IC}^j}^{\text{EAS}}$, which is made up of the first k largest nonzero pixel samples in $\mathbf{X}_{\text{IC}^j}^{\text{AS}}$ in \mathbf{W} .

the following anomaly-BKG detection problem:

$$\begin{aligned} H_0 : \mathbf{r} &\approx \text{BKG} \\ \text{versus} \\ H_1 : \mathbf{r} &\approx \text{anomlay} + \text{BKG} \end{aligned} \quad (97)$$

as a two-class binary classification problem where BKG and anomaly are specified by two distinct classes. By virtue of (97), the detection of BKG is no longer detection of noise but rather is as important as the detection of anomalies. Therefore, P_D is specifically described by ADP, and P_{ADP} , and P_F used in (1) can be now modified as $1-P_F$ to be considered as BKG P_D (BDP). Now, the 3D ROC analysis-based evaluation tool developed in [35] plots a 3D ROC curve as a function of three parameters, P_D , P_F , and τ , where τ is used as a threshold for (97) to determine whether a data sample is BKG or anomalous target pixel. Using such a 3D ROC curve, three 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) can be generated along with their respective AUC values, denoted by $\text{AUC}_{(D,F)}$, $\text{AUC}_{(D,\tau)}$, and $\text{AUC}_{(F,\tau)}$, to define the following detection measures.

A. Effectiveness of a Detector

$\text{AUC}_{(D,F)}$ is the most widely used criterion in signal detection. As noted in [35], it is specifically designed to evaluate the effectiveness of a detector but not TD and BS since P_D and P_F work jointly using the same threshold value τ but cannot work alone independently

$$0 \leq \text{AUC}_{(D,F)} \leq 1. \quad (98)$$

B. Anomaly Detection Probability

The AUC value of a 2D ROC curve of (P_D, τ) is commonly used to measure the effectiveness of a detector. It can be also used to evaluate how effective an AD is. That is, the higher the value of $\text{AUC}_{(D,\tau)}$, the better the AD. In this case, we can define a measure, called ADP of a detector as AUC_{ADP} by

$$0 \leq \text{AUC}_{\text{ADP}} \leq 1 \quad (99)$$

and

$$0 \leq \text{AUC}_{\text{JAD}} = \text{AUC}_{(D,F)} + \text{AUC}_{\text{ADP}} \leq 2. \quad (100)$$

C. Background Detection Probability

In TD, P_F is the false alarm probability and $1-P_F$ is the detection of noise. Thus, using $1-P_F$ to evaluate noise detection performance does not make sense. However, this is not true from a perspective medical diagnosis where $1-P_F$ is defined as specificity probability or true negative probability. Also, from the binary hypothesis formulation for AD specified by (97), P_F represents the probability of BKG data samples falsely detected as anomalies. With this interpretation, $1-P_F$ is indeed the probability of correctly detecting BKG samples. Thus, by virtue of the AUC value of 2D ROC curves of (P_F, τ) , $\text{AUC}_{(F,\tau)}$, we can define BDP, AUC_{BDP} , as a counterpart of ADP in (99) by

$$0 \leq \text{AUC}_{\text{BDP}} = 1 - \text{AUC}_{(F,\tau)} \leq 1 \quad (101)$$

and

$$0 \leq \text{AUC}_{\text{JBS}} = \text{AUC}_{(D,F)} + \text{AUC}_{\text{BDP}} \leq 2. \quad (102)$$

D. Joint Anomaly Detectability With Background Suppressibility

A 2D ROC curve of (P_D, P_F) has been widely used to measure the effectiveness of a detector [1]. As pointed out in Sections II and III, these P_D and P_F are actually calculated by the same threshold τ for their joint performance without accounting for ADP and BDP separately. In order to factor $\text{AUC}_{(F,\tau)}$ into $\text{AUC}_{(D,\tau)}$, we need to define the joint ADBS of a detector.

1) *AD in BKG (AD-BS)*: The first ADBS measure is to consider AD and BS working separately with different values of the threshold τ as an independent parameter. Since P_{FA} is caused by the probability of misdetecting BKG as an anomaly, this error probability should be subtracted from P_{AD} . To take care of this effect, a measure, factoring P_{FA} into P_{AD} , to be called ADBS, can be defined as AUC_{ADBS} by

$$-1 \leq \text{AUC}_{\text{ADBS}} = \text{AUC}_{(D,\tau)} - \text{AUC}_{(F,\tau)} \leq 1. \quad (103)$$

2) *Overall Anomaly Detection Probability*: Another important ADBS measure is derived from a classification point of view. It is called OADP defined by

$$0 \leq \text{AUC}_{\text{OADP}} = \text{AUC}_{\text{ADP}} + \text{AUC}_{\text{BDP}} \leq 2 \quad (104)$$

where AUC_{ADP} in (99) and AUC_{BDP} in (101) with $0 \leq \rho \leq 1$ being the probability to weigh AUC_{ADP} .

3) *Signal-to-Noise Probability Ratio*: The last but most effective detection measure comes from a similar idea that is widely used in communications/signal processing, SNR. It is called SNPR by

$$0 \leq \text{AUC}_{\text{SNPR}} = \frac{\text{AUC}_{(D,\tau)}}{\text{AUC}_{(F,\tau)}} \quad (105)$$

where P_F can be assumed to be caused by noise.

Most recently, Chang [71] provided a comprehensive analysis of 2D/3D ROC analysis for HAD with codes available codes on the website.¹

¹<https://wiki.umbc.edu/display/rssipl/10.+Download>

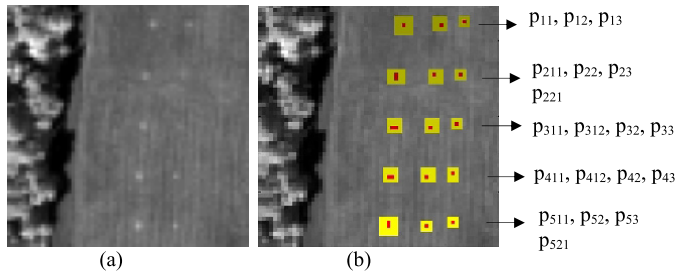


Fig. 3. (a) HYDICE 15-panel scene that contains 15 panels. (b) Ground-truth map of spatial locations of 19 R panel pixels.

VIII. IMAGE EXPERIMENTS

In order to demonstrate the performance of TAC ADs, three datasets were used for experiments.

A. HYDICE 15-Panel Scene

An airborne hyperspectral digital imagery collection experiment (HYDICE) scene is shown in Fig. 3. It was collected in August 1995 from a flight altitude of 10 000 ft with a ground sampling distance of approximately 1.56 m. This scene has been studied extensively by many reports, such as [1], [2], and [40].

It has a total of 169 bands that were used for the experiments with low signal/high noise bands: bands 1–3 and bands 202–210 and water vapor absorption bands: bands 101–112 and bands 137–153 were removed. There are 15 square panels in Fig. 3(a) with three different sizes, $3\text{ m} \times 3\text{ m}$, $2\text{ m} \times 2\text{ m}$, and $1\text{ m} \times 1\text{ m}$, respectively. Due to the ground sampling distance of approximately 1.56 m, each of the panels in the first column except the first row contains two panel pixels highlighted by red, p_{211} and p_{221} in row 2, p_{311} and p_{312} in row 3, p_{411} and p_{412} in row 4, and p_{511} and p_{521} in row 5, as shown in Fig. 6. All the remaining 11 panels in Fig. 6 contain one single panel pixel for each panel also highlighted by red, p_{11} , p_{12} , and p_{13} in row 1, p_{22} and p_{23} in row 2, p_{32} and p_{33} in row 3, p_{42} and p_{43} in row 4, and p_{52} and p_{53} in row 5. Therefore, there are a total of 19 red (R) panel pixels. Fig. 3(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. 4(a), which is available and can be downloaded from the Remote Sensing Lab website.²

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. 4(b) along with the ground-truth map shown in Fig. 4(c) where 21 pixels were identified as anomalies, which were cars and roofs, because they had spectra that differ from BKG.

²<https://rslab.ut.ac.ir/data>

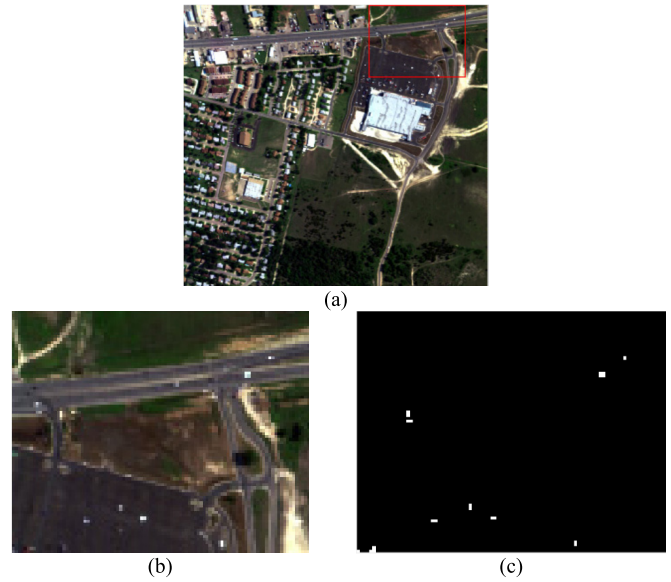


Fig. 4. HYDICE urban scene. (a) Pseudocolor image of the whole scene. (b) Pseudocolor image of the selected area. (c) Ground-truth map.

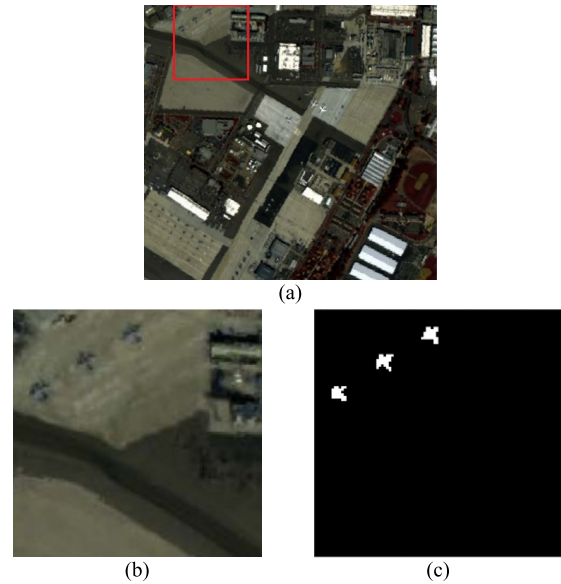


Fig. 5. San Diego Airport scene. (a) Pseudocolor image of the whole scene. (b) Pseudocolor image of the selected area. (c) Ground-truth map.

C. San Diego Airport Data

A third dataset used for experiments is available in [60] and was also an airborne visible/infrared imaging spectrometer (AVIRIS) acquired scene, which is a San Diego airport scene in CA, USA, with pseudocolor shown in Fig. 5(a).

It is an urban scene in which the main BKG materials are roof, shadow, and grass. It has a size of 400×400 pixels with a 3.5-m spatial resolution and 224 spectral channels in wavelengths ranging from 370 to 2510 nm. After removing the bands that correspond to the water absorption regions, low SNR, and bad bands (1–6, 33–35, 97, 107–113, 153–166, and 221–224), 189 available bands of the data were retained in the experiments. An area of 100×100 pixels at the upper

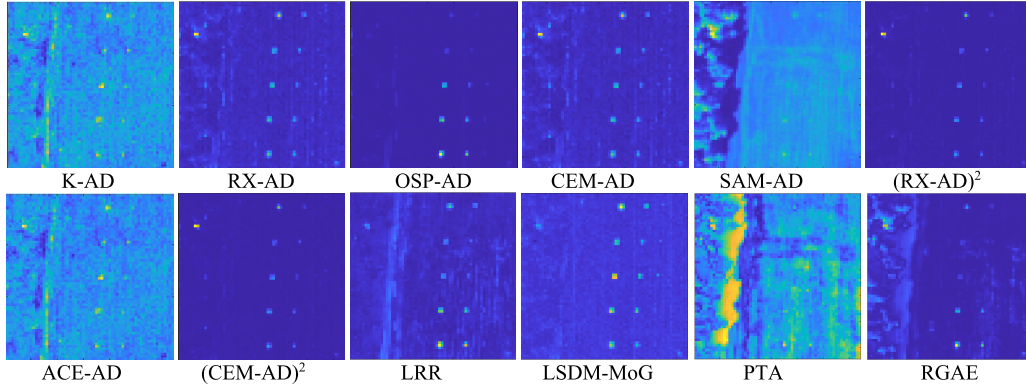


Fig. 6. Detection maps of various ADs in the original space for the HYDICE panel scene.

left corner of the scene was selected as the test image shown in Fig. 5(b) with its ground-truth map shown in Fig. 5(c).

IX. EXPERIMENTS AND DISCUSSION

In this section, we implemented eight TAC-AD listed in Table I with $\kappa = 1$, which are

$$\delta^{\text{LRT-AD}}(\mathbf{r}) = \delta^{\text{K-AD}}(\mathbf{r}) = \mathbf{r}^T \mathbf{K}^{-1} \mathbf{r} \quad (106)$$

$$\delta^{\text{RX-AD}}(\mathbf{r}) = (\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}) \quad (107)$$

$$\delta^{\text{OSP-AD}}(\mathbf{r}) = \mathbf{r}^T \mathbf{P}_U^\perp \mathbf{r} \quad (108)$$

$$\delta^{\text{CEM-AD}}(\mathbf{r}) = \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \quad (109)$$

$$\delta^{\text{SAM-AD}}(\mathbf{r}) = \|\mathbf{r}\|^2 \quad (110)$$

$$\delta^{\text{GDS-SNR-AD}}(\mathbf{r}) = (\delta^{\text{RX-AD}}(\mathbf{r}))^2 = ((\mathbf{r} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{r} - \boldsymbol{\mu}))^2 \quad (111)$$

$$\begin{aligned} \delta^{\text{GK-SNR-AD}}(\mathbf{r}) &= \delta^{\text{ACE-AD}}(\mathbf{r}) = \delta^{\text{ASD-AD}}(\mathbf{r}) \\ &= (\delta^{\text{K-AD}}(\mathbf{r}))^2 = (\mathbf{r}^T \mathbf{K}^{-1} \mathbf{r})^2 \end{aligned} \quad (112)$$

$$\delta^{\text{GR-SNR-AD}}(\mathbf{r}) = (\mathbf{r}^T \mathbf{R}^{-1} \mathbf{r})^2. \quad (113)$$

In addition to operating all the above eight ADs on the original data space, two constructions of EAS were also implemented, which are the IC-SC space in [34] specified by $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{SC}}^{\text{IC}^j}$ and the CDASC space in [33] specified by $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{CDASC}}^{\text{IC}^j}$, where the value of j was the same used in [33] and [34]. One note on the implementation of OSP-AD is particularly worth being mentioned. In the original data space \mathbf{X} , GoDec was implemented to produce the low-rank space \mathbf{L} , which was used as \mathbf{U} in OSP-AD. In this case, OSP-AD can be specifically referred to as OSP-GoDec-AD. In $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{SC}}^{\text{IC}^j}$, \mathbf{U} was made up of the remaining ICs after the first j ICs where OSP-AD can be specifically referred to as OSP-ICA-AD. In $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{CDASC}}^{\text{IC}^j}$, \mathbf{U} was specified by PC space produced by CDA. Thus, the resulting OSP-AD can be specifically referred to as OSP-CDA-AD.

Since RX-AD using CDA with SC (CDASC) [33] and CEM-AD using IC-SC [9] have been shown to perform at least or better than many other ADs, such as collaborative representation detector (CRD) [16], RPCA [28], and LRSMD-based MD-AD (LSMAD) [46], these approaches were not compared but rather referred to the abovementioned references. Instead, we selected four recently developed anomaly

detection approaches for our comparative analysis, each of which represents one particular category, such as LRR in [52] for low-rank and sparse representation, LSDM-MoG [61] for BKG estimation, PTA [60] for tensor approximation decomposition, and RGAE [51] for graph-based anomaly detection. Their implementation details along with parameter settings are available on the website. Thus, in the following experiments, a total of eight TAC-ADs [K-AD in (106), RX-AD in (107), CEM-AD in (108), OSP-AD in (109), SAM-AD in (110), GDS-SNR-AD in (111), GK-SNR-AD in (112), and GR-SNR-AD in (113)] were performed in two different spaces of EAS, $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{SC}}^{\text{IC}^j}$ and $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{CDASC}}^{\text{IC}^j}$, where their codes are available on websites,^{3,4} respectively, in comparison with the four selected AD methods: LRR, LSDM-MoG, PTA, and RGAE. It should be noted that DL network-based AD methods were not compared here because these methods are supervised and require training samples to train the networks, while the compared methods are unsupervised. Thus, it is inappropriate and unfair to compare training samples-based AD against AD without using training samples.

A. HYDICE Data

1) *Original Data Space*: Fig. 6 shows the detection maps produced by K-AD, RX-AD, CEM-AD, OSP-AD, SAM-AD, GDS-SNR-AD=(RX-AD)², GK-SNR-AD/ASD-AD/ACE-AD=(K-AD)², GR-SNR-AD=(CEM-AD)², LRR, LSDM-MoG, PTA, and RGAE. By visual inspection of Fig. 6, the best AD was LSDM-MoG, and the worst AD was PTA. Fig. 7 plots 3D ROC curves along with their corresponding 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) produced from the detection maps in Fig. 6. For quantitative comparison, Table II calculates the AUC values from Fig. 7, and eight detection measures derived in Section V were also used to evaluate the comparative performance among 12 ADs in Fig. 6, where the best results are boldfaced. In Table II, their computing times are also documented, where the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, required exceedingly high computational complexity, specifically, training-based

³<https://wiki.umbc.edu/download/attachments/26281078/EAS%20by%20Chein-I%20Chang.zip?api=v2>

⁴<https://wiki.umbc.edu/download/attachments/26281078/CDA%20by%20Chein-I%20Chang%20and%20Shuhan%20Chen.zip?api=v2>

TABLE II
AUC VALUES CALCULATED FROM FIG. 7 IN THE ORIGINAL DATA SPACE FOR THE HYDICE PANEL SCENE

| detector | $AUC_{(D,F)}$ | AUC_{ADP} | AUC_{BDP} | AUC_{JAD} | AUC_{JBS} | AUC_{ADBS} | AUC_{SNPR} | AUC_{OAPD} | time |
|-----------------------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|---------------|---------------|
| K-AD (106) | 0.9040 | 0.6029 | 0.6570 | 1.5069 | 1.5610 | 1.2600 | 1.7580 | 2.1639 | 0.0705 |
| RX-AD (107) | 0.9898 | 0.3574 | 0.9570 | 1.3472 | 1.9468 | 1.3145 | 8.3205 | 2.3042 | 0.0704 |
| OSP-AD (108) | 0.9901 | 0.3092 | 0.9963 | 1.2993 | 1.9865 | 1.3055 | 84.3372 | 2.2956 | 27.351 |
| CEM-AD (109) | 0.9900 | 0.3584 | 0.9565 | 1.3484 | 1.9465 | 1.3149 | 8.2341 | 2.3049 | 0.0708 |
| SAM-AD (110) | 0.9140 | 0.4517 | 0.7242 | 1.3656 | 1.6381 | 1.1758 | 1.6375 | 2.0898 | 0.0008 |
| (RX-AD) ² (111) | 0.9898 | 0.1866 | 0.9935 | 1.1764 | 1.9832 | 1.1801 | 28.5872 | 2.1698 | 0.0783 |
| ACE-AD (112) | 0.9040 | 0.5767 | 0.6849 | 1.4807 | 1.5889 | 1.2616 | 1.8304 | 2.1656 | 0.0765 |
| (CEM-AD) ² (113) | 0.9900 | 0.1872 | 0.9934 | 1.1772 | 1.9834 | 1.1806 | 28.3968 | 2.1706 | 0.0690 |
| LRR[52] | 0.9397 | 0.3564 | 0.9514 | 1.2961 | 1.8911 | 1.3079 | 7.3398 | 2.2475 | 9.3919 |
| LSDM-MoG [61] | 0.9932 | 0.4531 | 0.8742 | 1.4464 | 1.8675 | 1.3273 | 3.6026 | 2.3206 | 4.8089 |
| PTA [60] | 0.5954 | 0.4257 | 0.6391 | 1.0211 | 1.2344 | 1.0648 | 1.1795 | 1.6602 | 10.481 |
| RGAE [51] | 0.8169 | 0.2279 | 0.9672 | 1.0448 | 1.7841 | 1.1951 | 6.9518 | 2.0120 | 76.027 |

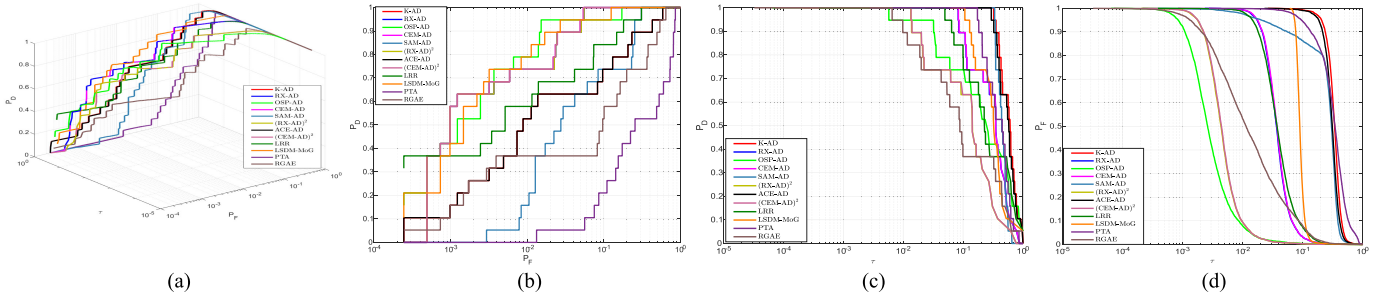


Fig. 7. 3D ROC curves along with three 2D ROC curves produced from the detection maps in Fig. 6. (a) 3D ROC curves. (b) 2D ROC curves of (P_D, P_F) . (c) 2D ROC curves of (P_D, τ) . (d) 2D ROC curves of (F_D, τ) .

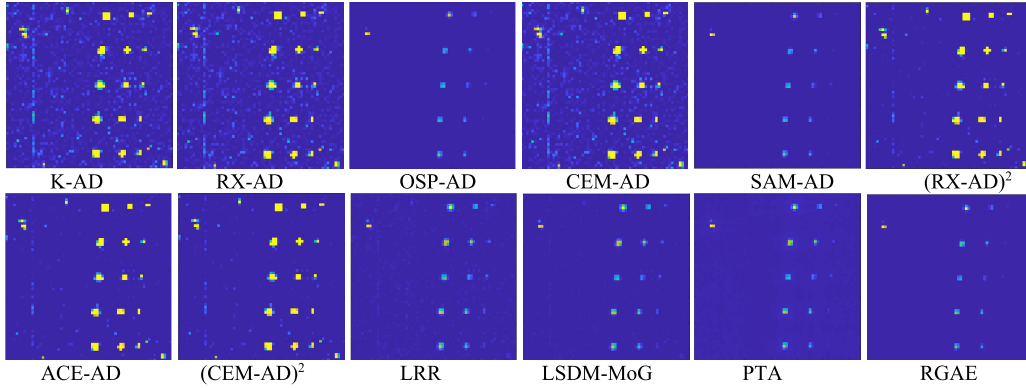


Fig. 8. Detection maps of various ADs in the IC-SC space for the HYDICE panel scene.

RGAE, which is expected. Based on the results in Table II, no AD can perform the best across the board in eight detection measures. In fact, for ADP, K-AD was the best as opposed to OSP-AD, which was the best in terms of BDP and JBS. Although LSDM-MoG produced the highest $AUC_{(D,F)}$ and was the best according to visual inspection in Fig. 6, it is neither the best in ADP nor BDP. These experiments simply demonstrated that solely relying on $AUC_{(D,F)}$ or visual inspection is not reliable and may be even misleading.

2) $EAS = IC-SC$ Space: In this section, the EAS was obtained by retaining the first j IC resulting from applying ATGP-ICA in [33] to the sphered data space $\tilde{\mathbf{X}}$ to yield $\mathbf{X}^{EAS} = \mathbf{X}_{SC}^{IC^j}$. We then operated the 12 ADs, K-AD, RX-AD, CEM-AD, OSP-AD, SAM-AD, (RX-AD)²,

(K-AD)², (CEM-AD)², LRR, LSDM-MoG, PTA, and RGAE in $\mathbf{X}^{EAS} = \mathbf{X}_{SC}^{IC^j}$. Fig. 8 shows their detection maps that had significantly improved the detection maps in Fig. 6. Interestingly, by visual inspection, the detection maps in Fig. 8 can be roughly divided into two categories: better ADP produced by a group of K-AD, RX-AD, CEM-AD, (RX-AD)², (K-AD)², and (CEM-AD)² and better BDP with very clean BS by a group of OSP-AD, SAM-AD, LRR, LSDM-MoG, PTA, and RGAE. The best AD was LSDM-MoG, and the worst AD was PTA. In order to conduct a quantitative analysis, Fig. 9 plots 3D ROC curves along with their corresponding 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) produced from the detection maps in Fig. 8. Table III calculates the AUC values from Fig. 9, and eight detection measures derived in Section V

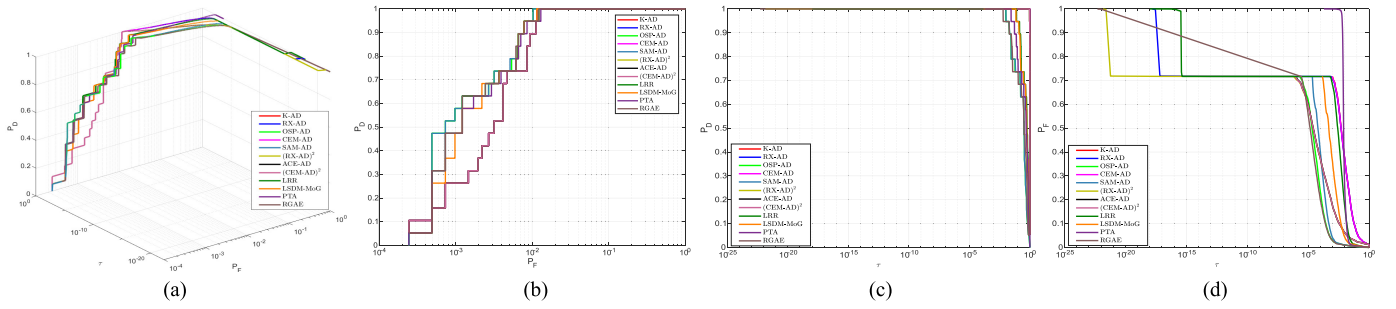


Fig. 9. 3D ROC curves along with three 2D ROC curves produced from the detection maps in Fig. 8. (a) 3D ROC curves. (b) 2D ROC curves of (P_D, P_F) . (c) 2D ROC curves of (P_D, τ) . (d) 2D ROC curves of (F_D, τ) .

TABLE III
AUC VALUES CALCULATED FROM FIG. 9 IN THE IC-SC SPACE FOR THE HYDICE PANEL SCENE

| detector | $AUC_{(D,F)}$ | AUC_{ADP} | AUC_{BDP} | AUC_{JAD} | AUC_{JBS} | AUC_{ADBS} | AUC_{SNPR} | AUC_{OADP} | time |
|--------------------|---------------|---------------|---------------|---------------|---------------|---------------|-----------------|---------------|---------------|
| K-AD (106) | 0.9958 | 0.9965 | 0.9629 | 1.9922 | 1.9587 | 1.9594 | 26.8567 | 2.9551 | 1.0757 |
| RX-AD (107) | 0.9958 | 0.9965 | 0.9631 | 1.9922 | 1.9589 | 1.9596 | 27.0057 | 2.9553 | 1.0694 |
| OSP-AD (108) | 0.9974 | 0.3237 | 0.9984 | 1.3211 | 1.9958 | 1.3221 | 206.0787 | 2.3195 | 1.0695 |
| CEM-AD (109) | 0.9958 | 0.9965 | 0.9629 | 1.9922 | 1.9587 | 1.9594 | 26.8845 | 2.9552 | 1.0702 |
| SAM-AD (110) | 0.9974 | 0.3241 | 0.9983 | 1.3215 | 1.9957 | 1.3225 | 196.4207 | 2.3199 | 0.9910 |
| $(RX-AD)^2$ (111) | 0.9958 | 0.9930 | 0.9818 | 1.9888 | 1.9776 | 1.9748 | 54.5035 | 2.9706 | 1.0603 |
| ACE-AD (112) | 0.9958 | 0.9930 | 0.9818 | 1.9888 | 1.9776 | 1.9748 | 54.5027 | 2.9706 | 1.0604 |
| $(CEM-AD)^2$ (113) | 0.9958 | 0.9930 | 0.9818 | 1.9888 | 1.9776 | 1.9748 | 54.5082 | 2.9706 | 1.0666 |
| LRR[52] | 0.9972 | 0.5310 | 0.9914 | 1.5282 | 1.9886 | 1.5224 | 61.6688 | 2.5196 | 2.5776 |
| LSDM-MoG [61] | 0.9971 | 0.5043 | 0.9951 | 1.5014 | 1.9922 | 1.4994 | 102.5741 | 2.4965 | 3.8916 |
| PTA [60] | 0.9970 | 0.4623 | 0.9884 | 1.4592 | 1.9854 | 1.4506 | 39.7652 | 2.4476 | 10.159 |
| RGAE [51] | 0.9972 | 0.3566 | 0.9984 | 1.3538 | 1.9956 | 1.3550 | 219.1824 | 2.3522 | 75.584 |

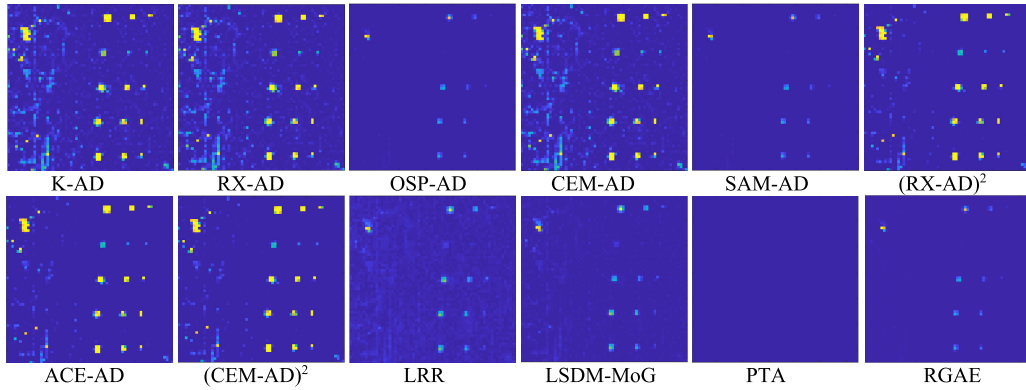


Fig. 10. Detection maps of various ADs in the CDASC space for the HYDICE panel scene.

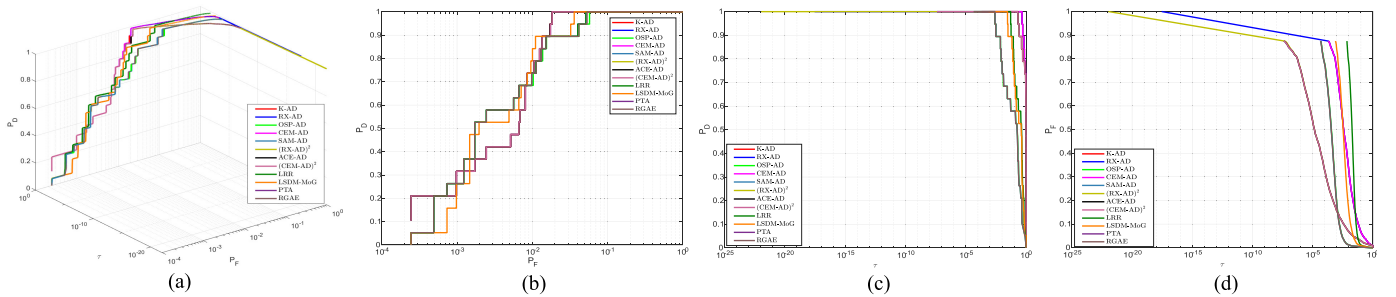


Fig. 11. 3D ROC curves along with three 2D ROC curves produced from the detection maps in Fig. 10. (a) 3D ROC curves. (b) 2D ROC curves of (P_D, P_F) . (c) 2D ROC curves of (P_D, τ) . (d) 2D ROC curves of (F_D, τ) .

were also used to evaluate the comparative performance among 12 ADs in Fig. 8 where the best results are boldfaced. In Table III, their computing times are also documented.

Like Table II, the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, also required exceedingly high computational complexity, specifically, training-based RGAE,

TABLE IV
AUC VALUES CALCULATED FROM FIG. 11 IN THE CDASC SPACE FOR THE HYDICE PANEL SCENE

| detector | AUC _(D,F) | AUC _{ADP} | AUC _{BDP} | AUC _{JAD} | AUC _{JBS} | AUC _{ADBS} | AUC _{SNPR} | AUC _{OADP} | time |
|-----------------------------|----------------------|--------------------|--------------------|--------------------|--------------------|---------------------|---------------------|---------------------|---------------|
| K-AD (106) | 0.9936 | 0.8952 | 0.9643 | 1.8888 | 1.9578 | 1.8595 | 25.0604 | 2.8530 | 1.3393 |
| RX-AD (107) | 0.9936 | 0.8949 | 0.9643 | 1.8885 | 1.9579 | 1.8592 | 25.0649 | 2.8528 | 1.3368 |
| OSP-AD (108) | 0.9908 | 0.2168 | 0.9981 | 1.2076 | 1.9889 | 1.2149 | 111.6061 | 2.2057 | 25.943 |
| CEM-AD (109) | 0.9936 | 0.8950 | 0.9643 | 1.8885 | 1.9578 | 1.8592 | 25.0593 | 2.8528 | 1.3641 |
| SAM-AD (110) | 0.9912 | 0.2122 | 0.9981 | 1.2035 | 1.9893 | 1.2103 | 111.5281 | 2.2016 | 1.2593 |
| (RX-AD) ² (111) | 0.9936 | 0.8301 | 0.9836 | 1.8236 | 1.9771 | 1.8136 | 50.5061 | 2.8072 | 1.3253 |
| ACE-AD (112) | 0.9936 | 0.8303 | 0.9836 | 1.8239 | 1.9771 | 1.8139 | 50.5124 | 2.8075 | 1.3283 |
| (CEM-AD) ² (113) | 0.9936 | 0.8301 | 0.9836 | 1.8237 | 1.9771 | 1.8137 | 50.5028 | 2.8072 | 1.3283 |
| LRR[52] | 0.9913 | 0.3771 | 0.9764 | 1.3684 | 1.9676 | 1.3535 | 15.9465 | 2.3447 | 4.6756 |
| LSDM-MoG [61] | 0.9927 | 0.3462 | 0.9914 | 1.3389 | 1.9841 | 1.3376 | 40.2136 | 2.3303 | 18.259 |
| PTA [60] | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| RGAE [51] | 0.9911 | 0.2124 | 0.9981 | 1.2035 | 1.9892 | 1.2105 | 111.5925 | 2.2016 | 82.662 |

TABLE V
AUC VALUES CALCULATED FROM FIG. 13 IN THE ORIGINAL DATA SPACE FOR THE HYDICE URBAN SCENE

| detector | AUC _(D,F) | AUC _{ADP} | AUC _{BDP} | AUC _{JAD} | AUC _{JBS} | AUC _{ADBS} | AUC _{SNPR} | AUC _{OADP} | time |
|-----------------------------|----------------------|--------------------|--------------------|--------------------|--------------------|---------------------|---------------------|---------------------|---------------|
| K-AD (106) | 0.9862 | 0.3126 | 0.9228 | 1.2988 | 1.9090 | 1.2354 | 4.0493 | 2.2216 | 0.2283 |
| RX-AD (107) | 0.9872 | 0.2575 | 0.9639 | 1.2448 | 1.9512 | 1.2215 | 7.1406 | 2.2087 | 0.2882 |
| OSP-AD (108) | 0.9877 | 0.3464 | 0.9756 | 1.3341 | 1.9633 | 1.3220 | 14.1883 | 2.3097 | 186.24 |
| CEM-AD (109) | 0.9872 | 0.2549 | 0.9641 | 1.2421 | 1.9513 | 1.2190 | 7.0965 | 2.2062 | 0.2974 |
| SAM-AD (110) | 0.6730 | 0.2121 | 0.8515 | 0.8851 | 1.5245 | 1.0636 | 1.4286 | 1.7366 | 0.0145 |
| (RX-AD) ² (111) | 0.9872 | 0.0943 | 0.9960 | 1.0816 | 1.9832 | 1.0903 | 23.3065 | 2.0775 | 0.2889 |
| ACE-AD (112) | 0.9862 | 0.1660 | 0.9754 | 1.1522 | 1.9616 | 1.1414 | 6.7463 | 2.1276 | 0.2721 |
| (CEM-AD) ² (113) | 0.9872 | 0.0927 | 0.996 | 1.0799 | 1.9831 | 1.0887 | 22.9858 | 2.0759 | 0.2633 |
| LRR[52] | 0.9787 | 0.4830 | 0.9542 | 1.4618 | 1.9330 | 1.4373 | 10.5584 | 2.4160 | 18.589 |
| LSDM-MoG [61] | 0.9801 | 0.3141 | 0.9014 | 1.2942 | 1.8815 | 1.2155 | 3.1868 | 2.1956 | 12.913 |
| PTA [60] | 0.8175 | 0.4474 | 0.7920 | 1.2649 | 1.6095 | 1.2394 | 2.1507 | 2.0569 | 23.019 |
| RGAE [51] | 0.8153 | 0.2771 | 0.9241 | 1.0924 | 1.7393 | 1.2012 | 3.6499 | 2.0164 | 127.15 |

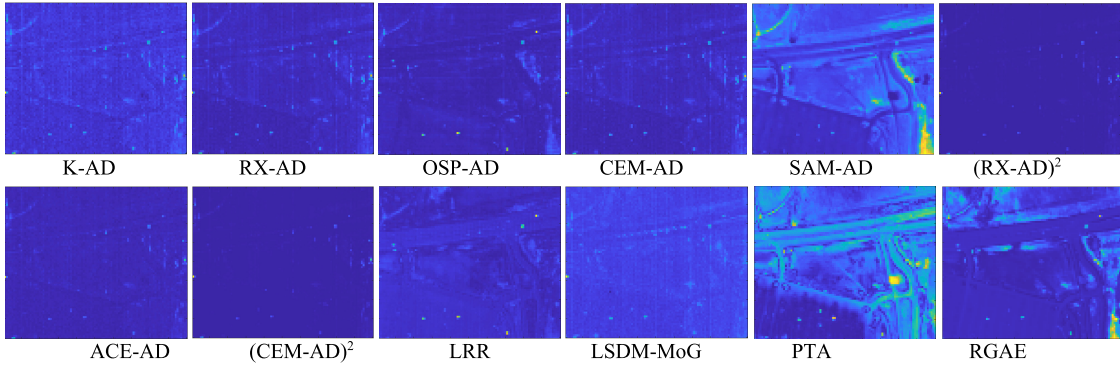


Fig. 12. Detection maps of various ADs in the IC-SC space for the HYDICE urban scene.

which is expected. From the results in Table III, OSP-AD was the best according to four out of eight detection measures. Surprisingly, LRR, LSDM-MoG, PTA, and RGAE were significantly improved by being operated in $\mathbf{X}_{SC}^{EAS} = \mathbf{X}_{SC}^{IC^j}$ but still could not compete against most of TAC-ADs. These experiments demonstrated an intriguing finding that recently developed BKG-modeled or BKG-processed anomaly detection methods may not necessarily perform better than simple TAC-ADs. This mainly because the residual space obtained by removing BKG-estimated or reconstructed by LRR, LSDM-MoG, PTA, and RGAE was not effective compared to the IC-SC space, which removes the first two orders of statistics including Gaussian-like BKG and Gaussian noise.

3) $EAS = CDASC$ Space: As an alternative to the use of $\mathbf{X}_{EAS} = \mathbf{X}_{SC}^{IC^j}$, we used another EAS obtained by retaining the first j IC resulting from applying CDA to the original data space to yield $\mathbf{X}_{EAS} = \mathbf{X}_{CDASC}^{IC^j}$. We then operated the 12 ADs, K-AD, RX-AD, CEM-AD, OSP-AD, SAM-AD, (RX-AD)², (K-AD)², (CEM-AD)², LRR, LSDM-MoG, PTA, and RGAE in $\mathbf{X}_{EAS} = \mathbf{X}_{CDASC}^{IC^j}$. Fig. 10 shows their detection maps that had improved the detection maps in Fig 6 but not as significantly as they did in $\mathbf{X}_{EAS} = \mathbf{X}_{SC}^{IC^j}$ shown in Fig. 8. By visual inspection, the detection maps in Fig. 10 seemed to be not as good as the detection maps in Fig. 8. Specifically, PTA was NA to CDASC since PC component space was removed by CDA, and PTA could not perform tensor

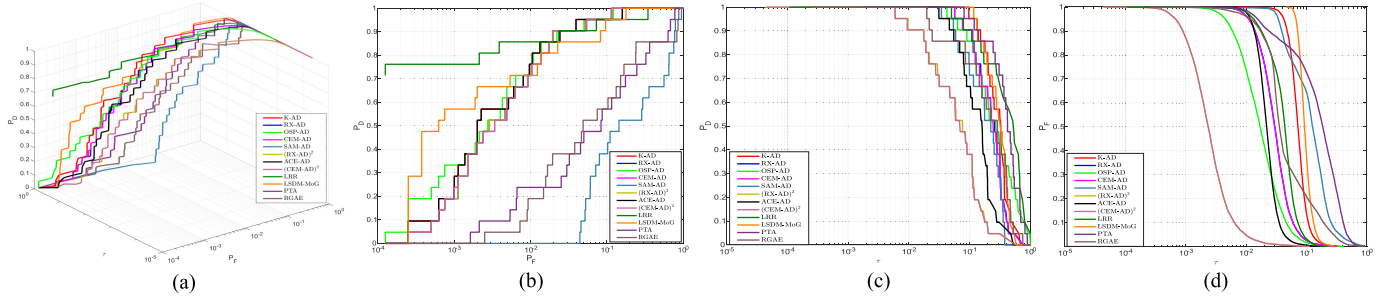


Fig. 13. 3D ROC curves along with three 2D ROC curves produced from the detection maps in Fig. 12. (a) 3D ROC curves. (b) 2D ROC curves of (P_D, P_F) . (c) 2D ROC curves of (P_D, τ) . (d) 2D ROC curves of (F_D, τ) .

TABLE VI
AUC VALUES CALCULATED FROM FIG. 15 IN THE IC-SC SPACE FOR THE HYDICE URBAN SCENE

| detector | AUC _(D,F) | AUC _{ADP} | AUC _{BDP} | AUC _{JAD} | AUC _{JBS} | AUC _{ADBS} | AUC _{SNPR} | AUC _{OADP} | time |
|-----------------------------|----------------------|--------------------|--------------------|--------------------|--------------------|---------------------|---------------------|---------------------|---------------|
| K-AD (106) | 0.9956 | 0.5785 | 0.9767 | 1.5741 | 1.9723 | 1.5552 | 24.797 | 2.5508 | 1.8706 |
| RX-AD (107) | 0.9956 | 0.5784 | 0.9767 | 1.5740 | 1.9723 | 1.5551 | 24.8078 | 2.5507 | 1.8861 |
| OSP-AD (108) | 0.9955 | 0.1461 | 0.9988 | 1.1415 | 1.9943 | 1.1449 | 126.487 | 2.1404 | 1.8823 |
| CEM-AD (109) | 0.9956 | 0.5784 | 0.9767 | 1.5740 | 1.9723 | 1.5551 | 24.8004 | 2.5507 | 1.8897 |
| SAM-AD (110) | 0.9954 | 0.1465 | 0.9987 | 1.1419 | 1.9942 | 1.1452 | 114.4002 | 2.1407 | 1.6229 |
| (RX-AD) ² (111) | 0.9956 | 0.3536 | 0.9945 | 1.3492 | 1.9901 | 1.3481 | 64.0116 | 2.3437 | 1.8602 |
| ACE-AD (112) | 0.9956 | 0.3537 | 0.9945 | 1.3494 | 1.9901 | 1.3482 | 64.005 | 2.3438 | 1.8665 |
| (CEM-AD) ² (113) | 0.9956 | 0.3536 | 0.9945 | 1.3493 | 1.9901 | 1.3481 | 64.0085 | 2.3437 | 1.8918 |
| LRR [52] | 0.9949 | 0.3309 | 0.9938 | 1.3258 | 1.9887 | 1.3247 | 53.0056 | 2.3196 | 6.0641 |
| LSDM-MoG [61] | 0.9951 | 0.2901 | 0.9954 | 1.2852 | 1.9904 | 1.2854 | 62.5288 | 2.2805 | 16.243 |
| PTA [60] | 0.9902 | 0.2482 | 0.9874 | 1.2384 | 1.9776 | 1.2356 | 19.6636 | 2.2258 | 24.989 |
| RGAE [51] | 0.9957 | 0.1618 | 0.9989 | 1.1575 | 1.9946 | 1.1607 | 148.6861 | 2.1564 | 129.87 |

decomposition on SC. Nevertheless, Fig. 10 still performed much better than Fig. 6. To conduct a quantitative analysis, Fig. 11 plots 3D ROC curves along with their corresponding 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) produced from the detection maps in Fig. 10. Table IV calculates the AUC values from Fig. 11, and eight detection measures were also used to evaluate the comparative performance among 12 ADs in Fig. 10 where the best results are boldfaced. In Table IV, their computing times are also documented. In analogy with Tables II and III, the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, required exceedingly high computational complexity, specifically, training-based RGAE, which is expected. Comparing Table III and Table IV, it is obvious to see that using $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{SC}}^{\text{IC}^j}$ is more effective than using $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{CDASC}}^{\text{IC}^j}$. Once again, using $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{CDASC}}^{\text{IC}^j}$ also improved LRR, LSDM-MoG, PTA, and RGAE operating in the original data space, but its improvement still could not compete against most of TAC-ADs. These experiments demonstrated that simple TAC-ADs could do a better job than complicated BKG-modeled anomaly detection methods that are generally developed to estimate, separate, and reconstruct BKG for removal prior to anomaly detection.

B. HYDICE Urban Scene

The same experiments conducted for the HYDICE panel data scene were also performed for HYDICE urban data scene.

1) *Original Data Space*: Fig. 12 shows the detection maps produced by K-AD, RX-AD, CEM-AD, OSP-AD, SAM-AD, (RX-AD)², (K-AD)², (CEM-AD)², LRR, LSDM-MoG, PTA,

and RGAE, where the best ones were OSP-AD and LRR, while the worst ones were SAM-AD and PTA by visual inspection. Fig. 13 plots 3D ROC curves along with their corresponding 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) produced from the detection maps in Fig. 12, and Table V calculates the AUC values from Fig. 13 for eight detection measures to evaluate the comparative performance among 12 ADs in Fig. 12 where the best results are boldfaced. In Table V, their computing times are also documented. As expected, the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, required exceedingly high computational complexity, specifically, training-based RGAE, which is much worse than the three others. Interestingly, the best ones were produced by (CEM-AD)², which had the best BDP and JBS values compared to LRR, which had the best JAD and ADBS values. Overall speaking, except SAM-AD using vector length, all other TAC-ADs performed comparably or even better than LSDM-MoG and PTA.

2) *EAS = IC-SC Space*: Fig. 14 shows the detection maps of the 12 ADs, K-AD, RX-AD, CEM-AD, OSP-AD, SAM-AD, (RX-AD)², (K-AD)², (CEM-AD)², LRR, LSDM-MoG, PTA, and RGAE operated in $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{SC}}^{\text{IC}^j}$, where all ADs in Fig. 14 were improved to some extent. Specifically, those which did not work well in Fig. 12 had been significantly improved in Fig. 14. From Fig. 14, it is difficult to assess the performance by visual inspection. Fig. 15 plots 3D ROC curves along with their corresponding 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) produced from the detection maps in Fig. 14, and Table VI calculates the AUC values from

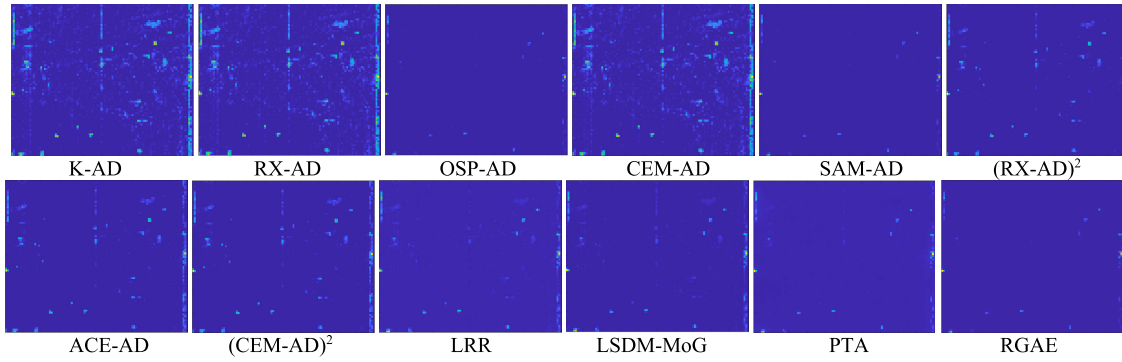


Fig. 14. Detection maps of various ADs using the IC-SC space for the HYDICE urban scene.

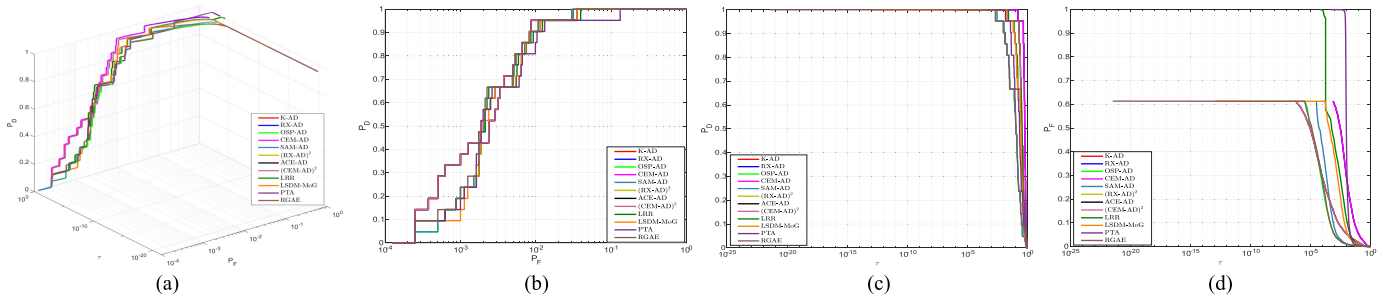


Fig. 15. 3D ROC curves along with three 2D ROC curves produced from the detection maps in Fig. 14. (a) 3D ROC curves. (b) 2D ROC curves of (P_D, P_F) . (c) 2D ROC curves of (P_D, τ) . (d) 2D ROC curves of (F_D, τ) .

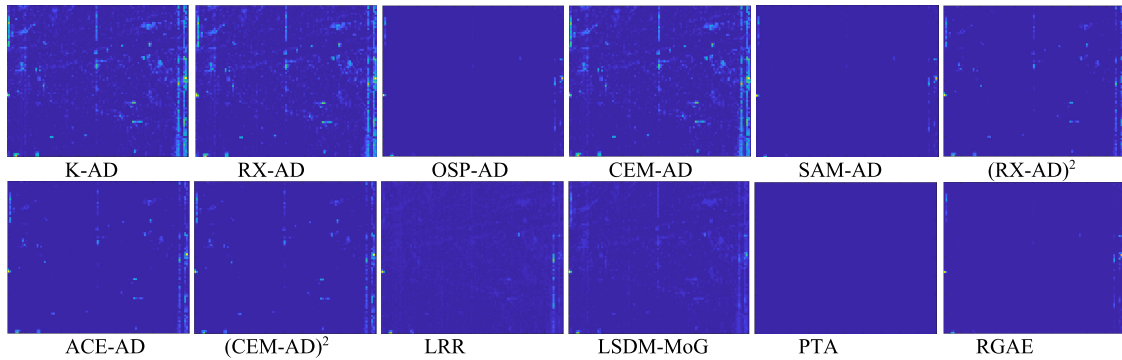


Fig. 16. Detection maps of various ADs in the CDASC space for the HYDICE urban scene.

Fig. 15 for eight detection measures to evaluate the comparative detection performance in Fig. 14 where the best results are boldfaced. In Table VI, their computing times are also documented. As expected, the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, required exceedingly high computational complexity, specifically, training-based RGAE, which is much worse than three others. According to the results in Table VI, K-AD performed the best in terms of ADP, JAD, ADBS, and OADP, while RGAE did the best in BDP, JBS, and SNPR in terms of four out of eight detection measures. It is very interesting to discover that K-AD and RGAE were indeed complemented each other. Surprisingly, LRR, LSDM-MoG, and PTA were significantly improved in BDP and JBS at the expense of ADP and JAD. Nevertheless, they still could not still compete against most of TAC-ADs.

3) $EAS = CDASC$: In analogy with experiments done for the IC-SC space, Fig. 16 shows the detection maps of the 12 ADs, K-AD, RX-AD, CEM-AD, OSP-AD, SAM-AD, $(RX-AD)^2$, $(K-AD)^2$, $(CEM-AD)^2$, LRR, LSDM-MoG, PTA, and RGAE operated in $\mathbf{X}^{EAS} = \mathbf{X}^{IC/CDASC}$, where all ADs in Fig. 12 were also improved. Specifically, those which did not work well in BDP in Fig. 12 had been significantly improved in Fig. 16. Fig. 17 plots 3D ROC curves along with their corresponding 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) produced from the detection maps in Fig. 16, and Table VII calculates the AUC values from Fig. 17 for eight detection measures to evaluate the comparative detection performance in Fig. 14 where the best results are boldfaced. In Table VII, their computing times are also documented. As expected, the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, required exceedingly high computational complexity,

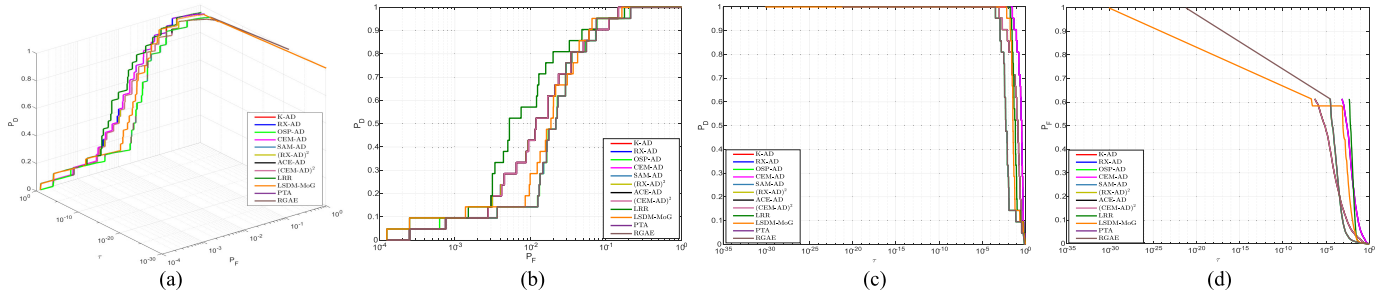


Fig. 17. 3D ROC curves along with three 2D ROC curves produced from the detection maps in Fig. 16. (a) 3D ROC curves. (b) 2D ROC curves of (P_D, P_F) . (c) 2D ROC curves of (P_D, τ) . (d) 2D ROC curves of (F_D, τ) .

TABLE VII
AUC VALUES CALCULATED FROM FIG. 17 IN THE CDASC SPACE FOR THE HYDICE URBAN SCENE

| detector | $AUC_{(D,F)}$ | AUC_{ADP} | AUC_{BDP} | AUC_{JAD} | AUC_{JBS} | AUC_{ADBS} | AUC_{SNPR} | AUC_{OADP} | time |
|--------------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|---------------|---------------|
| K-AD (106) | 0.9726 | 0.3507 | 0.9766 | 1.3233 | 1.9491 | 1.3273 | 14.9594 | 2.2999 | 2.0896 |
| RX-AD (107) | 0.9726 | 0.3506 | 0.9766 | 1.3232 | 1.9492 | 1.3272 | 14.9628 | 2.2998 | 2.0927 |
| OSP-AD (108) | 0.9662 | 0.0573 | 0.9984 | 1.0235 | 1.9647 | 1.0557 | 36.6631 | 2.0220 | 165.73 |
| CEM-AD (109) | 0.9726 | 0.3507 | 0.9766 | 1.3232 | 1.9491 | 1.3272 | 14.9599 | 2.2998 | 2.0871 |
| SAM-AD (110) | 0.9660 | 0.0572 | 0.9984 | 1.0232 | 1.9644 | 1.0556 | 36.2650 | 2.0216 | 1.8493 |
| $(RX-AD)^2$ (111) | 0.9726 | 0.1667 | 0.9947 | 1.1393 | 1.9673 | 1.1614 | 31.4018 | 2.1340 | 2.0967 |
| ACE-AD (112) | 0.9726 | 0.1668 | 0.9947 | 1.1394 | 1.9673 | 1.1615 | 31.4031 | 2.1341 | 2.0900 |
| $(CEM-AD)^2$ (113) | 0.9726 | 0.1668 | 0.9947 | 1.1393 | 1.9673 | 1.1614 | 31.4038 | 2.1340 | 2.0869 |
| LRR [52] | 0.9727 | 0.1324 | 0.9885 | 1.1051 | 1.9612 | 1.1209 | 11.4989 | 2.0936 | 8.8905 |
| LSDM-MoG [61] | 0.9689 | 0.1117 | 0.9926 | 1.0805 | 1.9615 | 1.1043 | 15.1804 | 2.0732 | 63.663 |
| PTA [60] | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| RGAE [51] | 0.9659 | 0.0572 | 0.9984 | 1.0231 | 1.9644 | 1.0556 | 36.4100 | 2.0216 | 124.80 |

specifically, training-based RGAE, which is much worse than three others. According to the results in Table VII, K-AD and RX-AD performed the best in terms of ADP, JAD, ADBS, and OADP, while RGAE did the best in BDP and SNPR. The same conclusions drawn from Table VI are also applied to Table VII. Nonetheless, comparing Table VI and Table VII, using $\mathbf{X}^{EAS} = \mathbf{X}_{CDASC}^{IC^j}$ is slightly less effective than using $\mathbf{X}^{EAS} = \mathbf{X}_{SC}^{IC^j}$.

C. San Diego Airport Scene

As noted in [34], technically speaking, the San Diego airport scene is not suitable for anomaly detection. This is because the three airplanes are so large and visible that they can be hardly viewed as anomalies. As expected, the conclusions drawn from this scene will be different from that of the previous HYDICE data scenes.

1) *Original Data Space*: Following the same experiments conducted for the previous two data scenes, Fig. 18 shows the detection maps of the 12 ADs, K-AD, RX-AD, CEM-AD, OSP-AD, SAM-AD, $(RX-AD)^2$, $(K-AD)^2$, $(CEM-AD)^2$, LRR, LSDM-MoG, PTA, and RGAE operated in the original data space \mathbf{X} , where only OSP-AD, SAM-AD, LRR, LSDM-MoG, PTA, and RGAE could detect three airplanes by visual inspection. It also clearly showed that PTA performed very poorly on BS even though it clearly detected the three airplanes. To conduct quantitative analysis, Fig. 19 plots 3D ROC curves along with their corresponding 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) produced from the detection maps in Fig. 18, and Table VIII calculates the AUC values from

Fig. 19 for eight detection measures to evaluate the quantitative detection performance in Fig. 18 where the best results are boldfaced. In Table VIII, their computing times are also documented. As expected, the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, required exceedingly high computational complexity, specifically, training-based RGAE, which is much worse than the three others. The results in Table VIII clearly showed that PTA was the best in terms of $AUC_{(D,F)}$, ADP, JAD, ADBS, and OADP, and had completely opposite results against the results concluded in the previous two data scenes. Also, PTA was the second worst in BDP next to SAM-AD, which was the worst performance across the board. However, it should be noted that SAM-AD is simply to measure the vector length of each data sample vector. As expected, it would not be effective in anomaly detection. Thus, if SAM-AD is excluded from the comparison, PTA was the worst in BDP, as also shown in Fig. 18. To the contrary, OSP-AD was the best in terms of JBS and SNPR. On the other hand, if PTA were used for target detection, it would have been the best target detector. However, when it was used as an AD, as shown in the experiments in this section, it became one of the worst ADs. These simple experiments demonstrated that a good target detector is not necessary to be an effective AD. In other words, RX-AD and CEM-AD working as ADs are supposed to detect anomalies. They did not pick up the three airplanes as anomalies. This made sense because the three airplanes are too large to be considered anomalies. This was further confirmed by their very high values of $AUC_{(D,F)}$, AUC_{ADP} , and AUC_{JAD} but very lower values of AUC_{BDP} in Table VIII. Thus, it is not surprised by the conclusions made

TABLE VIII
AUC VALUES CALCULATED FROM FIG. 19 IN THE ORIGINAL DATA SPACE FOR THE SAN DIEGO AIRPORT SCENE

| detector | $AUC_{(D,F)}$ | AUC_{ADP} | AUC_{BDP} | AUC_{JAD} | AUC_{JBS} | AUC_{ADBS} | AUC_{SNPR} | AUC_{OADP} | time |
|--------------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|---------------|---------------|
| K-AD (106) | 0.8137 | 0.1546 | 0.9008 | 0.9683 | 1.7146 | 1.0555 | 1.5592 | 1.8692 | 0.4249 |
| RX-AD (107) | 0.8334 | 0.0910 | 0.9547 | 0.9244 | 1.7881 | 1.0457 | 2.0101 | 1.8791 | 0.4561 |
| OSP-AD (108) | 0.9418 | 0.1365 | 0.9890 | 1.0783 | 1.9307 | 1.1255 | 12.3658 | 2.0673 | 336.03 |
| CEM-AD (109) | 0.8291 | 0.0940 | 0.9499 | 0.9231 | 1.7790 | 1.0439 | 1.8778 | 1.8730 | 0.4190 |
| SAM-AD (110) | 0.2743 | 0.1740 | 0.6644 | 0.4483 | 0.9387 | 0.8383 | 0.5183 | 1.1127 | 0.0019 |
| $(RX-AD)^2$ (111) | 0.8334 | 0.0244 | 0.9939 | 0.8578 | 1.8272 | 1.0183 | 3.9727 | 1.8516 | 0.4602 |
| ACE-AD (112) | 0.8137 | 0.0434 | 0.9824 | 0.8572 | 1.7962 | 1.0259 | 2.4732 | 1.8396 | 0.4082 |
| $(CEM-AD)^2$ (113) | 0.8291 | 0.0244 | 0.9935 | 0.8534 | 1.8226 | 1.0179 | 3.7460 | 1.8469 | 0.4191 |
| LRR[52] | 0.8749 | 0.2501 | 0.9275 | 1.1249 | 1.8023 | 1.1776 | 3.4482 | 2.0524 | 21.303 |
| LSDM-MoG [61] | 0.8385 | 0.2300 | 0.8290 | 1.0685 | 1.6675 | 1.0590 | 1.3449 | 1.8975 | 7.6864 |
| PTA [60] | 0.9846 | 0.6908 | 0.8062 | 1.6754 | 1.7908 | 1.4969 | 3.5640 | 2.4816 | 28.288 |
| RGAE [51] | 0.8984 | 0.1523 | 0.9749 | 1.0507 | 1.8733 | 1.1273 | 6.0812 | 2.0256 | 154.08 |

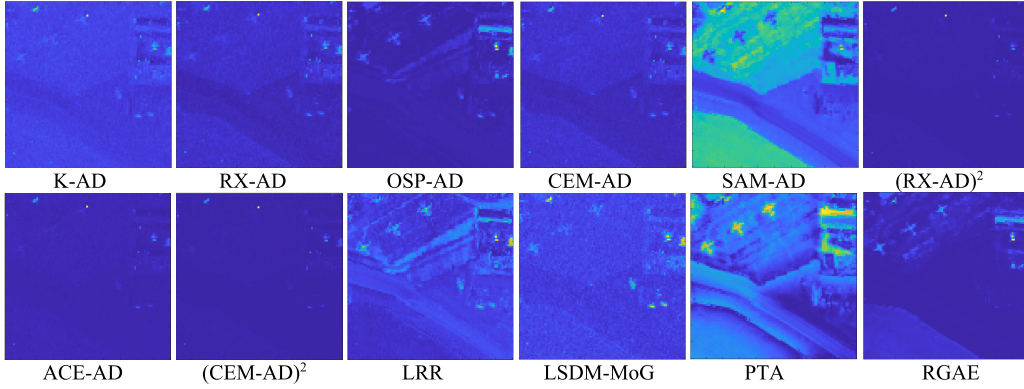


Fig. 18. Detection maps of various ADs in the original data space for the San Diego airport scene.

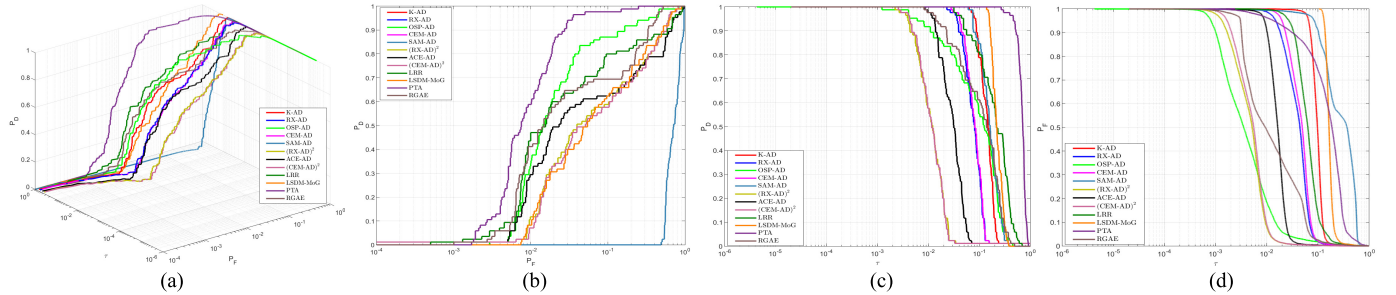


Fig. 19. 3D ROC curves along with three 2D ROC curves produced from the detection maps in Fig. 18. (a) 3D ROC curves. (b) 2D ROC curves of (P_D, P_F) . (c) 2D ROC curves of (P_D, τ) . (d) 2D ROC curves of (P_F, τ) .

in [46], [50], and [57] to claim that these detectors performed well. What was missing in these works was that they did not include AUC_{BDP} to compare their BS, which is BDP. Thus, using high TD to imply high anomaly detectability is misleading.

2) $EAS = IC-SC$ Space: It should be noted that the IC-SC space, $\mathbf{X}^{EAS} = \mathbf{X}_{SC}^{IC/}$, is particularly constructed for anomaly detection by DS, which removes the first and second orders of data statistics, while using ICA to retain anomalies. As mentioned in the beginning, the airplanes are too large to be considered as anomalies but rather as visible targets. Thus, by treating these airplanes as anomalies to perform anomaly detection in $\mathbf{X}^{EAS} = \mathbf{X}_{SC}^{IC/}$, it is expected that the performance will be worse than performing target detection in the original data space. The following experiments confirmed this conjecture. Fig. 20 shows the detection maps of the 12 ADs, K-AD, RX-AD, CEM-AD, OSP-AD, SAM-AD, $(RX-AD)^2$, $(K-AD)^2$, $(CEM-AD)^2$, LRR, LSDM-MoG, PTA,

and RGAE operated in $\mathbf{X}^{EAS} = \mathbf{X}_{SC}^{IC/}$, where only K-AD, RX-AD, and CEM-AD could barely detect three airplanes by visual inspection. It also clearly showed that most BKG in the detection maps in Fig. 20 was significantly suppressed in $\mathbf{X}^{EAS} = \mathbf{X}_{SC}^{IC/}$. Fig. 21 plots 3D ROC curves along with their corresponding 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) produced from the detection maps in Fig. 20, and Table IX calculates the AUC values from Fig. 21 for eight detection measures to evaluate the comparative detection performance in Fig. 20 where the best results are boldfaced. In Table IX, their computing times are also documented. As expected, the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, required exceedingly high computational complexity, specifically, training-based RGAE, which is much worse than three others. As anticipated, the results in Table IX were worse than that in Table VIII. In particular, ADP, JAD, and OADP that correspond to TD were much worse than their counterparts in Table VIII.

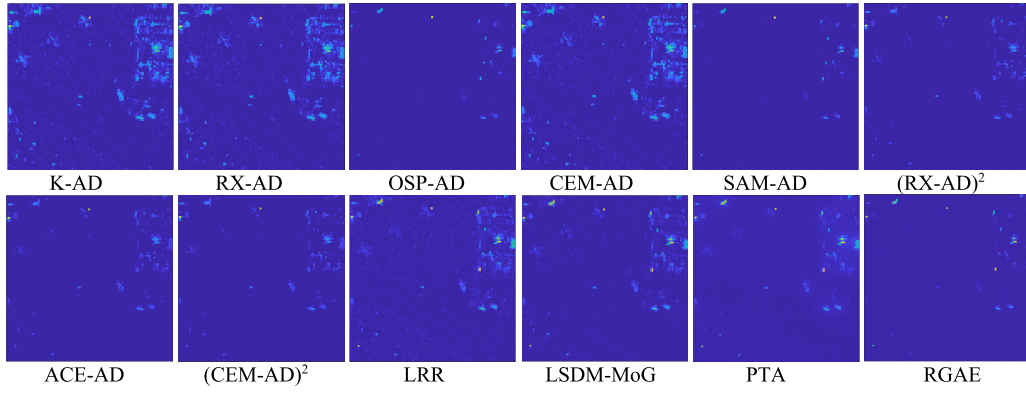


Fig. 20. Detection maps of various ADs in the ICASC space for the San Diego airport scene.

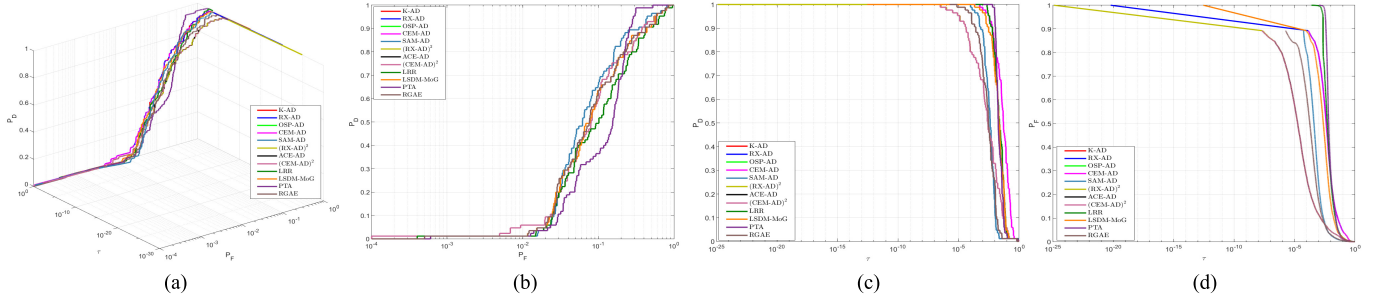


Fig. 21. 3D ROC curves along with three 2D ROC curves produced from the detection maps in Fig. 20. (a) 3D ROC curves. (b) 2D ROC curves of (P_D, P_F) . (c) 2D ROC curves of (P_D, τ) . (d) 2D ROC curves of (P_F, τ) .

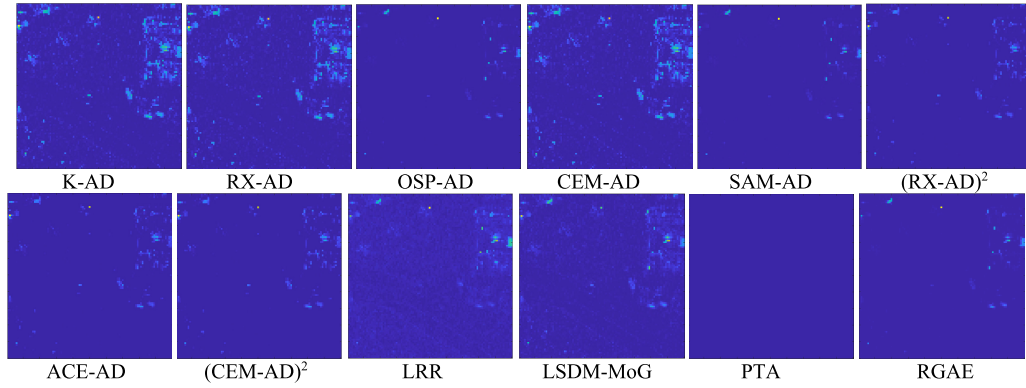


Fig. 22. Detection maps of various ADs in the CDASC space for the HYDICE San Diego airport scene.

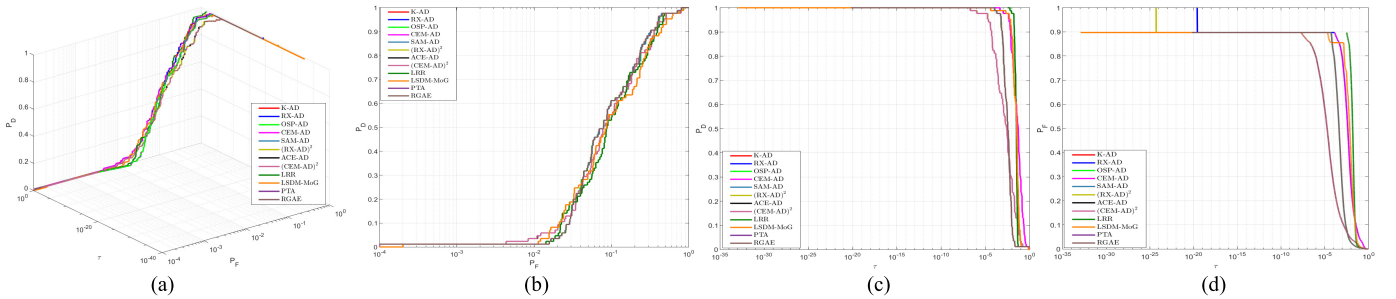


Fig. 23. 3D ROC curves along with three 2D ROC curves produced from the detection maps in Fig. 22. (a) 3D ROC curves. (b) 2D ROC curves of (P_D, P_F) . (c) 2D ROC curves of (P_D, τ) . (d) 2D ROC curves of (P_F, τ) .

The above experiments also demonstrated a crucial fact that $\mathbf{X}^{\text{EAS}} = \mathbf{X}_{\text{SC}}^{\text{IC}}$ is only effective for anomaly detection not for target detection, as also noted in [34]. For those who are interested in finding more details, we encourage them to compare the results presented in this section to the results in [18, Sec. IX.B] to see the significant and crucial differences between target detection and anomaly detection.

3) $\text{EAS} = \text{CDASC}$: Like the IC-SC space, CDASC removes the PC space from the original data space while using ICA to produce ICs. Thus, the same issues arising in the IC-SC space also occur in the CDASC space. To see this, the same experiment conducted for the IC-SC space was repeated for the CDASC space. Fig. 20 shows the detection maps of the 12 ADs, K-AD, RX-AD, CEM-AD, OSP-AD,

TABLE IX
AUC VALUES CALCULATED FROM FIG. 21 IN THE IC-SC FOR THE SAN DIEGO AIRPORT SCENE

| detector | AUC _(D,F) | AUC _{ADP} | AUC _{BDP} | AUC _{JAD} | AUC _{JBS} | AUC _{ADBS} | AUC _{SNPR} | AUC _{OADP} | time |
|-----------------------------|----------------------|--------------------|--------------------|--------------------|--------------------|---------------------|---------------------|---------------------|---------------|
| K-AD (106) | 0.8523 | 0.0962 | 0.9802 | 0.9485 | 1.8326 | 1.0764 | 4.8666 | 1.9288 | 2.4515 |
| RX-AD (107) | 0.8527 | 0.0962 | 0.9802 | 0.9489 | 1.8329 | 1.0765 | 4.8706 | 1.9291 | 2.4745 |
| OSP-AD (108) | 0.8763 | 0.0170 | 0.9972 | 0.8934 | 1.8735 | 1.0142 | 6.0112 | 1.8905 | 2.5186 |
| CEM-AD (109) | 0.8523 | 0.0962 | 0.9802 | 0.9485 | 1.8326 | 1.0764 | 4.8679 | 1.9288 | 2.4919 |
| SAM-AD (110) | 0.8763 | 0.0170 | 0.9972 | 0.8934 | 1.8735 | 1.0142 | 6.0112 | 1.8905 | 2.1012 |
| (RX-AD) ² (111) | 0.8527 | 0.0255 | 0.9964 | 0.8782 | 1.8491 | 1.0220 | 7.1401 | 1.8746 | 2.5392 |
| ACE-AD (112) | 0.8523 | 0.0255 | 0.9964 | 0.8779 | 1.8488 | 1.0219 | 7.1357 | 1.8743 | 2.4837 |
| (CEM-AD) ² (113) | 0.8523 | 0.0255 | 0.9964 | 0.8779 | 1.8488 | 1.0219 | 7.1371 | 1.8743 | 2.5144 |
| LRR[52] | 0.8214 | 0.0432 | 0.9859 | 0.8646 | 1.8072 | 1.0291 | 3.0590 | 1.8505 | 17.932 |
| LSDM-MoG [61] | 0.8496 | 0.0411 | 0.9895 | 0.8908 | 1.8391 | 1.0306 | 3.9180 | 1.8803 | 65.414 |
| PTA [60] | 0.8541 | 0.0363 | 0.9853 | 0.8904 | 1.8394 | 1.0216 | 2.4743 | 1.8757 | 3.0224 |
| RGAE [51] | 0.8507 | 0.0135 | 0.9969 | 0.8642 | 1.8475 | 1.0104 | 4.3387 | 1.8611 | 155.95 |

SAM-AD, (RX-AD)², (K-AD)², (CEM-AD)², LRR, LSDM-MoG, PTA, and RGAE operated in $\mathbf{X}^{\text{EAS}} = \mathbf{X}^{\text{IC}/\text{CDASC}}$, where the detection maps were very similar to those in Fig. 20, and only K-AD, RX-AD, and CEM-AD could barely detect three airplanes by visual inspection with most BKG in the detection maps in Fig. 22 being significantly suppressed in $\mathbf{X}^{\text{EAS}} = \mathbf{X}^{\text{IC}/\text{CDASC}}$. Fig. 23 plots 3D ROC curves along with their corresponding 2D ROC curves of (P_D, P_F) , (P_D, τ) , and (P_F, τ) produced from the detection maps in Fig. 22, and Table X calculates the AUC values from Fig. 23 for eight detection measures to evaluate the comparative detection performance in Fig. 22 where the best results are boldfaced. In Table X, their computing times are also documented. As expected, the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, required exceedingly high computational complexity, specifically, training-based RGAE, which is much worse than three others.

The similar conclusions drawn from Table IX were also applied to Table X. That is, $\mathbf{X}^{\text{EAS}} = \mathbf{X}^{\text{IC}/\text{CDASC}}$ is only effective for anomaly detection not for target detection. Comparing the results in Table IX and the results in Table X, $\mathbf{X}^{\text{EAS}} = \mathbf{X}^{\text{IC}/\text{SC}}$ worked better than $\mathbf{X}^{\text{EAS}} = \mathbf{X}^{\text{IC}/\text{CDASC}}$. However, there is an interesting finding between Tables IX and X. In Table IX, OSP-AD was the best in BS in terms of BDP and JBS and also the best in AUC_(D,F). RGAE was the second best but its results were very close and nearly the same. In Table X, the conclusion was reversed where RGAE performed the best in BS in terms of BDP and JBS and also the best in AUC_(D,F), and OSP-AD was the second best with nearly the same results as RGAE. This indicated that OSP-AD and RGAE performed very similarly in both $\mathbf{X}^{\text{EAS}} = \mathbf{X}^{\text{IC}/\text{SC}}$ and $\mathbf{X}^{\text{EAS}} = \mathbf{X}^{\text{IC}/\text{CDASC}}$, and can be considered as the same AD.

D. Discussion

The experiments conducted above provide hard evidence that a target detector can be indeed converted to an AD, TAC-AD, which can still work effectively, provided that its working data space is specified by EAS. This is because anomaly detection does not have prior target knowledge as a target detector does and must rely on BKG annihilation as a preprocessing step to bring out anomalies to light from BKG. EAS is particularly designed to meet this need. As shown

in the above experiments, TAC-AD operating on EAS can perform comparably or even better than many existing BKG-modeled-based AD methods due to the fact that anomalies cannot be learned from models. As an alternative to approaches that use models to characterize BKG, such as LRR, LSDM-MoG, PTA, and RGAE, many DL networks recently developed in the literature have used training samples to learn BKG. As a result, their performances are completely determined by how well BKG can be learned to find anomalies from the resulting BKG-suppressed or removed residual space. However, if BKG is very complicated and cannot be learned from training samples, DL networks will break down. What is worse is their very high computational complexity, as tabulated in Tables II–X, where their required computing times were extremely high compared to that required by TAC-AD. Thus, compared to BKG-modeled and BKG-learned DL network AD approaches, TAC-AD simply uses DS to clean up second-order characterized BKG, which includes mixtures of Gaussians-modeled BKG, and then uses ICA to produce an anomaly space, which can be further constrained by SC to construct EAS. Consequently, TAC-AD is simply an AD operating on EAS and does not need to learn BKG via models or training samples.

Many interesting findings are worth noting and summarized as follows.

- 1) Unlike a target detector that can work in the original data space \mathbf{X} , as shown in [18], a TAC-AD does not work for anomaly detection in the same space \mathbf{X} . This is because a target detector can offset the effect resulting from BKG interference via a matched filter using a specific target signature to be detected as its matched signature. Anomaly detection does not have such prior knowledge. In this case, EAS constructs a data space that removes all potential Gaussian-like BKG interference or noise that may be mistakenly considered anomalies.
- 2) As shown by the above experiments, all test ADs operating in the original data space could be improved in EAS. The level of significance of improvement is determined by the test AD. For example, if an AD performs poorly in the original space, it will be significantly improved in EAS. If an AD already performed very well, it could

TABLE X
AUC VALUES CALCULATED FROM FIG. 23 IN THE CDASC SPACE FOR THE SAN DIEGO AIRPORT SCENE

| detector | AUC _(D,F) | AUC _{ADP} | AUC _{BDP} | AUC _{JAD} | AUC _{JBS} | AUC _{ADBS} | AUC _{SNPR} | AUC _{OADP} | time |
|-----------------------------|----------------------|--------------------|--------------------|--------------------|--------------------|---------------------|---------------------|---------------------|---------------|
| K-AD (106) | 0.8490 | 0.0863 | 0.9804 | 0.9354 | 1.8295 | 1.0668 | 4.4126 | 1.9158 | 2.7092 |
| RX-AD (107) | 0.8492 | 0.0863 | 0.9805 | 0.9355 | 1.8297 | 1.0668 | 4.4165 | 1.9160 | 2.6919 |
| OSP-AD (108) | 0.8607 | 0.0162 | 0.9972 | 0.8769 | 1.8579 | 1.0134 | 5.7566 | 1.8741 | 314.90 |
| CEM-AD (109) | 0.8490 | 0.0863 | 0.9804 | 0.9353 | 1.8294 | 1.0668 | 4.4138 | 1.9157 | 2.7544 |
| SAM-AD (110) | 0.8613 | 0.0163 | 0.9972 | 0.8775 | 1.8584 | 1.0134 | 5.7275 | 1.8747 | 2.3058 |
| (RX-AD) ² (111) | 0.8492 | 0.0231 | 0.9965 | 0.8723 | 1.8457 | 1.0195 | 6.5350 | 1.8688 | 2.7167 |
| ACE-AD (112) | 0.849 | 0.0231 | 0.9965 | 0.8721 | 1.8455 | 1.0195 | 6.5307 | 1.8686 | 2.7761 |
| (CEM-AD) ² (113) | 0.8490 | 0.0231 | 0.9965 | 0.8721 | 1.8454 | 1.0195 | 6.5340 | 1.8685 | 2.6885 |
| LRR[52] | 0.8426 | 0.0418 | 0.9812 | 0.8844 | 1.8238 | 1.0230 | 2.2221 | 1.8656 | 16.558 |
| LSDM-MoG [61] | 0.8348 | 0.0431 | 0.9859 | 0.8779 | 1.8207 | 1.0290 | 3.0480 | 1.8638 | 59.083 |
| PTA [60] | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| RGAE [51] | 0.8622 | 0.0163 | 0.9972 | 0.8784 | 1.8594 | 1.0134 | 5.7366 | 1.8756 | 154.03 |

be still improved in EAS, but the room for improvement would not be great.

- 3) In all experiments, the computing times required for the four compared AD methods, LRR, LSDM-MoG, PTA, and RGAE, are extremely high compared to the other eight ADs, especially, training-based RGAE, which is far worse than the three methods because it needs an unsupervised training process to train AE.
- 4) Despite that SAM-AD did not work effectively in the original data space \mathbf{X} , it turned out to work very well when it is working sphered or whitened data space, such as DS-based TAC-ADs that include RX-AD and (RX-AD)², covariance matrix \mathbf{K} -whitened-based ADs that include LRT-AD = K-AD and ACE-AD = (K-AD)², and correlation matrix \mathbf{R} -whitened-based ADs that include CEM-AD and (CEM-AD)².
- 5) Most surprisingly, OSP-AD worked very well, particularly, in BS, JBS, and SNPR if the undesired target signature matrix \mathbf{U} is appropriately selected, such as OSP-GoDec-AD, which uses the GoDec-generated low-rank space \mathbf{L} as \mathbf{U} , OSP-ICA-AD, which uses the remaining ICs after the first j ICs as \mathbf{U} , and OSP-CDA-AD, which uses the PC space produced by CDA as \mathbf{U} .
- 6) According to conducted experiments, RGAE had a similar with slightly worse performance than OSP-AD. The advantage of OSP-AD over RGAE is that OSP-AD is much easier to be implemented than RGAE because there are no parameters for OSP-AD needed to be tuned.
- 7) As shown in the conducted experiments, all the derived TAC-ADs performed either comparably or even better than many existing ADs, including LRR, LSDM-MoG, PTA, and RGAE. This indicates that the statistical ADs, TAC-ADs, can perform more effectively in EAS than many model-based anomaly detection methods, specifically, when BKG is very complicated, and training samples or network learning cannot provide sufficient knowledge about BKG. On the contrary, EAS first makes use of DS to clean up the first two orders of data statistics, including Gaussian-like BKG and Gaussian noise, and then applies ICA to retain anomalies while using SC to remove non-Gaussian noises.
- 8) Most importantly, the best advantage of TAC-ADs is no parameters that are required to be tuned. In addition,

there is also no prior target knowledge, or the use of training samples is needed. The best of all is that TAC-AD is very easy to understand and implement without appealing to very sophisticated network structures.

- 9) Finally, the San Diego airport data experiments further confirmed the findings in [34] that this dataset can be used for target detection but may not be appropriate for anomaly detection due to the fact that the airplanes are too large to be considered anomalies. This evidence has been shown by K-AD, RX-AD, and CEM-AD, which could not detect the airplanes as anomalies. As a result, the conclusions of AD made on this dataset may be very likely misleading.

As final remarks, for those who are interested in comparative studies and analyses of other ADs, such as CRD, RPCA, and LSMAD, we refer to [29], [31], [32], and [33], where these ADs were shown to be inferior to RX-AD or CEMAD working in their appropriate defined spaces. In this article, we further showed that classic RX-AD and CEM-AD performed even better in EAS. Since all the TAC-ADs have very simple detector structures, they are very easy to implement compared to model- or DL-based networks, which must tune many parameters by training samples.

X. CONCLUSION

Target detection and anomaly detection are two completely separate tasks due to the fact that the former is active with required prior target knowledge, while the latter is passive without knowing any target knowledge. Thus, one cannot imply another. This article presents an interesting approach, which converts a target detector to its corresponding AD, to be called target-to-anomaly conversion-based AD (TAC-AD). Its key is to use a novel idea of DVT proposed in [24], which introduces a dummy variable as a UST to bridge the gap between a known target and an unknown anomaly. By virtue of DVT, many target detectors derived in [18] are readily converted to their corresponding TAC-ADs. However, due to its nature in characteristics derived from anomalies, their data statistics generally cannot be characterized by the first two orders of data statistics. In this case, what works for a target detector may not also work for its corresponding TAC-AD. To resolve this issue, the concept of EAS introduced in [34] plays a crucial role in making TAC-ADs work effectively.

Experiments showed that TAC-ADs using EAS perform at least as well or even better than many existing ADs. This article can be considered as a companion paper of [18] and [23], and offers a transition from target detection to anomaly detection.

A final remark is noteworthy. A potential application of TAC-AD to object detection [72] and infrared small target detection [73] is currently under investigation. Specifically, the traditional 2D ROC curve analysis can be replaced by 3D ROC analysis to evaluate objects and infrared targets. This leads to a posteriori 3D ROC curve-based detection measures for object detection and infrared small target detection.

ACKNOWLEDGMENT

The author would like to thank Dr. Shuhan Chen with Zhejiang University, China, for her help in implementing all experiments presented in this article.

REFERENCES

- [1] H. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY, USA: Springer-Verlag, 1991.
- [2] 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.
- [3] D. Manolakis, M. Pieper, E. Truslow, T. Cooley, M. Brueggeman, and S. Lipson, "The remarkable success of adaptive cosine estimator in hyperspectral target detection," *Proc. SPIE*, vol. 8743, May 2013, Art. no. 874302, doi: [10.1117/12.2015392](https://doi.org/10.1117/12.2015392).
- [4] D. Manolakis, G. Shaw, and N. Keshava, "Comparative analysis of hyperspectral adaptive matched filter detectors," *Proc. SPIE*, vol. 4049, Aug. 2000, Art. no. 410332, doi: [10.1117/12.410332](https://doi.org/10.1117/12.410332).
- [5] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001.
- [6] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
- [7] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.
- [8] 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.
- [9] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.
- [10] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-RX-AD and linear filter-based RX-AD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.
- [11] 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.
- [12] 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.
- [13] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2014.
- [14] 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. 658–664, Apr. 2013.
- [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] 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.
- [18] C.-I. Chang, "Hyperspectral target detection: Hypothesis testing, signal-to-noise ratio, and spectral angle theories," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5505223.
- [19] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–218, Jan. 1992.
- [20] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.
- [21] S. Kraut, L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.
- [22] C.-I. Chang, *Real-Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection*. New York, NY, USA: Springer, 2016.
- [23] C.-I. Chang, "Hyperspectral anomaly detection: A dual theory of hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5511720.
- [24] C.-I. Chang, "Constrained energy minimization anomaly detection for hyperspectral imagery via dummy variable trick," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517119.
- [25] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Ph.D. dissertation, Dept. Elect. Eng., Univ. Maryland Baltimore County, Baltimore, MD, USA, 1993.
- [26] 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, no. 1, pp. 64–76, Jan. 1997.
- [27] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA, USA: Kluwer, 2003.
- [28] 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.
- [29] 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, Mar. 2021.
- [30] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. 28th Int. Conf. Mach. Learn.*, Bellevue, WA, USA, 2011.
- [31] C.-I. Chang, H. Cao, and M. Song, "Orthogonal subspace projection target detector for hyperspectral anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4915–4932, 2021.
- [32] C.-I. Chang and J. Chen, "Hyperspectral anomaly detection by data sphering and sparsity density peaks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5526321.
- [33] S. Chen, C.-I. Chang, and X. Li, "Component decomposition analysis for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5516222.
- [34] C.-I. Chang, "Effective anomaly space for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5526624.
- [35] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, Jun. 2021.
- [36] T. M. Tu, C.-H. Chen, and C.-I. Chang, "A posteriori least squares orthogonal subspace projection approach to weak signature extraction and detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 127–139, Jan. 1997.
- [37] J. J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1045–1046, Jul. 1996.
- [38] 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.
- [39] 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.
- [40] C.-I. Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*. Hoboken, NJ, USA: Wiley, 2013.

- [41] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2538–2541, Sep. 1999.
- [42] B. Picinbono, "On deflection as a performance criterion in detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 3, pp. 1072–1081, Jul. 1995.
- [43] 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.
- [44] J. Zhong, W. Xie, Y. Li, J. Lei, and Q. Du, "Characterization of background-anomaly separability with generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 6017–6028, Jul. 2021.
- [45] W. Xie, S. Fan, J. Qu, X. Wu, Y. Lu, and Q. Du, "Spectral distribution-aware estimation network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.
- [46] J. Qu, Q. Du, Y. Li, L. Tian, and H. Xia, "Anomaly detection in hyperspectral imagery based on Gaussian mixture model," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9504–9517, Nov. 2021.
- [47] W. Xie, B. Liu, Y. Li, J. Lei, and Q. Du, "Autoencoder and adversarial-learning-based semisupervised background estimation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5416–5427, Aug. 2020.
- [48] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503314.
- [49] 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.
- [50] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 3123–3134, Dec. 2015.
- [51] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
- [52] 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.
- [53] H. Su, Z. Wu, A.-X. Zhu, and Q. Du, "Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 195–211, Nov. 2020.
- [54] W. Xie, X. Zhang, Y. Li, J. Lei, J. Li, and Q. Du, "Weakly supervised low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 3889–3900, Aug. 2021.
- [55] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1027–1063, 2009.
- [56] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018.
- [57] N. Vaswani, Y. Chi, and T. Bouwmans, "Rethinking PCA for modern data sets: Theory, algorithms, and applications," *Proc. IEEE*, vol. 106, no. 8, pp. 1274–1276, Aug. 2018.
- [58] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, Jan. 2017.
- [59] A. Sobral, T. Bouwmans, and E. Zahzah, "LRSLibrary: Low-rank and sparse tools for background modeling and subtraction in videos," in *Handbook on Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*, T. Bouwmans, N. Aybat, and E. Zahzah, Eds. Boca Raton, FL, USA: CRC Press, May 2016.
- [60] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022.
- [61] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of Gaussian for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4363–4372, Sep. 2020.
- [62] S. Feng, S. Tang, C. Zhao, and Y. Cui, "A hyperspectral anomaly detection method based on low-rank and sparse decomposition with density peak guided collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.
- [63] C.-I. Chang and J. Chen, "Orthogonal subspace projection using data sphering and low-rank and sparse matrix decomposition for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8704–8722, Oct. 2021.
- [64] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ, USA: Wiley, 2001.
- [65] A. Hyvärinen and E. Oja, "A fast fixed-point for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, 1997.
- [66] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
- [67] 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.
- [68] 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.
- [69] 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.
- [70] 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.
- [71] C.-I. Chang, "Comprehensive analysis of receiver operating characteristic (ROC) curves for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, early access.
- [72] Q. Wang, Y. Liu, Z. Xiong, and Y. Yuan, "Hybrid feature aligned network for salient object detection in optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5624915.
- [73] P. Yang, L. Dong, and W. Xu, "Infrared small maritime target detection based on integrated target saliency measure," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2369–2386, 2021.



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

He has been a Professor with the Department of Computer Science and Electrical Engineering with the University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA, since 1987. He has authored four books: *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Kluwer Academic Publishers, 2003), *Hyperspectral Data Processing: Algorithm Design and Analysis* (John Wiley & Sons, 2013), *Real Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection* (Springer, 2016), and *Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation* (Springer, 2017). He has also edited three books, *Recent Advances in Hyperspectral Signal and Image Processing* (2006), *Hyperspectral Data Exploitation: Theory and Applications* (John Wiley & Sons, 2007), and *Advances in Hyperspectral Image Processing Techniques* (Wiley, 2023), and co-edited with A. Plaza the book *High Performance Computing in Remote Sensing* (CRC Press, 2007). His research interests include hyperspectral image processing and medical imaging.

Dr. Chang is a fellow of the Society for Photo-optical Instrumentation Engineers (SPIE). He received the National Research Council Senior Research Associateship Awards from 2002 to 2003. He currently serves as an Associate Editor for *Remote Sensing* and *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*.