

# Multipixel Anomaly Detection With Unknown Patterns for Hyperspectral Imagery

Jun Liu<sup>1b</sup>, *Senior Member, IEEE*, Zengfu Hou, Wei Li<sup>1b</sup>, *Senior Member, IEEE*,  
Ran Tao<sup>1b</sup>, *Senior Member, IEEE*, Danilo Orlando, *Senior Member, IEEE*,  
and Hongbin Li<sup>1b</sup>, *Fellow, IEEE*

**Abstract**—In this article, anomaly detection is considered for hyperspectral imagery in the Gaussian background with an unknown covariance matrix. The anomaly to be detected occupies multiple pixels with an unknown pattern. Two adaptive detectors are proposed based on the generalized likelihood ratio test design procedure and *ad hoc* modification of it. Surprisingly, it turns out that the two proposed detectors are equivalent. Analytical expressions are derived for the probability of false alarm of the proposed detector, which exhibits a constant false alarm rate against the noise covariance matrix. Numerical examples using simulated data reveal how some system parameters (e.g., the background data size and pixel number) affect the performance of the proposed detector. Experiments are conducted on five real hyperspectral data sets, demonstrating that the proposed detector achieves better detection performance than its counterparts.

**Index Terms**—Anomaly detection (AD), constant false alarm rate, generalized likelihood ratio test (GLRT), hyperspectral imagery (HSI), multipixel target.

## I. INTRODUCTION

**S**IGNAL detection in hyperspectral imagery (HSI) has been a topic of long-standing interest due to its wide applications [1]–[5], such as cancer detection [6], man-made object surveillance [7]–[10], and environmental monitoring [11]–[14]. Hyperspectral imaging systems collect data into 3-D cubes with spatial and spectral dimensions, and information in such dimensions can be exploited to solve object or activity

detection problems. In HSI, the radiance of any observed material changes with wavelength, and a spectral vector with a large number of elements is formed for each pixel in a scene of interest. Hence, different materials can be distinguished by using their spectral features.

In HSI, anomalies are considered as outlier observations that are unusual with respect to their surroundings. Anomaly detection (AD) is to search for such outliers without knowing their spectral radiance measures [15]. It is worth noting that no *a priori* knowledge about spectral features is available in the AD. On the contrary, in target detection, the spectral signature of a specific material, which is known, is exploited. In this study, we restrict ourselves to the AD problem.

In the open literature, extensive investigations have been conducted on the AD problem, and many detectors have been designed under different models. One class of anomaly detectors is nonparametric, which does not make any assumption on the data distribution. In [16], a support vector data description (SVDD) detector was developed for the AD in HSI. It was a nonparametric data-driven method that provides a tight boundary around the local background by only using a small number of support vectors lying on the hypersphere surface. The pixels outside the hypersphere are identified as anomalies. In [17], a new nonparametric algorithm, named the collaborative representation detector (CRD), was proposed on the basis of the concept that anomalies cannot be approximated by the linear combination of neighboring pixels, while each pixel in the background can. Note that these nonparametric algorithms adaptively estimate the background without requiring *a priori* knowledge about the statistical properties of the data. Other nonparametric detectors can also be found in [18]–[23].

Another class of anomaly detectors relies on statistical assumptions made for the data at the design stage, which imposes specific distributions on HSI data. Based on different statistical models, various detectors were designed according to the hypothesis testing theory, including Mahalanobis distance-based detectors [24]–[26], subspace projection-based detectors [27], and fractional Fourier entropy-based detector [28]. In [24] and [25], the well-known Reed–Xiaoli detector (RXD) was introduced according to the principle of the generalized likelihood ratio test (GLRT), on the assumption that each pixel has a Gaussian distribution. It has been found that the resulting GLRT is exactly the Mahalanobis distance

Manuscript received November 14, 2020; revised February 22, 2021; accepted March 31, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1801105; in part by the National Natural Science Foundation of China under Grant 61871469, Grant 61922013, and Grant U1833203; and in part by the Youth Innovation Promotion Association CAS under Grant CX2100060053. (Corresponding author: Wei Li.)

Jun Liu is with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: junliu@ustc.edu.cn).

Zengfu Hou, Wei Li, and Ran Tao are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China, and also with Beijing Key Lab of Fractional Signals and Systems, Beijing Institute of Technology, Beijing 100081, China (e-mail: zengf.hou@bit.edu.cn; liwei089@ieee.org; rantao@bit.edu.cn).

Danilo Orlando is with the Università degli Studi Niccolò Cusano, 00166 Rome, Italy (e-mail: danilo.orlando@unicusano.it).

Hongbin Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: hongbin.li@stevens.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TNNLS.2021.3071026>.

Digital Object Identifier 10.1109/TNNLS.2021.3071026

2162-237X © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

between the pixel under test and the background. In [29], global and local versions of RXD have been investigated. In the global Reed–Xiaoli (GRX) detector, the statistical properties (i.e., mean and covariance matrix) of the background are obtained by using the entire image, while, in the local Reed–Xiaoli (LRX) detector, the local image is utilized to estimate the background information. The estimated covariance matrix is used for background suppression. Hence, the performance of the global or local RXD highly depends on the covariance matrix estimation quality. To alleviate the computational load of the RXDs, dimensionality-reduction-based RXD through random projections was designed [26]. It was found that a Mahalanobis distance in the random-projection domain is similar to that in the original domain. It should be pointed out that the covariance matrix estimation is directly performed over the pixels under test (PUTs) in the original RXD [24] when the anomaly to be detected occupies multiple pixels. It means that background data are not used in the multipixel case.

In practice, background data are collected from the neighbor region around the PUTs. In [30], a background data-assisted anomaly detector was proposed for the multipixel case where the anomaly pattern is known *a priori*. This detector exhibits better performance than the classic RXD since additional background data are exploited by the former. It is worth noting that the anomaly pattern has to be known for the detectors in [24], [26], and [30]. In reality, the pattern of the multipixel anomaly may be unknown since the amplitude of each pixel in the anomaly may vary at different angles, positions, and illumination during hyperspectral imaging.

In this article, we address the problem of detecting multiple-pixel anomalies with an unknown pattern in the presence of the Gaussian noise. Two detectors are proposed by exploiting the GLRT-based design criterion and its two-step modification.<sup>1</sup> Interestingly, the 2S-GLRT for the AD problem is found to be equivalent to the 1S-GLRT, but the former has a lower computational complexity. Analytical expressions for the probability of false alarm are derived, which indicates that the proposed detector bears a constant false alarm rate against the noise covariance matrix. Experiments based on real data show that the proposed detector outperforms its counterparts.

The remainder of this article is organized as follows. In Section II, the data model is introduced, and the problem of interest is formulated. Two adaptive detectors are devised in Section III. In Section IV, analysis of the threshold setting of the proposed strategy is provided. Illustrative results are presented in Section V, and finally, this article is concluded in Section VI.

**Notations:** Vectors (matrices) are denoted by boldface lower (upper) case letters. Superscript  $(\cdot)^T$  denotes transpose,

<sup>1</sup>In the one-step GLRT (1S-GLRT), all the unknown parameters are estimated by using both the test data and background data, and then, these estimates take the place of the corresponding parameters in the likelihood ratio. In the two-step GLRT (2S-GLRT), the covariance matrix is assumed known, and a GLRT is designed by only using the test data. Then, background data are employed to estimate the covariance matrix, which is substituted into the previous GLRT.

and  $\mathbb{R}^{m \times n}$  is a real matrix space of dimension  $m \times n$ .  $\mathbf{I}_n$  stands for an identity matrix of  $n \times n$ , and  $\mathbf{0}_{m \times n}$  represents a null matrix of dimension  $m \times n$ . For notational simplicity, we sometimes drop the explicit indexes in  $\mathbf{I}_n$  and  $\mathbf{0}_{m \times n}$  if no confusion exists. The notation  $\sim$  means “be distributed as,” and  $\mathcal{N}(\boldsymbol{\mu}_r, \mathbf{R}_r)$  represents a real-valued Gaussian distribution with mean  $\boldsymbol{\mu}_r$  and covariance matrix  $\mathbf{R}_r$ .  $\mathcal{B}(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1}dt$  is the incomplete (lower) beta function, and  $\Gamma(\cdot)$  is the Gamma function. The matrix  $|\cdot|$  represents the modulus of a scalar and the determinant of a matrix, when the argument is a scalar and matrix, respectively.  $\text{tr}(\cdot)$  denotes the trace of a square matrix, and  $j = \sqrt{-1}$ .  $\mathcal{P}\{\cdot\}$  and  $\lambda_{\max}\{\cdot\}$  are the principal eigenvector and the corresponding maximum eigenvalue of a square matrix, respectively.

## II. PROBLEM FORMULATION

Assume that a hyperspectral image to be processed consists of  $N$  spectral bands, where an anomaly occupies  $K$  pixels. The  $k$ th pixel denoted by  $\mathbf{x}_k \in \mathbb{R}^{N \times 1}$  in the anomaly can be expressed as

$$\mathbf{x}_k = b_k \mathbf{a} + \mathbf{n}_k, \quad k = 1, \dots, K \quad (1)$$

where  $b_k$  is an unknown amplitude of the  $k$ th pixel,  $\mathbf{a} \in \mathbb{R}^{N \times 1}$  is an unknown spectral signature of the anomaly, and  $\mathbf{n}_k \in \mathbb{R}^{N \times 1}$  denotes the spectral vector of the noise in the  $k$ th pixel. As justified in [24]–[26], the noise vector  $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ , for  $k = 1, 2, \dots, K$ , where  $\mathbf{R}$  is an unknown covariance matrix. For ease of reference,  $\mathbf{x}_k$  for  $k = 1, 2, \dots, K$  are referred to as the test data, and  $b_k$  for  $k = 1, 2, \dots, K$  are called the pattern of anomaly. It is worth noting that  $b_k$  may vary at different pixels, because the amplitude depends on the composition of the pixel, as well as the angle, position, and illumination during the hyperspectral imaging. Define  $\mathbf{b} = [b_1, b_2, \dots, b_K]^T \in \mathbb{R}^{K \times 1}$ , which denotes the spatial pattern of the anomaly.

To estimate the covariance matrix, a set of homogeneous background data  $\{\mathbf{y}_l\}_{l=1}^L$ , where no anomaly exists and  $L \geq N$  is the number of background data, is assumed available, i.e.,

$$\mathbf{y}_l = \mathbf{n}_{K+l} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \quad (2)$$

for  $l = 1, 2, \dots, L$ . Suppose further that the noise vectors  $\mathbf{n}_l$  for  $l = 1, 2, \dots, L + K$  are independent identically distributed (IID).

Let the null hypothesis ( $H_0$ ) and alternative hypothesis ( $H_1$ ) be that the anomaly is absent and, respectively, present in the test data. The AD problem to be solved can be written as a binary hypothesis test of  $H_1$  against  $H_0$ :

$$H_0 : \begin{cases} \mathbf{x}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), & k = 1, \dots, K \\ \mathbf{y}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), & l = 1, 2, \dots, L \end{cases} \quad (3a)$$

$$H_1 : \begin{cases} \mathbf{x}_k \sim \mathcal{N}(b_k \mathbf{a}, \mathbf{R}), & k = 1, \dots, K \\ \mathbf{y}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), & l = 1, 2, \dots, L \end{cases} \quad (3b)$$

where  $b_k$ ,  $\mathbf{a}$ , and  $\mathbf{R}$  are unknown.

Note that, when  $K = 1$ , the detection problem in (3) reduces to that addressed in [24], where the classic RXD was proposed. When the pattern of anomaly represented by

$\mathbf{b} = [b_1, b_2, \dots, b_K]^T$  is known, the detection problem in (3) becomes that addressed in [30]. In practice, the pattern of anomaly is rarely known. Thus, we consider a more practically motivated case where the pattern of anomaly is unknown.

### III. PROPOSED DETECTION FRAMEWORK

Write the test and background data into matrix forms, i.e.,

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K] \in \mathbb{R}^{N \times K} \quad (4)$$

and

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_L] \in \mathbb{R}^{N \times L}. \quad (5)$$

Due to the independence among the noise, the joint probability density functions (pdfs) of the test data  $\mathbf{X}$  and the background data  $\mathbf{Y}$  under  $H_0$  and  $H_1$  can be written as

$$f(\mathbf{X}, \mathbf{Y}; \mathbf{R}, H_0) = \frac{\exp[-\frac{1}{2}\text{tr}(\mathbf{R}^{-1}\mathbf{F}_0)]}{(2\pi)^{N(L+K)/2}|\mathbf{R}|^{(L+K)/2}} \quad (6)$$

and

$$f(\mathbf{X}, \mathbf{Y}; \mathbf{a}, \mathbf{b}, \mathbf{R}, H_1) = \frac{\exp[-\frac{1}{2}\text{tr}(\mathbf{R}^{-1}\mathbf{F}_1)]}{(2\pi)^{N(L+K)/2}|\mathbf{R}|^{(L+K)/2}} \quad (7)$$

respectively, where

$$\mathbf{F}_0 = \mathbf{Y}\mathbf{Y}^T + \mathbf{X}\mathbf{X}^T \in \mathbb{R}^{N \times N} \quad (8)$$

and

$$\mathbf{F}_1 = \mathbf{Y}\mathbf{Y}^T + (\mathbf{X} - \mathbf{a}\mathbf{b}^T)(\mathbf{X} - \mathbf{a}\mathbf{b}^T)^T \in \mathbb{R}^{N \times N}. \quad (9)$$

According to the Neyman–Pearson criterion, the optimum solution to the hypothesis testing (3) is given by comparing the likelihood ratio test (LRT) with an appropriate detection threshold. However, the optimum solution cannot be achieved because these likelihood functions  $f(\mathbf{X}, \mathbf{Y}; \mathbf{R}, H_0)$  and  $f(\mathbf{X}, \mathbf{Y}; \mathbf{a}, \mathbf{b}, \mathbf{R}, H_1)$  include unknown parameters  $\mathbf{a}, \mathbf{b}$ , and  $\mathbf{R}$ . Alternatively, two adaptive detectors are designed according to the criterion of 1S-GLRT and its two-step variation, where each unknown parameter is replaced by its own maximum likelihood (ML) estimate.

#### A. One-Step GLRT

According to the principle of the 1S-GLRT, the ML estimates of all the parameters are employed to replace the corresponding unknown parameters in the LRT in one step. Specifically, the 1S-GLRT detector is obtained by

$$\frac{\max_{\{\mathbf{a}, \mathbf{b}, \mathbf{R}\}} f(\mathbf{X}, \mathbf{Y}; \mathbf{a}, \mathbf{b}, \mathbf{R}, H_1)}{\max_{\{\mathbf{R}\}} f(\mathbf{X}, \mathbf{Y}; \mathbf{R}, H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \xi_0 \quad (10)$$

where  $\xi_0$  is a detection threshold.

Based on (6) and (7), the ML estimates (up to a scalar) of the noise covariance matrix  $\mathbf{R}$  are  $\mathbf{F}_1$  and  $\mathbf{F}_0$ , respectively. Replacing  $\mathbf{R}$  by  $\mathbf{F}_1$  and  $\mathbf{F}_0$  under  $H_1$  and  $H_0$ , respectively, we have

$$\begin{cases} \max_{\mathbf{R}} f(\mathbf{X}, \mathbf{Y}; \mathbf{a}, \mathbf{b}, \mathbf{R}, H_1) = \frac{1}{(2e\pi)^{N(L+K)/2}|\mathbf{F}_1|^{(L+K)/2}} \\ \max_{\mathbf{R}} f(\mathbf{X}, \mathbf{Y}; \mathbf{R}, H_0) = \frac{1}{(2e\pi)^{N(L+K)/2}|\mathbf{F}_0|^{(L+K)/2}} \end{cases} \quad (11)$$

Then, (10) can be rewritten as

$$\frac{|\bar{\mathbf{R}} + \mathbf{X}\mathbf{X}^T|}{\min_{\{\mathbf{a}, \mathbf{b}\}} |\bar{\mathbf{R}} + (\mathbf{X} - \mathbf{a}\mathbf{b}^T)(\mathbf{X} - \mathbf{a}\mathbf{b}^T)^T|} \underset{H_0}{\overset{H_1}{\gtrless}} \xi_1 \quad (12)$$

where  $\xi_1$  is a suitable modification of the original detection threshold, and

$$\bar{\mathbf{R}} = \mathbf{Y}\mathbf{Y}^T \in \mathbb{R}^{N \times N}. \quad (13)$$

Under the condition of  $L \geq N$ , (12) can be recast to

$$\frac{|\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}|}{\min_{\{\mathbf{a}, \mathbf{b}\}} |\mathbf{I}_K + (\tilde{\mathbf{X}} - \bar{\mathbf{R}}^{-1/2} \mathbf{a}\mathbf{b}^T)^T (\tilde{\mathbf{X}} - \bar{\mathbf{R}}^{-1/2} \mathbf{a}\mathbf{b}^T)|} \underset{H_0}{\overset{H_1}{\gtrless}} \xi_1 \quad (14)$$

where

$$\tilde{\mathbf{X}} = \bar{\mathbf{R}}^{-1/2} \mathbf{X} \in \mathbb{R}^{N \times K}. \quad (15)$$

As derived in Appendix A, the value of  $\mathbf{b}$  minimizing the determinant is given by

$$\hat{\mathbf{b}} = \tilde{\mathbf{X}}^T \bar{\mathbf{R}}^{-1/2} \mathbf{a}. \quad (16)$$

Using (53), we have

$$\frac{|\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}|}{\min_{\{\mathbf{a}\}} |\mathbf{Q}|} = \frac{1}{\min_{\{\mathbf{a}\}} [1 - g_2(\mathbf{a})]} \underset{H_0}{\overset{H_1}{\gtrless}} \xi_1 \quad (17)$$

where  $\mathbf{Q}$  is defined in (51), and we use [31, eq. (A1-3)] to obtain

$$g_2(\mathbf{a}) = \mathbf{a}^T \bar{\mathbf{R}}^{-1/2} \tilde{\mathbf{X}} (\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \bar{\mathbf{R}}^{-1/2} \mathbf{a}. \quad (18)$$

It is obvious that (17) can be equivalently written as

$$\max_{\{\mathbf{a}\}} g_2(\mathbf{a}) \underset{H_0}{\overset{H_1}{\gtrless}} \xi_2 \quad (19)$$

where  $\xi_2$  is a detection threshold.

From (49), we know that  $\bar{\mathbf{R}}^{-1/2} \mathbf{a}$  is a unit vector. According to the Rayleigh–Ritz theorem [32, p. 176], the value of  $\bar{\mathbf{R}}^{-1/2} \mathbf{a}$  (up to a scaling factor) maximizing  $g_2(\mathbf{a})$  is given by

$$\bar{\mathbf{R}}^{-1/2} \hat{\mathbf{a}} = \mathcal{P}\{\tilde{\mathbf{X}}(\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T\}. \quad (20)$$

Then

$$g_2(\hat{\mathbf{a}}) = \lambda_{\max}\{\tilde{\mathbf{X}}(\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T\}. \quad (21)$$

Hence, the 1S-GLRT detector can be expressed as

$$t_1 = \lambda_{\max}\{\tilde{\mathbf{X}}(\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T\} \underset{H_0}{\overset{H_1}{\gtrless}} \xi_{1\text{S-GLRT}} \quad (22)$$

where  $\xi_{1\text{S-GLRT}}$  is a detection threshold. This detector is, thus, referred to as 1S-GLRT.

#### B. Two-Step GLRT

In the first step, the covariance matrix  $\mathbf{R}$  is assumed known. Under this assumption, the GLRT detector is further given by

$$\frac{\max_{\{\mathbf{a}, \mathbf{b}\}} f(\mathbf{X}|H_1)}{f(\mathbf{X}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \xi_3 \quad (23)$$

where  $\xi_3$  is a detection threshold, and  $f(\mathbf{X}|H_1)$  and  $f(\mathbf{X}|H_0)$  denote the pdfs of the test data  $\mathbf{X}$  under  $H_1$  and  $H_0$ , respectively. Due to the independence among the test data, the pdf of  $\mathbf{X}$  under  $H_1$  can be represented as

$$f(\mathbf{X}|H_1) = \frac{1}{(2\pi)^{NK/2}|\mathbf{R}|^{K/2}} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{R}^{-1}\bar{\mathbf{F}}_1)\right] \quad (24)$$

where

$$\bar{\mathbf{F}}_1 = (\mathbf{X} - \mathbf{a}\mathbf{b}^T)(\mathbf{X} - \mathbf{a}\mathbf{b}^T)^T. \quad (25)$$

In addition, the pdf of  $\mathbf{X}$  under  $H_0$  is

$$f(\mathbf{X}|H_0) = \frac{1}{(2\pi)^{NK/2}|\mathbf{R}|^{K/2}} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{R}^{-1}\mathbf{X}\mathbf{X}^T)\right]. \quad (26)$$

It can be checked that the maximization of  $f(\mathbf{X}|H_1)$  with respect to  $\mathbf{b}$  can be achieved at

$$\hat{\mathbf{b}}_1 = \frac{\mathbf{X}^T \mathbf{R}^{-1} \mathbf{a}}{\mathbf{a}^T \mathbf{R}^{-1} \mathbf{a}}. \quad (27)$$

As a result

$$\max_{\{\mathbf{b}\}} f(\mathbf{X}|H_1) = \frac{1}{(2\pi)^{NK/2}|\mathbf{R}|^{K/2}} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{R}^{-1}\mathbf{X}\mathbf{X}^T)\right] \times \exp\left[\frac{1}{2}\text{tr}\left(\frac{\mathbf{R}^{-1}\mathbf{X}\mathbf{X}^T \mathbf{R}^{-1} \mathbf{a} \mathbf{a}^T}{\mathbf{a}^T \mathbf{R}^{-1} \mathbf{a}}\right)\right]. \quad (28)$$

Substituting (28) into (23) produces

$$\max_{\{\mathbf{a}\}} \frac{\mathbf{a}^T \mathbf{R}^{-1} \mathbf{X} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{a}}{\mathbf{a}^T \mathbf{R}^{-1} \mathbf{a}} \underset{H_0}{\overset{H_1}{\geq}} \xi_4 \quad (29)$$

where  $\xi_4$  is a detection threshold. According to the Rayleigh–Ritz theorem [32, p. 176], we can obtain that

$$\max_{\{\mathbf{a}\}} \frac{\mathbf{a}^T \mathbf{R}^{-1} \mathbf{X} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{a}}{\mathbf{a}^T \mathbf{R}^{-1} \mathbf{a}} = \lambda_{\max}\{\mathbf{R}^{-1/2} \mathbf{X} \mathbf{X}^T \mathbf{R}^{-1/2}\}. \quad (30)$$

In the second step, an estimate of the covariance matrix is needed to take the place of  $\mathbf{R}$  in (30). Based on the background data, the ML estimate (up to a scalar) of the covariance matrix is given as  $\bar{\mathbf{R}}$  in (13). Using  $\bar{\mathbf{R}}$  to replace  $\mathbf{R}$  in (30), the 2S-GLRT detector is obtained as

$$t_2 = \lambda_{\max}\{\mathbf{X}^T \bar{\mathbf{R}}^{-1} \mathbf{X}\} \underset{H_0}{\overset{H_1}{\geq}} \xi_{2\text{S-GLRT}} \quad (31)$$

where  $\xi_{2\text{S-GLRT}}$  is the detection threshold, and  $\bar{\mathbf{R}}$  is given in (13). This detector is referred to as 2S-GLRT.

To conclude this section, some interesting remarks can be described as follows.

- 1) As shown in Appendix B, we have proven that the proposed 1S-GLRT is equivalent to the proposed 2S-GLRT.
- 2) Comparing (22) and (31), the computational load of the proposed 2S-GLRT is lower than that of the proposed 1S-GLRT. Hence, the proposed 2S-GLRT is preferred in practice.
- 3) When  $K = 1$ , the proposed 2S-GLRT detector reduces to the classic RXD for the single-pixel AD [24]. That is to say, the proposed detector is the generalization of the classic RXD from the single-pixel case to the multipixel case.

#### IV. THRESHOLD SETTING

To facilitate the threshold setting, analytical expressions are derived for the proposed 2S-GLRT detector. Under  $H_0$ ,  $\mathbf{x}_k, \mathbf{y}_l \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$  for  $k = 1, \dots, K$  and  $l = 1, \dots, L$ . According to the theorem in [33], the probability of false alarm of the proposed 2S-GLRT detector can be given by

$$P_{\text{FA}} = 1 - C(z, s, r) \sqrt{\left| \mathbf{A} \left( \frac{\xi_{2\text{S-GLRT}}}{1 + \xi_{2\text{S-GLRT}}} \right) \right|} \quad (32)$$

where

$$z = \min(N, K) \quad (33)$$

$$s = (|K - N| - 1)/2 \quad (34)$$

$$r = (L - N - 1)/2 \quad (35)$$

$$C(z, s, r) = \pi^{z/2} \prod_{i=1}^z \frac{\Gamma\left(\frac{i+2s+2r+z+2}{2}\right)}{\Gamma\left(\frac{i}{2}\right) \Gamma\left(\frac{i+2s+1}{2}\right) \Gamma\left(\frac{i+2r+1}{2}\right)}. \quad (36)$$

When  $z$  is even,  $\mathbf{A}(\theta)$  is a  $z \times z$  skew-symmetric matrix, and the  $(i, j)$ -element in  $\mathbf{A}(\theta)$  is

$$a_{i,j}(\theta) = \mathcal{E}(\theta; s + j, s + i) - \mathcal{E}(\theta; s + i, s + j) \quad (37)$$

where  $i, j = 1, 2, \dots, z$ , and

$$\mathcal{E}(x; a, b) = \int_0^x t^{a-1} (1-t)^b \mathcal{B}(t; b, r+1) dt. \quad (38)$$

When  $z$  is odd,  $\mathbf{A}(\theta)$  is a  $(z+1) \times (z+1)$  skew-symmetric matrix, and the elements are as in (37) with the additional elements

$$a_{i,z+1}(\theta) = \mathcal{B}(\theta; s + i, r + 1), \quad i = 1, \dots, z \quad (39)$$

$$a_{z+1,j}(\theta) = -a_{j,z+1}(\theta), \quad j = 1, \dots, z \quad (40)$$

and

$$a_{z+1,z+1}(\theta) = 0. \quad (41)$$

Note that  $a_{i,j}(\theta) = -a_{j,i}(\theta)$  and  $a_{i,i}(\theta) = 0$ . It is obvious that the probability of false alarm is irrelevant to the noise covariance matrix, which means that the proposed detector exhibits a constant false alarm rate (CFAR) property against the noise covariance matrix.

Note that the exact expression given in (32) is somewhat complicated. If high accuracy is not required, we provide an approximate expression for the probability of false alarm to alleviate the computational load [33], i.e.,

$$P_{\text{FA}} \approx 1 - \gamma \left( \tau, \frac{\ln \xi_{2\text{S-GLRT}} - \mu + \sigma \alpha}{\delta \sigma} \right) \quad (42)$$

where  $\tau = 46.446$ ,  $\delta = 0.186054$ ,  $\alpha = 9.84801$ , and

$$\mu = 2 \ln \tan\left(\frac{\varphi + \phi}{2}\right) \quad (43)$$

$$\sigma^3 = \frac{16}{(L + K - 1)^2 \sin^2(\varphi + \phi) \sin \varphi \sin \phi} \quad (44)$$

$$\varphi = \arccos\left(\frac{L + K - 2N}{L + K - 1}\right) \quad (45)$$

$$\phi = \arccos\left(\frac{L - K}{L + K - 1}\right). \quad (46)$$

As will be shown in Section V, the accuracy of the approximate expression in (42) is acceptable.



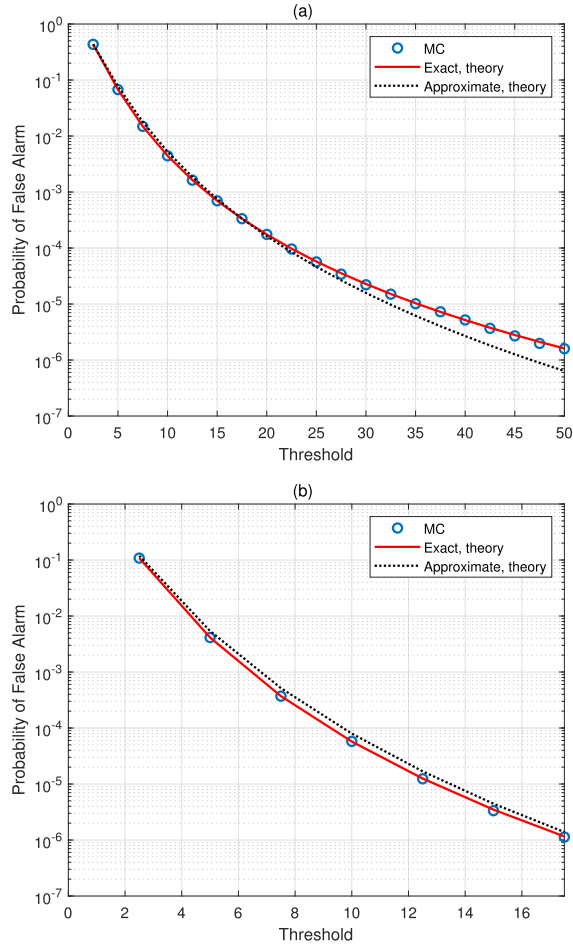


Fig. 1. Detection probability versus SNR with  $K = 4$  and  $N = 10$ . (a)  $L = 20$ . (b)  $L = 25$ .

## V. NUMERICAL EXAMPLES AND ANALYSIS

In this section, numerical examples and experimental analysis are provided to evaluate the detection performance of the proposed detector.

### A. Simulated Data

The spectral signature  $\mathbf{a}$  is randomly generated and then normalized to have unit norm. Specifically, each element is sampled from the uniform distribution between 0 and 1. The pattern vector  $\mathbf{b}$  is chosen as  $\mathbf{b} = \sigma_1[0.5, 1, 0.5, 1, \dots, 0.5, 1]^T$ . Assume that the  $(i, j)$ th element of the noise covariance matrix is chosen as  $[\mathbf{R}]_{i,j} = \sigma_0^2 0.95^{|i-j|}$ , where  $\sigma_0^2$  is the noise power. The signal-to-noise ratio (SNR) in decibel is defined by

$$\text{SNR} = 10 \log_{10} \frac{\sigma_1^2}{\sigma_0^2}. \quad (47)$$

For comparison purposes, we consider the adaptive detector designed in [30] with a known pattern. In practice, the pattern is usually unknown. Thus, the adaptive detector in [30] cannot be used in such a case. It just provides a benchmark of the best achievable detection performance.

In Fig. 1, the probability of false alarm of the proposed detector as a function of the detection threshold is plotted,

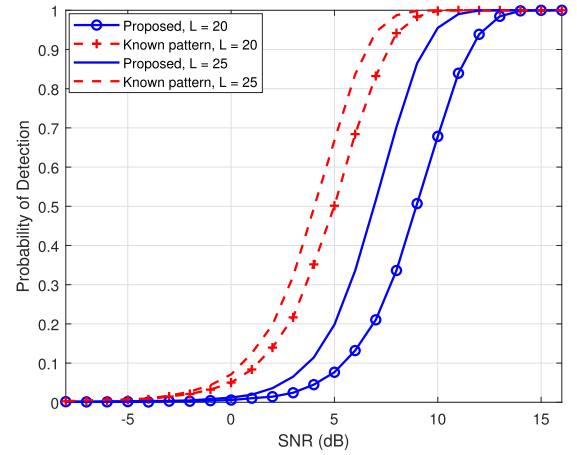


Fig. 2. Detection probability versus SNR with  $K = 4$  and  $N = 10$ .

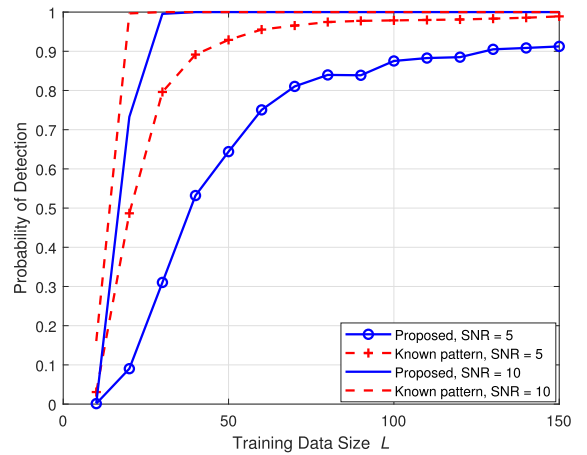


Fig. 3. Detection probability versus  $L$  with  $K = 4$  and  $N = 10$ .

where  $K = 4$  and  $N = 10$ . The solid and dotted lines denote the results obtained using the exact and approximate expressions [see (32) and (42)], respectively, while the symbols “o” denote the Monte Carlo (MC) results. We can see that the exact results match the MC ones, and the approximation accuracy is very high.

In Fig. 2, the detection probability is depicted with respect to the SNR for different  $L$ 's, where  $K = 4$ ,  $N = 10$ , and the probability of false alarm is set to be  $10^{-3}$ . It can be observed that the performance of the proposed detector is worse than that of its counterpart. This is because the latter exploits *a priori* knowledge about the pattern. As the background data size  $L$  increases from 20 to 25, the performance of both detectors improves, and the performance loss of the proposed detector compared with its counterpart becomes smaller.

The detection probabilities versus the background data size  $L$  for different SNRs are presented in Fig. 3, where  $K = 4$ ,  $N = 10$ , and  $P_{FA} = 10^{-3}$ . The results highlight that the performance of the detectors improves as  $L$  increases. This is because the covariance matrix estimation accuracy improves when more background data are used.

In Fig. 4, the detection probability is plotted as a function of the pixel number  $K$  for different  $L$ 's, where  $N = 10$  and

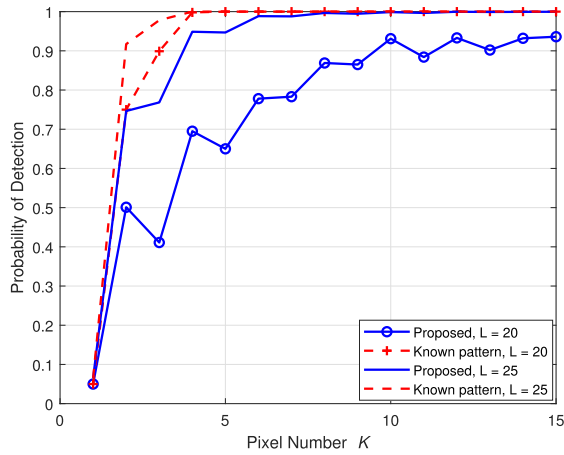


Fig. 4. Detection probability versus  $K$  with  $N = 10$  and  $\text{SNR} = 10$  dB.

$\text{SNR} = 10$  dB. Interestingly, the performance of the proposed detector does not necessarily improve when more pixels are used. This can be explained as follows. Recall that the pattern vector  $\mathbf{b}$  is chosen as  $\mathbf{b} = \sigma_1[0.5, 1, 0.5, 1, \dots, 0.5, 1]^T$ . It means that the signal powers in the pixels with odd indices are lower than those in the pixels with even indices. When the pixel with a small power is added in the detection, the performance may be deteriorated, especially when the background data size  $L$  is small (namely, the noise cannot be suppressed enough). This means that the best performance of the proposed detector may not be achieved with too small or large  $K$ ; especially, the conventional RXD with  $K = 1$  is not optimal. In practice, we have to empirically choose an appropriate value of  $K$  for a specific situation.

### B. Real Hyperspectral Data

The scene of the first image is the Los Angeles area, which was obtained by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor on November 9, 2011 [34]. The image consists of  $100 \times 100$  pixels with a spatial resolution of 7.1 m and consists of 205 spectral bands after removing the bands (i.e., the water absorption regions and low SNR bands) unsuitable for detection, where the anomalies come from the airplanes. The scene and the ground-truth map are shown in Fig. 5.

The scene of the second image is the Cat Island area [34], and the spatial size of this data set is  $150 \times 150$ , with a spatial resolution of 17.2 m. It was obtained by the hyperspectral airborne AVIRIS sensor on September 12, 2010. After removing the unsuitable bands, 188 bands are retained. Anomalies consist of an airplane. The image and the ground truth of anomalies are illustrated in Fig. 6.

The third image was obtained in the Texas Coast on August 29, 2010, which was also obtained by the airborne AVIRIS sensor [34]. It consists of  $100 \times 100$  pixels with a spatial resolution of 17.2 m and consists of 204 spectral bands after removing the unsuitable bands, where the anomalies come from airplanes. The scene and the ground-truth map are shown in Fig. 7.

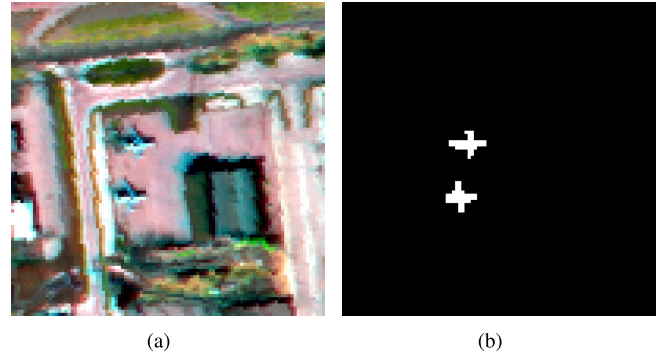


Fig. 5. Illustration of the Los Angeles data set: (a) pseudocolor (RGB: 150, 60, and 25) and (b) ground-truth map.

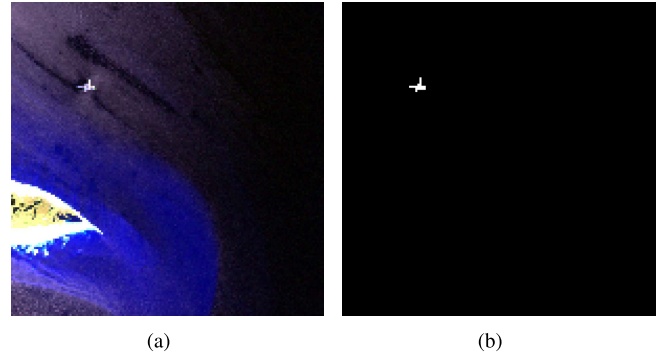


Fig. 6. Illustration of the Cat Island data set: (a) pseudocolor (RGB: 130, 70, and 15) and (b) ground-truth map.

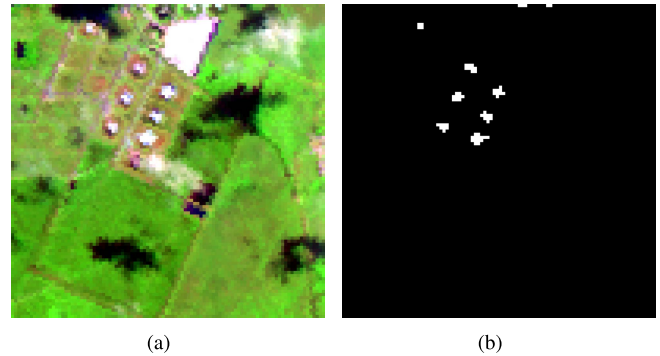


Fig. 7. Illustration of the Texas Coast data set: (a) pseudocolor (RGB: 140, 60, and 25) and (b) ground-truth map.

The scene of the fourth image is the Pavia Centre, Italy [35]. The scene was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) airborne sensor during a flight campaign over Pavia, Northern Italy. The spatial size of this data set is  $150 \times 150$ , which has a geometric resolution of 1.3 m and contains 115 bands. After removal of the unsuitable bands, 102 bands are retained. The main backgrounds in the image are water and a bridge, and the main anomalies are vehicles on the bridge. The scene and the ground-truth map are shown in Fig. 8.

The last data set was acquired by the Nuance Cri Hyperspectral sensor [21], [36]. This sensor can acquire imagery with a spectral resolution of 10 nm covering the wavelengths

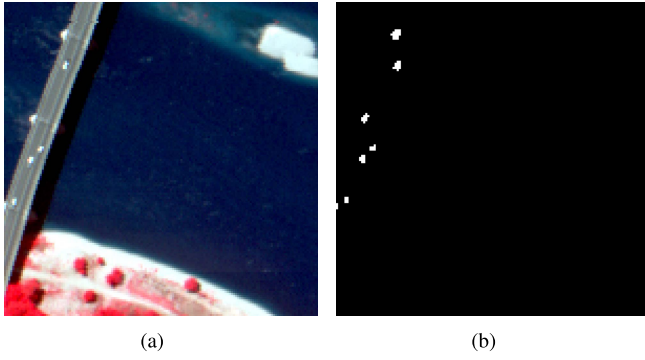


Fig. 8. Illustration of the Pavia Centre data set: (a) pseudocolor (RGB: 80, 60, and 40) and (b) ground-truth map.

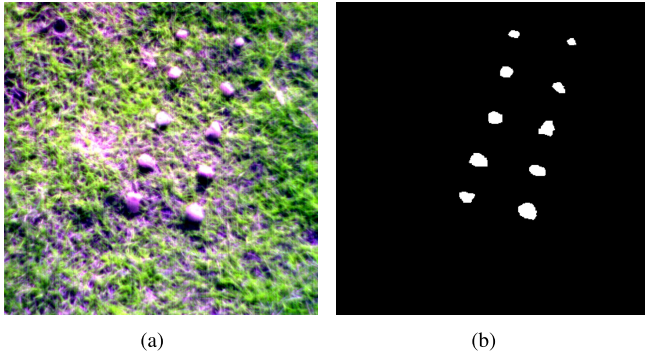


Fig. 9. Illustration of the Cri data set: (a) pseudocolor (RGB: 30, 15, and 1) and (b) ground-truth map.

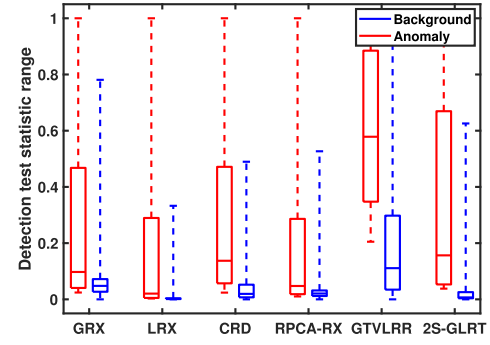
from 650 to 1100 nm. The image contains  $400 \times 400$  pixels with 46 spectral bands, where the background is grass and the anomalies are 10 rocks consisting of 2261 pixels. The pseudocolor image and the position information of anomalies are presented in Fig. 9.

### C. Detection Performance

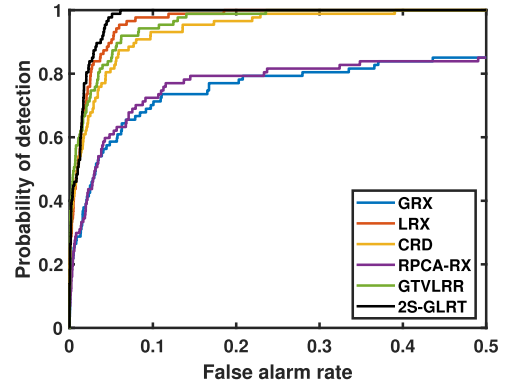
AD performance of the proposed 2S-GLRT is compared with that of some other detectors. In these experiments, five widely used anomaly detectors, such as the GRX [24], LRX [29], state-of-the-art CRD [17], robust principal component analysis with Reed–Xiaoli (RPCA-RX) detector [37], and graph total variation regularized low rank representation (GTVLRR) [38], are employed. For qualitative and quantitative comparisons, the statistical separability analysis [39], [40], receiver operating characteristic (ROC), and area under the curve (AUC) [41] metric are utilized as the main criteria for evaluation.

A standard approach (i.e., a double concentric sliding window [11]) is employed in the proposed detection procedure. The sliding window consists of the inner and outer windows. The background data are selected from the outer window region. The pixels in the inner window region are jointly exploited to detect the multipixel anomaly. When the value of the proposed 2S-GLRT exceeds a precalculated threshold, it is declared that the central pixel is an anomaly.

It is worth noting that the initial choices of different parameters are important for the LRX, CRD, RPCA-RX, GTVLRR, and the proposed 2S-GLRT. Therefore, to make fully release



(a)



(b)

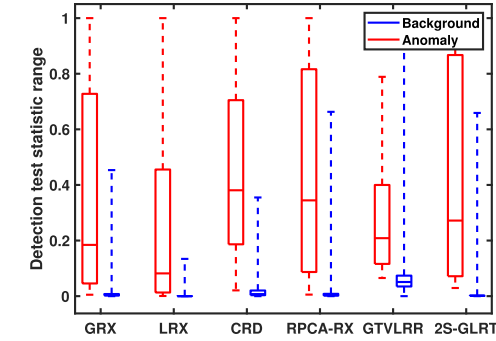
Fig. 10. ROC curves of different methods in the Los Angeles data set. (a) Statistical separability analysis. (b) ROC curves.

the performance of various detectors, the local optimal parameters of various detectors are selected for comparative experiments. In the LRX and 2S-GLRT, by varying the outer window size  $\omega_{out}$  from 5 to 25 and the inner window size  $\omega_{in}$  from 3 to 15, the optimal detection performance under different window sizes ( $\omega_{out}, \omega_{in}$ ) is collected. Similarly, for the CRD, by fixing  $\lambda$  as  $10^{-6}$  as suggested [17], [34] and varying  $\omega_{out}$  from 5 to 25 and  $\omega_{in}$  from 3 to 15, the optimal detection performance under different window sizes is selected. For the RPCA-RX, the optimal parameter  $\lambda$ , which ranges from  $10^{-4}$  to  $10^4$ , is selected for comparative experiments. In the GTVLRR, the choice of dictionary is very important. Therefore, the K-means method is used to cluster into six categories during the construction of dictionary according to the recommendation in [38], and then, the largest 20 elements in each category are selected as atoms of dictionary. Empirical parameters ( $\lambda$ ,  $\beta$ , and  $\gamma$ ), as suggested in [38], for GTVLRR are first set to (0.5, 0.2, 0.05). Then, by fixing  $\beta$  and  $\gamma$ , and varying  $\lambda$  in [0.005, 0.05, 0.1, 0.3, 0.5, 0.7, 1], the optimal  $\lambda$  under different values is selected. Similarly, by fixing the selected  $\lambda$  and default  $\gamma$ , and varying  $\beta$  in [0.005, 0.05, 0.1, 0.2, 0.4, 0.7, 1], the optimal  $\beta$  is selected. In the same way, the optimal  $\gamma$  is selected in [0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5]. The optimal parameters of different detection methods are listed in Table I.

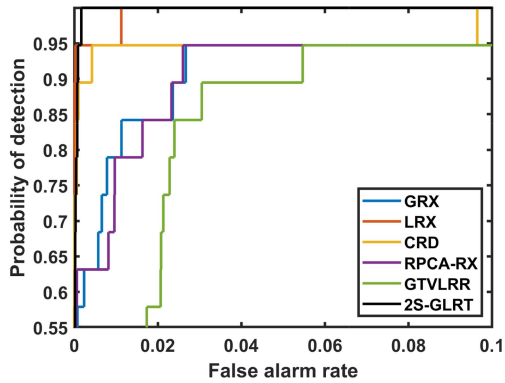
The statistical separability analysis, also known as boxplot [39], is employed for performance assessment. It is mainly used to reflect the distribution characteristics and can also

TABLE I  
OPTIMAL PARAMETERS OF VARIOUS DETECTORS IN DIFFERENT DATA SETS

Methods	LRX	CRD	RPCA-RX	GTVLRR	2S-GLRT
Los Angeles	(19,15)	(17,15,10 <sup>-6</sup> )	0.01	(0.5,0.2,0.1)	(19,15)
Cat Island	(17,13)	(25,15,10 <sup>-6</sup> )	0.01	(1,0.005,0.005)	(25,3)
Texas Coast	(11,5)	(9,7,10 <sup>-6</sup> )	0.01	(0.005,1,0.5)	(9,5)
Pavia Centre	(25,5)	(7,5,10 <sup>-6</sup> )	0.01	(0.7,0.4,0.05)	(21,5)
Cri	(19,15)	(19,9,10 <sup>-6</sup> )	0.001	(0.005,1,0.5)	(25,15)

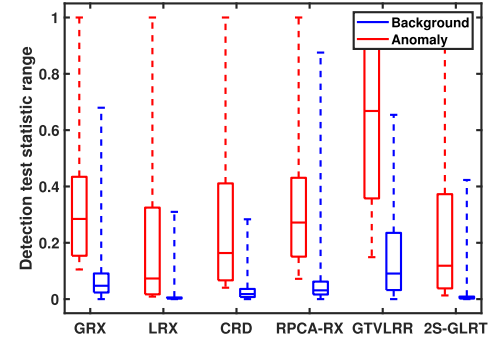


(a)

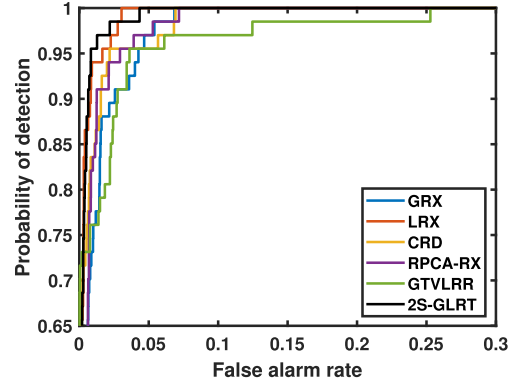


(b)

Fig. 11. ROC curves of different methods in the Cat Island data set. (a) Statistical separability analysis. (b) ROC curves.



(a)



(b)

Fig. 12. ROC curves of different methods in the Texas Coast data set. (a) Statistical separability analysis. (b) ROC curves.

be used to compare distribution characteristics of multiple groups of data. Essentially, the separability analysis is a statistical distribution map of detection results corresponding to anomalous and background pixels. In statistical separability analysis, the horizontal line in the middle of red or blue boxes represents the median of anomaly or background pixels and has a positive correlation with the mean to some extent. The upper and lower boundaries of boxes are 90% and 10% of the statistical interval, respectively, while the 0%–10% and 90%–100% intervals are represented by dotted lines. The red box represents anomalous targets, and the blue box represents background pixels, where the interval between the red box and the blue box represents the separability between anomalous and background pixels. The height of the blue box represents the suppression degree of these methods to the background. Generally, the lower the blue box height is, the stronger the background suppression is, which also translates into a better separation between the background and the anomaly.

For the Los Angeles data set, as shown in Fig. 10(a), the interval between the red and blue boxes of the proposed 2S-GLRT is larger than those of the GRX, LRX, CRD, RPCA-RX, and GTVLRR, which indicates that the proposed method can separate anomalies from the background more effectively. In Fig. 10(b), the black curve associated with the 2S-GLRT is closest to the upper right corner, which demonstrates that the proposed 2S-GLRT has better detection performance than the other alternatives.

In the Cat Island data set, as shown in Fig. 11(a), it is concluded that the six detectors, i.e., the GRX, LRX, CRD, RPCA-RX, GTVLRR, and 2S-GLRT, can effectively separate anomalies from the background, where the CRD achieves a better separability than the others. However, in the ROC curves shown in Fig. 11(b), the proposed 2S-GLRT is the best detector.

For the Texas Coast data set, as shown in Fig. 12(a), the height of the blue boxes associated with the 2S-GLRT

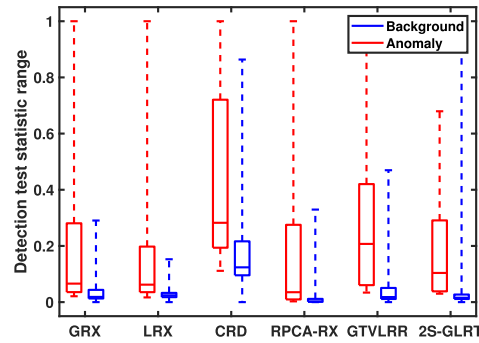


TABLE II  
AUC VALUES (%) OF VARIOUS ANOMALY DETECTORS USING DIFFERENT DATA SETS

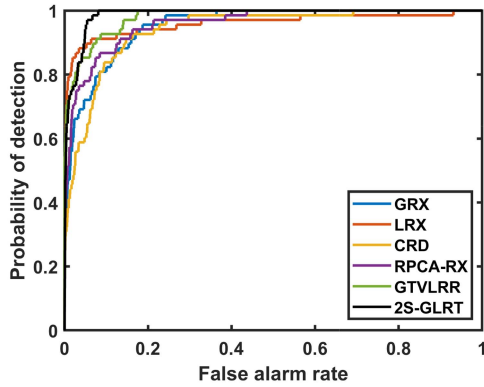
Methods	GRX	LRX	CRD	RPCA-RX	GTVLRR	2S-GLRT
Los Angeles	84.038	98.207	96.809	84.737	97.770	<b>98.735</b>
Cat Island	98.072	99.936	99.458	98.320	97.461	<b>99.968</b>
Texas Coast	99.065	99.691	99.404	99.265	98.836	<b>99.697</b>
Pavia Centre	95.380	95.973	94.072	95.906	97.950	<b>98.676</b>
Cri	91.342	75.251	87.267	91.478	93.429	<b>95.403</b>

TABLE III  
EXECUTION TIME OF VARIOUS ANOMALY DETECTORS USING DIFFERENT DATA SETS (UNIT: SECONDS)

Methods	GRX	LRX	CRD	RPCA-RX	GTVLRR	2S-GLRT
Los Angeles	0.11	43.24	5.72	5.56	67.68	161.57
Cat Island	0.16	81.83	254.82	5.68	95.34	99.25
Texas Coast	0.07	37.21	2.78	4.66	54.34	33.88
Pavia Centre	0.09	54.21	4.02	2.21	98.93	38.13
Cri	0.32	87.32	821.30	8.51	933.77	3903.11



(a)

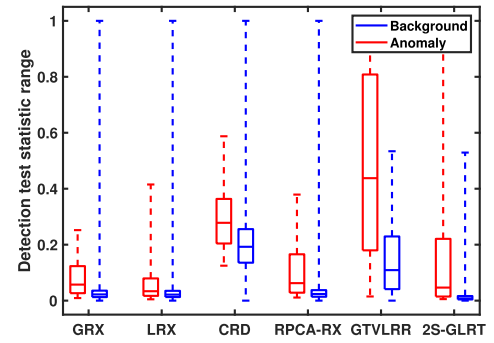


(b)

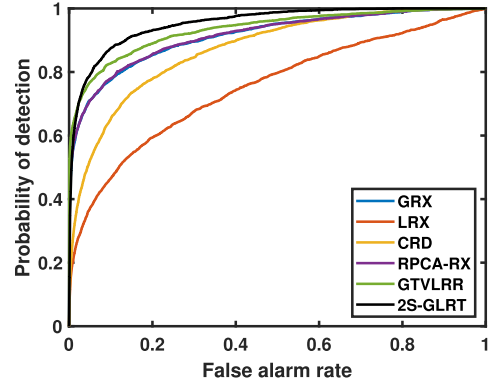
Fig. 13. ROC curves of different methods in the Pavia Centre data set. (a) Statistical separability analysis. (b) ROC curves.

and LRX is lower than the others, which indicates that LRX and 2S-GLRT can effectively suppress background, and the interval between the red box and the blue box in the 2S-GLRT is larger than the LRX, which indicates that the proposed 2S-GLRT can better separate anomalies from the background. Similarly, in Fig. 12(b), the 2S-GLRT presents better performance than the other detectors.

For the Pavia Centre data set, as shown in Fig. 13(a), a similar conclusion can be drawn that the proposed 2S-GLRT has the best ability to separate anomalies from the background.



(a)



(b)

Fig. 14. ROC curves of different methods in the Cri data set. (a) Statistical separability analysis. (b) ROC curves.

The ROC curves shown in Fig. 13(b) also confirm that the 2S-GLRT yields the best performance.

Similarly, for the Cri data set, as shown in Fig. 14, the interval between the red and blue boxes of the proposed 2S-GLRT is slightly larger than others, and the black ROC curve is closest to the upper right corner. All these results indicate that the proposed 2S-GLRT can effectively separate anomalies from the background, and the designed detector has a better performance than these traditional detectors, i.e., the GRX, LRX, CRD, RPCA-RX, and GTVLRR.

For quantitative comparisons, AUC values are employed to evaluate the detection performance, as listed in Table II. From the AUC values of the five real hyperspectral data sets, a conclusion can be drawn that the performance of the proposed 2S-GLRT is the best, especially for the Cri data set. Compared with the other detectors, the improvement of 2S-GLRT is approximately 2%.

Table III provides the computational costs of various detectors. Although the execution time of the proposed 2S-GLRT is less than that of the LRX in the Texas Coast and Pavia Centre data sets and is less than that of the CRD in the Cat Island data set. Its computational cost is still high, especially for the Los Angeles and Cri data sets. The reason is due to that a larger window size with more pixels is used by the detector, which makes the matrix inverse computationally more intensive.

## VI. CONCLUSION

In this work, we addressed the multipixel AD problem with the Gaussian background for the HSI, where the spatial pattern of the anomaly is unknown. The proposed 1S-GLRT and 2S-GLRT detectors have been mathematically studied according to the GLRT principle. It is found that the 1S-GLRT is equivalent to the 2S-GLRT, but the 2S-GLRT has a less computational load. Analytical expressions for the probability of false alarm of the proposed 2S-GLRT were derived, which have demonstrated that the proposed detector has the CFAR property against the noise covariance matrix. Numerical examples based on simulated data demonstrated that the performance of the proposed detector improves as the background data size increases. Five sets of real data results confirm that the proposed detector is better than several popular prior solutions. Future work may focus on the multipixel AD problem in non-Gaussian backgrounds.

## APPENDIX A DERIVATIONS OF (16)

Here, we derive the optimal  $\mathbf{b}$  that minimizes the determinant in (14). Let us proceed by defining

$$g_1(\mathbf{a}, \mathbf{b}) = |\mathbf{I}_K + (\tilde{\mathbf{X}} - \tilde{\mathbf{R}}^{-1/2}\mathbf{a}\mathbf{b}^T)^T(\tilde{\mathbf{X}} - \tilde{\mathbf{R}}^{-1/2}\mathbf{a}\mathbf{b}^T)|. \quad (48)$$

It is obvious that, if  $\mathbf{a}_0$  and  $\mathbf{b}_0$  are the optimal values minimizing  $g_1$ , then  $c\mathbf{a}_0$  and  $\mathbf{b}_0/c$  are also the optimal solutions. To avoid such an ambiguity, we impose a constraint on  $\mathbf{a}$ , i.e.,

$$\mathbf{a}^T \tilde{\mathbf{R}}^{-1} \mathbf{a} = 1. \quad (49)$$

Under such a constraint,  $g_1(\mathbf{a}, \mathbf{b})$  can be rewritten as

$$g_1(\mathbf{a}, \mathbf{b}) = |\mathbf{Q} + (\mathbf{b} - \tilde{\mathbf{X}}^T \tilde{\mathbf{R}}^{-1/2} \mathbf{a})(\mathbf{b} - \tilde{\mathbf{X}}^T \tilde{\mathbf{R}}^{-1/2} \mathbf{a})^T| \quad (50)$$

where

$$\mathbf{Q} = \mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} - \tilde{\mathbf{X}}^T \tilde{\mathbf{R}}^{-1/2} \mathbf{a} \mathbf{a}^T \tilde{\mathbf{R}}^{-1/2} \tilde{\mathbf{X}}. \quad (51)$$

Furthermore, we have

$$\begin{aligned} g_1(\mathbf{a}, \mathbf{b}) &= |\mathbf{Q}||\mathbf{I}_K + \mathbf{Q}^{-1}(\mathbf{b} - \tilde{\mathbf{X}}^T \tilde{\mathbf{R}}^{-1/2} \mathbf{a})(\mathbf{b} - \tilde{\mathbf{X}}^T \tilde{\mathbf{R}}^{-1/2} \mathbf{a})^T| \\ &= |\mathbf{Q}||1 + (\mathbf{b} - \tilde{\mathbf{X}}^T \tilde{\mathbf{R}}^{-1/2} \mathbf{a})^T \mathbf{Q}^{-1}(\mathbf{b} - \tilde{\mathbf{X}}^T \tilde{\mathbf{R}}^{-1/2} \mathbf{a})| \end{aligned} \quad (52)$$

where the second equality is obtained using [42, p. 59]. Thus, the optimal  $\mathbf{b}$  minimizing  $g_1(\mathbf{a}, \mathbf{b})$  is given as  $\hat{\mathbf{b}}$  in (16), and

$$g_1(\mathbf{a}, \hat{\mathbf{b}}) = |\mathbf{Q}|. \quad (53)$$

## APPENDIX B EQUIVALENCE BETWEEN 1S-GLRT AND 2S-GLRT

Now, we show that the proposed 1S-GLRT is equivalent to the proposed 2S-GLRT. Denote by  $\mathbf{z}$  the principal eigenvector of  $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ , i.e.,

$$\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \mathbf{z} = t_2 \mathbf{z} \quad (54)$$

where  $t_2$  defined in (31) is the maximal eigenvalue corresponding to the principal eigenvector of  $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ . Adding  $\mathbf{z}$  at both sides of (54) yields

$$(\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}) \mathbf{z} = (1 + t_2) \mathbf{z}. \quad (55)$$

Furthermore, we have

$$\mathbf{z} = (1 + t_2)(\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \mathbf{z}. \quad (56)$$

Multiplying by  $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$  at both sides of (56) leads to

$$\begin{aligned} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \mathbf{z} &= (1 + t_2) \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} (\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \mathbf{z} \\ &= t_2 \mathbf{z} \end{aligned} \quad (57)$$

where the second equation is obtained using (54). It follows from (57) that:

$$\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} (\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \mathbf{z} = \frac{t_2}{1 + t_2} \mathbf{z}. \quad (58)$$

Since  $(t_2/1 + t_2)$  is a monotonically increasing function of  $t_2$ , we can obtain that

$$\frac{t_2}{1 + t_2} = \lambda_{\max}\{\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} (\mathbf{I}_K + \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}\}. \quad (59)$$

Combining (22) and (59), we have

$$t_1 = \frac{t_2}{1 + t_2} \quad (60)$$

which proves that the proposed 1S-GLRT is equivalent to the proposed 2S-GLRT.

## REFERENCES

- [1] S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "Models and methods for automated background density estimation in hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2837–2852, May 2013.
- [2] D. Ma, Y. Yuan, and Q. Wang, "Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation," *Remote Sens.*, vol. 10, no. 5, pp. 1–21, 2018.
- [3] Q. Ling, Y. Guo, Z. Lin, L. Liu, and W. An, "A constrained sparse-representation-based binary hypothesis model for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1933–1947, Jun. 2019.
- [4] H. Kwon and N. M. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 178–194, Feb. 2006.
- [5] L. Ru, B. Du, and C. Wu, "Multi-temporal scene classification and scene change detection with correlation based fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 1382–1394, 2021.
- [6] R. Pike, G. Lu, D. Wang, Z. G. Chen, and B. Fei, "A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 653–663, Mar. 2016.

- [7] X. Hao, Y. Wu, and P. Wang, "Angle distance-based hierarchical background separation method for hyperspectral imagery target detection," *Remote Sens.*, vol. 12, no. 697, pp. 1–23, 2020.
- [8] 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.*, early access, Feb. 25, 2020, doi: [10.1109/TCYB.2020.2968750](https://doi.org/10.1109/TCYB.2020.2968750).
- [9] R. Zhao, B. Du, L. Zhang, and L. Zhang, "A robust background regression based score estimation algorithm for hyperspectral anomaly detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 122, pp. 126–144, Dec. 2016.
- [10] S. Chang, B. Du, and L. Zhang, "A subspace selection-based discriminative forest method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4033–4046, Jun. 2020.
- [11] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–27, Jul. 2010.
- [12] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature approximation for anomaly detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.
- [13] Y. Xu, B. Du, L. Zhang, and S. Chang, "A low-rank and sparse matrix decomposition-based dictionary reconstruction and anomaly extraction framework for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1248–1252, Jul. 2020.
- [14] 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.*, early access, Dec. 9, 2020, doi: [10.1109/TNNLS.2020.3038659](https://doi.org/10.1109/TNNLS.2020.3038659).
- [15] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.
- [16] 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.
- [17] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [18] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3123–3134, Dec. 2016.
- [19] Q. Ling, Y. Guo, Z. Lin, and W. An, "A constrained sparse representation model for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2358–2371, Apr. 2019.
- [20] 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.
- [21] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.
- [22] R. Zhao, B. Du, and L. Zhang, "Hyperspectral anomaly detection via a sparsity score estimation framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3208–3222, Jun. 2017.
- [23] T. Veracini, S. Matteoli, M. Diani, and G. Corsini, "Nonparametric framework for detecting spectral anomalies in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 666–670, Jul. 2011.
- [24] 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.
- [25] X. Yu, I. S. Reed, and A. D. Stocker, "Comparative performance analysis of adaptive multispectral detectors," *IEEE Trans. Signal Process.*, vol. 41, no. 8, pp. 2639–2656, Aug. 1993.
- [26] J. E. Fowler and Q. Du, "Anomaly detection and reconstruction from random projections," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 184–195, Jan. 2012.
- [27] C.-I. Chang, "Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.
- [28] R. Tao, X. Zhao, W. Li, H.-C. Li, and Q. Du, "Hyperspectral anomaly detection by fractional Fourier entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.
- [29] J. M. Molero, E. M. Garzon, I. Garcia, 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.
- [30] J. Liu, Y. Feng, W. Liu, D. Orlando, and H. Li, "Training data assisted anomaly detection of multi-pixel targets in hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 68, pp. 3022–3032, 2020.
- [31] E. J. Kelly and K. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," Lincoln Lab., MIT, Cambridge, MA, USA, Tech. Rep. 848, 1989.
- [32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [33] J. Liu, J. Chen, J. Li, and W. Liu, "Persymmetric adaptive detection of distributed targets with unknown steering vectors," *IEEE Trans. Signal Process.*, vol. 68, pp. 4123–4134, 2020.
- [34] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [35] A. W. Bitar, L.-F. Cheong, and J.-P. Ovarlez, "Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5239–5251, Aug. 2019.
- [36] 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.
- [37] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [38] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jan. 2020.
- [39] K. Tan, Z. Hou, F. Wu, Q. Du, and Y. Chen, "Anomaly detection for hyperspectral imagery based on the regularized subspace method and collaborative representation," *Remote Sens.*, vol. 11, no. 11, p. 1318, Jun. 2019.
- [40] K. Tan, Z. Hou, D. Ma, Y. Chen, and Q. Du, "Anomaly detection in hyperspectral imagery based on low-rank representation incorporating a spatial constraint," *Remote Sens.*, vol. 11, no. 13, p. 1578, Jul. 2019.
- [41] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
- [42] G. A. Seber, *A Matrix Handbook for Statisticians*. Hoboken, NJ, USA: Wiley, 2008.