Training Data Assisted Anomaly Detection of Multi-Pixel Targets In Hyperspectral Imagery

Jun Liu, Senior Member, IEEE, Yutong Feng, Weijian Liu, Senior Member, IEEE, Danilo Orlando, Senior Member, IEEE, and Hongbin Li, Fellow, IEEE

Abstract—In this paper, we investigate the anomaly detection problem for widespread targets with known spatial patterns under a local Gaussian model when training data are available. Three adaptive detectors are proposed based on the principles of the generalized likelihood ratio test, the Rao test, and the Wald test, respectively. We prove that these tests are statistically equivalent to each other. In addition, analytical expressions for the probability of false alarm and probability of detection of the proposed detectors are obtained, which are verified through Monte Carlo simulations. It is shown that these detectors have a constant false alarm rate against the covariance matrix. Finally, numerical examples using synthetic and real hyperspectral data demonstrate that these training data assisted detectors have better detection performance than their counterparts without training data.

Index Terms—Anomaly detection, hyperspectral images, constant false alarm rate, generalized likelihood ratio test, Rao test, Wald test.

I. INTRODUCTION

HYPERSONTICAL imaging sensors collect data in hundreds of contiguous spectral bands for each pixel of a scene [1]–[3]. The fine spectral resolution makes it possible to identify objects or features in the scene based on their spectral signatures. One application of hyperspectral data is target detection which has been widely investigated and found to be very useful, e.g., defense-and-surveillance, search-and-rescue, and mine exploration systems [2], [4], [5]. From a detection-theoretic point of view, target detection problem is a binary hypothesis test. A typical solution is the Neyman-Pearson (NP) approach [2], which maximizes the detection probability for a preassigned false alarm probability. In recent years, spectral signature-based target detection has been studied by many researchers [3], [6]–[8]. In most of such works, it is assumed that the target spectral signatures are known. The goal is to detect a signal with known signature but unknown amplitude from the background clutter. Many solutions have been developed to solve similar problems in radar target detection, such as Kelly’s generalized likelihood ratio test (GLRT) [9], the adaptive matched filter [10], the adaptive coherence estimator test [11], and the adaptive Rao test [12]. However, in real applications, the effects of atmospheric fluctuations and some other complicating factors may have a strong influence on the detection results [2], [13], [14]. This has led to the consideration of anomaly detection methods which are designed to distinguish unusual targets (anomalies) from the background without target signature references [1], [15].

In hyperspectral imaging applications, anomaly detection can be considered as a particular case of target detection. It tries to detect rare objects (anomalies) whose spectral signatures are different from those of their surroundings, when no a priori information about the target spectral signatures is available. Anomalies can be defined with respect to a model of the background. Local spectral anomalies can be defined as observations that deviate from the neighboring clutter background, while global spectral anomalies are defined as the pixels whose signatures spectrally distinct from the image-wide clutter background [8], [16]. Both approaches have their advantages and disadvantages (for more detailed discussion, see [1], [8]). The local model method is useful for the background clutter characterization. However, it may be susceptible to false alarms that are isolated anomalies. The global anomaly detection algorithms are not likely to generate this type of false alarms, but it may be incapable of properly detecting isolated targets. In this paper, we focus on the anomaly detection problem based on the local approach.

Due to mathematical tractability, most local methods for hyperspectral data rely on a Gaussian model which assumes that the background clutter obeys a real-valued Gaussian multivariate distribution with an unknown covariance matrix. In real hyperspectral data, the Gaussian assumption was found suitable after the data is processed by a spatial sliding window approach [17], [18]. Based on the local Gaussian model, the well-known Reed-Xiaoli detector (RXD) was derived in [18] for detecting multi-pixel targets with known spatial patterns [8], [18]. If the spatial pattern is neglected (i.e., just one anomaly pixel is intended to be detected), the original RXD test variable reduces to the Mahalanobis distance which is also widely used
for anomaly detection [1], [8], [15], [19]. The RXD has been successfully applied in many real hyperspectral applications and is considered as a benchmark anomaly detector for hyperspectral data. When the anomaly to be detected is multiple-pixel, covariance matrix estimation is directly performed over the pixels under test (PUTs) in the original RXD [17], [18], [20]. That is, no additional training data are used in the multi-pixel case. In practice, additional training data are often available, which can be collected using a guard window method [3], [8], [21]–[23], and it is a standard practice to employ training data for solving detection problems in hyperspectral imagery [1], [2], [15].

In this paper, the anomaly detection problem under a local Gaussian model is investigated. We consider the situation for detecting widespread targets with a known spatial pattern in hyperspectral images when training data are available. The corresponding GLRT, Rao test and Wald test are derived. Interestingly, we find that the three tests are coincident with each other. In addition, detection and false alarm probabilities of the proposed detectors are obtained and validated by numerical simulations. It is found from these expressions that the proposed detectors have the CFAR property. Finally, experiments are conducted on synthetic and real hyperspectral images showing that the proposed detectors outperform the RXD.

The remainder of this paper is organized as follows. Section II provides signal models and formulates the detection problem. In Section III, we design the GLRT, Rao test and Wald test, and prove that the three tests coincide. Analytical performance is evaluated in Section IV. Numerical simulations and experiments on synthetic hyperspectral data are carried out in Section V. Finally, the paper is summarized in Section VI.

II. PROBLEM FORMULATION

We consider the anomaly detection problem for widespread targets in hyperspectral images when the training data are available. The problem can be formulated as a binary hypothesis test which is to decide between the null hypothesis \( H_0 \) and the alternative one \( H_1 \)

\[
H_0 : \begin{cases} 
\mathbf{r}_n = \mathbf{x}_n, & n = 1, \ldots, N, \\
\mathbf{r}_k = \bar{\mathbf{x}}_k, & k = 1, \ldots, K,
\end{cases}
\]

(1)

and

\[
H_1 : \begin{cases} 
\mathbf{r}_n = \mathbf{x}_n + s a(n), & n = 1, \ldots, N, \\
\mathbf{r}_k = \bar{\mathbf{x}}_k, & k = 1, \ldots, K,
\end{cases}
\]

(2)

where

- \( \mathbf{r}_n \in \mathbb{R}^{M \times 1}, n = 1, \ldots, N \), are the observed spectral data containing potential targets (i.e., anomalies), and \( \bar{\mathbf{r}}_k \in \mathbb{R}^{M \times 1}, k = 1, \ldots, K \) are training data;
- \( s \in \mathbb{R}^{M \times 1} \) denotes the spectral signature and \( a(n), n = 1, \ldots, N \), are the signal amplitudes;
- \( M \) and \( N \) denote the number of spectral bands and the number of PUTs, respectively, \( K \) is the number of available training data with \( K + N > M \);
- the training data \( \bar{\mathbf{r}}_k \in \mathbb{R}^{M \times 1}, k = 1, \ldots, K \), and the residual background vectors \( \mathbf{x}_n \in \mathbb{R}^{M \times 1}, n = 1, \ldots, N \), are assumed to be independent identically distributed (i.i.d.) samples from a real-valued multivariate zero-mean Gaussian distribution with covariance matrix \( \mathbf{M} \). That is

\[
\begin{cases} 
\mathbf{x}_n \sim \mathcal{N}(0_{M,1}, \mathbf{M}), & n = 1, \ldots, N, \\
\bar{\mathbf{r}}_k \sim \mathcal{N}(0_{M,1}, \mathbf{M}), & k = 1, \ldots, K.
\end{cases}
\]

The covariance matrix \( \mathbf{M} \) is assumed to be unknown and needs to be estimated.

By defining

\[
\begin{align*}
\mathbf{R} &= [\mathbf{r}_1 \mathbf{r}_2 \ldots \mathbf{r}_N] \in \mathbb{R}^{M \times N}, \\
\mathbf{X} &= [\mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_N] \in \mathbb{R}^{M \times N}, \\
\mathbf{R}_L &= [\bar{\mathbf{r}}_1 \bar{\mathbf{r}}_2 \ldots \bar{\mathbf{r}}_K] \in \mathbb{R}^{M \times K}, \\
\mathbf{X}_L &= [\bar{\mathbf{x}}_1 \bar{\mathbf{x}}_2 \ldots \bar{\mathbf{x}}_K] \in \mathbb{R}^{M \times K}, \\
\mathbf{a} &= [a(1) a(2) \ldots a(N)]^T \in \mathbb{R}^{N \times 1},
\end{align*}
\]

we can recast the binary hypothesis test in a compact form:

\[
\begin{cases} 
H_0 : \mathbf{R} = \mathbf{X}, & \mathbf{R}_L = \mathbf{X}_L, \\
H_1 : \mathbf{R} = \mathbf{X} + s \mathbf{a}^T, & \mathbf{R}_L = \mathbf{X}_L,
\end{cases}
\]

(4)

where the vector \( \mathbf{a} \) stands for the known spatial pattern of targets, and the spectral signature \( s \) is unknown.

Note that when \( K = 0 \) (which means that the training data are not available), the model stated above is exactly the one considered by Reed and Yu in [18]. In [18], where the generalized likelihood ratio test (GLRT) criterion was used to derive the well-known RXD, which takes the form

\[
T_{\text{RXD}} = \frac{(\mathbf{R}_L^T \mathbf{R} \mathbf{R}_L)^{-1} (\mathbf{R}_L^T \mathbf{R} \mathbf{a})}{\mathbf{a}^T \mathbf{a}} \geq \lambda_{\text{RXD}},
\]

(5)

where \( \lambda_{\text{RXD}} \) is a detection threshold. It should be pointed out that the constraint \( M < N \) has to be satisfied in the RXD, which might be restrictive in practice.

To the best of our knowledge, the detection problem in (4) for the case of \( K > 0 \) has not been considered before. The purpose of this work is to address the anomaly detection problem by exploiting training data.

III. DETECTOR DESIGN

In this section, we derive the GLRT, the Rao and Wald tests for the detection problem in (4). As to be shown at the end of this section, these tests coincide with each other.

A. GLRT

In the GLRT procedure, we replace the unknown parameters by their maximum likelihood (ML) estimates under each of the
two hypotheses in the likelihood ratio [9], [24], [25], that is

\[ T_{\text{GLRT}} = \frac{\max_{(s,M)} f_2(R, R_L)}{\max_{(s,M)} f_0(R, R_L)} \geq \lambda_G, \]  

where \( \lambda_G \) is a detection threshold, \( f_0(R, R_L) \) and \( f_1(R, R_L) \) are the joint probability density functions (PDFs) of \( R \) and \( R_L \), under hypothesis \( H_0 \) and \( H_1 \), respectively.

According to the model in (4), the joint PDF \( f_0(R, R_L) \) under hypothesis \( H_0 \) can be written as

\[ f_0(R, R_L) = (2\pi)^{-\frac{M(N+K)}{2}} |M|^\frac{N+K}{2} \times \exp \left\{ -\frac{1}{2} \text{tr} \left[ M^{-1} \left( RR^T + R_L R_L^T \right) \right] \right\}. \]  

(7)

Thus, the ML estimate of the unknown covariance matrix \( M \) under hypothesis \( H_0 \) can be calculated as

\[ \hat{M}_0 = \frac{RR^T + R_L R_L^T}{N + K}. \]  

(8)

Similarly, the joint PDF \( f_1(R, R_L) \) under hypothesis \( H_1 \) takes the form

\[ f_1(R, R_L) = (2\pi)^{-\frac{M(N+K)}{2}} |\hat{M}_1|^\frac{N+K}{2} \times \exp \left\{ -\frac{1}{2} \text{tr} \left[ \hat{M}_1^{-1} \left( R - s a^T \right) \left( R - s a^T \right)^T + M^{-1} R_L R_L^T \right] \right\}. \]  

(9)

After some algebra, one can verify that the ML estimates of \( M \) and \( s \) under hypothesis \( H_1 \) are

\[ \hat{M}_1 = \frac{(R - s a^T)(R - s a^T)^T + R_L R_L^T}{N + K}, \]  

(10)

and

\[ \hat{s}_1 = \frac{Ra}{a^T a}, \]  

(11)

respectively.

Substituting these ML estimates of \( M \) into the joint PDFs \( f_0(R, R_L) \) and \( f_1(R, R_L) \) in (7) and (9), respectively, we have the following identities

\[ f_0(R, R_L) = (2\pi)^{-\frac{M(N+K)}{2}} |\hat{M}_0|^\frac{N+K}{2} \exp \left\{ -\frac{M(N+K)}{2} \right\}, \]  

(12)

and

\[ f_1(R, R_L) = (2\pi)^{-\frac{M(N+K)}{2}} |\hat{M}_1|^\frac{N+K}{2} \exp \left\{ -\frac{M(N+K)}{2} \right\}. \]  

(13)

Thus, plugging (12) and (13) into the GLRT defined in (6) and neglecting the exponent \((N+K)/2\), result in

\[ T_{\text{GLRT}} = \frac{\hat{M}_1}{\hat{M}_0} \geq \lambda_{\text{GLRT}}, \]  

(14)

where \( \lambda_{\text{GLRT}} = \lambda^2_{\text{G}} \). Finally, a substitution of (11) into (14) yields the explicit test

\[ T_{\text{GLRT}} = \frac{RR^T + R_L R_L^T}{RR^T + R_L R_L^T - \frac{Raa^T R^T}{a^T a}} \geq \lambda_{\text{GLRT}}. \]  

(15)

Note that when the training data are not available (i.e., \( K = 0 \)), the term \( R_L R_L^T \) disappears. As a result, the GLRT with \( K = 0 \) in (15) reduces to the RXD.

B. Rao Test

Define a parameter vector

\[ \theta = \left[ \theta_T, \theta_s^T \right]^T = \left[ s^T, \text{vec}(M) \right]^T, \]  

(16)

where \( \theta_s = s \) is the relevant parameter and \( \theta_a = \text{vec}(M) \) is the nuisance parameter. The real-valued Rao test with nuisance parameters is given by [26]

\[ T_{\text{Rao}} = \frac{\partial \ln f_1(R, R_L)}{\partial \theta_r} \bigg|_{\theta = \tilde{\theta}_0} \left[ \Gamma^{-1}(\tilde{\theta}_0) \right]_{\theta_\theta, \theta_r, \theta_r}, \]  

(17)

where \( \lambda_{\text{Rao}} \) is the detection threshold, \( \tilde{\theta}_0 = \left[ \theta_T, \theta_s^T \right] \) is the ML estimate of \( \theta \) under hypothesis \( H_0 \), \( I(\theta) \) is the Fisher information matrix (FIM), defined as

\[ I(\theta) = E \left[ \frac{\partial \ln f(R; s, \Sigma)}{\partial \theta} \frac{\partial \ln f(R; s, \Sigma)}{\partial \theta^T} \right], \]  

(18)

which can be partitioned into the following form

\[ I(\theta) = \begin{bmatrix} I_{\theta_\theta, \theta_\theta} & I_{\theta_\theta, \theta_s} \\ I_{\theta_s, \theta_\theta} & I_{\theta_s, \theta_s} \end{bmatrix}. \]  

Thus, the term \([\Gamma^{-1}(\theta)]_{\theta_\theta, \theta_r}\), in (17) can be expressed as

\[ [\Gamma^{-1}(\theta)]_{\theta_\theta, \theta_r} = \begin{bmatrix} I_{\theta_\theta, \theta_\theta}^{-1} & I_{\theta_\theta, \theta_s} \\ I_{\theta_s, \theta_\theta} & I_{\theta_s, \theta_s} \end{bmatrix}. \]  

(20)

Taking the logarithm of (9) and calculating its derivative with respect to \( \theta_r \) and \( \theta_T \), we have the following two identities

\[ \frac{\partial \ln f_1(R, R_L)}{\partial \theta_r} = M^{-1} \left( R - sa^T \right) a; \]  

(21)

and

\[ \frac{\partial \ln f_1(R, R_L)}{\partial \theta_T} = a^T \left( R - sa^T \right)^T M^{-1}. \]  

(22)

Plugging (21) and (22) into (18) results in

\[ I_{\theta_\theta, \theta_\theta} = a^TMa^{-1}. \]  

(23)

In a similar manner, one can readily verify that in (19), \( I_{\theta_\theta, \theta_s} \) is a null matrix. As a consequence, we have

\[ [\Gamma^{-1}(\theta)]_{\theta_\theta, \theta_s} = I_{\theta_\theta, \theta_s} = \begin{bmatrix} M \\ a^T a \end{bmatrix}. \]  

(24)

Note that the ML estimate of \( \theta \) under hypothesis \( H_0 \) is

\[ \tilde{\theta}_0 = \left[ \theta_T, \text{vec}(\hat{M}_0) \right]^T, \]  

(25)

where \( \hat{M}_0 \) is the ML estimate of the covariance matrix \( M \) under hypothesis \( H_0 \), defined in (8).

Combining (8), (21), (22), (24) and (25) and (17), with the constant scalar dropped, yields the Rao test

\[ T_{\text{Rao}} = \frac{(Ra)^T \left( RR^T + R_L R_L^T \right)^{-1} (Ra)}{a^T a} \geq \lambda_{\text{Rao}}, \]  

(26)

where \( \lambda_{\text{Rao}} = \lambda_{\text{R}}/(N + K) \).
C. Wald Test

The real-valued Wald test is given by [26]

$$T_{\text{Wald}} = a^T \left[ \left( I^{-1}(\hat{\theta}_1)_{\theta, \theta} \right) \right]^{-1} \hat{\theta}_1 \overset{H_0}{\sim} \lambda_w,$$  
(27)

where $\lambda_w$ denotes a detection threshold, $\hat{\theta}_1 = [\theta_{r1,1}, \theta_{s1,1}^T]$ is the ML estimate of $\theta$ under hypothesis $H_1$, $(I^{-1}(\theta)_{\theta, \theta})^{-1}$ is the inversion of (20).

Following similar derivation of FIM for the Rao test, one can verify that

$$\left( \left( I^{-1}(\hat{\theta}_1)_{\theta, \theta} \right) \right) = I_{\theta, \theta}(\hat{\theta}_1).$$  
(28)

Moreover, the ML estimate of $\theta$ under hypothesis $H_1$ is given by

$$\hat{\theta}_1 = \left[ \hat{s}_1^T, \text{vec}^T(\hat{M}_1) \right]^T,$$  
(29)

where $\hat{s}_1$ and $\hat{M}_1$ are the ML estimates of the parameters $s$ and $M$ under hypothesis $H_1$, shown in (10) and (11), respectively.

Plugging (10), (11), (23), (28) and (29) into (27), and putting the scalar into the threshold, we can get the Wald test as

$$T_{\text{Wald}} = \frac{(Ra)^T (RP_a R^T + R_s R_s^T)^{-1} (Ra)}{a^T a} \overset{H_1}{\overset{H_0}{\sim}} \lambda_{\text{Wald}},$$  
(30)

where $\lambda_{\text{Wald}} = \lambda_w/(N+K)$, $P_a^\perp$ is a projection operator, defined as

$$P_a^\perp = I_N - a(a^T a)^{-1} a^T.$$  
(31)

In summary, we have derived the GLRT, Rao test and Wald test for the detection problem in (4). The structures of the three proposed detectors appear different from each other. However, we prove in Appendix A that these three detectors are statistically equivalent. Therefore, the GLRT, Rao test and Wald test coincide for the detection problem in (4). In addition, we can observe that the proposed detector is computationally less efficient than the conventional RXD, which is the cost to achieve detection performance improvements.

IV. ANALYTICAL PERFORMANCE

In this section, we derive analytical expressions of the probability of false alarm (PFA) and probability of detection (PD) for the proposed detectors. Due to the equivalence of the three proposed tests, we consider the Wald test, and derive its analytical performance.

In order to obtain analytical expressions of the PFA and PD for the Wald test in (30), we first derive an explicit form for the Wald test. We start by noting that

$$\begin{align*}
\text{cov}[r_n | H_1] &= M, \quad i = 0, 1, \\
\text{cov}[R_k | H_1] &= M, \quad i = 0, 1, \\
E[R | H_0] &= 0_{M,N}, \\
E[R | H_1] &= sa^T,
\end{align*}$$  
(32)

and $E[X_L | H_0] = 0_{M,K}$, $i = 0, 1$. Perform a whitening procedure on $R$ and $R_k$ by letting

$$Z = [z_1, z_2, \ldots, z_N] = M^{-\frac{1}{2}} R,$$  
(34)

and

$$Z_L = [z_1, z_2, \ldots, z_K] = M^{-\frac{1}{2}} R_L,$$  
(35)

respectively. Then, we have

$$\begin{align*}
\text{cov}[z_n | H_1] &= I_M, \quad i = 0, 1, \\
\text{cov}[z_k | H_1] &= I_M, \quad i = 0, 1.
\end{align*}$$  
(36)

In addition,

$$\begin{align*}
E[Z | H_0] &= 0_{M,N}, \\
E[Z | H_1] &= M^{-\frac{1}{2}} sa^T,
\end{align*}$$  
(37)

and $E[Z_L | H_1] = 0_{M,K}, i = 0, 1$. After the whitening procedure, it is evident that the Wald test in (30) takes the form

$$T_{\text{Wald}} = \frac{a^T Z^T (PD_a^\perp Z^T + Z_L Z_L^T)^{-1} Z}{a^T a} \overset{H_1}{\overset{H_0}{\sim}} \lambda_{\text{Wald}},$$  
(38)

As derived in Appendix A, the Wald test can be written as

$$T_{\text{Wald}} = \frac{\nu^2}{\tau} \Psi^{-1} v_1,$$  
(39)

where $\Psi$ is defined in (60). After the transformations, the covariance matrices of $v_n, n = 1, 2, \ldots, N$ and $v_k, k = 1, 2, \ldots, K$ are the same to those of $z_n, n = 1, 2, \ldots, N$ and $z_k, k = 1, 2, \ldots, K$, whereas the mean values are changed under hypothesis $H_1$. That is

$$\begin{align*}
\text{cov}[v_n | H_1] &= I_M, \quad i = 0, 1, \\
\text{cov}[v_k | H_1] &= I_M, \quad i = 0, 1,
\end{align*}$$  
(40)

and $E[v_n | H_1] = E[v_k | H_1] = 0_{M,1}$, $n = 2, 3, \ldots, N$, $E[v_n | H_1] = E[v_k | H_1] = 0_{M,1}$, $n = 1, 2, \ldots, K$.

For further simplification, the test statistic in (39) can be recast as follows (see Appendix B for the detailed derivations)

$$T_{\text{Wald}} = \frac{\nu}{\tau} H_1 \overset{H_0}{\sim} \lambda_{\text{Wald}},$$  
(42)

where

- the random variables $\nu$ and $\tau$ are independent to each other, and are defined in (76) and (79), respectively;
- the distributions of the numerator $\nu$ under hypotheses $H_0$ and $H_1$ are

$$\nu \sim \begin{cases} 
X_M^2, & \text{under } H_0, \\
X_M^2(\sigma), & \text{under } H_1,
\end{cases}$$  
(43)

where the non-centrality parameter $\sigma$ is defined as the generalized signal-to-noise ratio (GSNR), given by

$$\sigma = E[v_1^T H_1 | E[v_1 | H_1] = (s^T M^{-1} s)^2.$$  
(44)

- the denominator $\tau$ obeys a central Chi-squared distribution with $N + K - M$ degrees of freedom under both hypotheses $H_0$ and $H_1$, that is $\tau \sim X_{M+K-M}^2$.

In the following, we derive the PFA and PD for the test in (42). According to [27, p. 52, corollary 2], we can obtain the PDF of the Wald test under hypothesis $H_1$ as

$$f(T_{\text{Wald}} | H_1) = \frac{x^{\frac{M+K-M}{2}} e^{-\frac{x}{2}} \frac{1}{B}}{B \left( \frac{N+K-M}{2}, M \right) (1 + x)^{\frac{M+K-M}{2}}},$$  
(45)

where $B(a, b)$ denotes the beta function, and $1 F_1(a; b; x)$ is the confluent hypergeometric function. If no signal is present, i.e.,
σ = 0, (45) reduces to the PDF of the Wald test under hypothesis 
H_0, namely,
\[ f(T_{Wald}|H_0) = \frac{x^{\frac{M-2}{2}}I_1(N+K; \frac{M}{2}; 0)}{B\left(\frac{N+K-M}{2}, \frac{M}{2}\right) (1+x)^{\frac{M+K}{2}}} \cdot (46) \]

Hence, in terms of the PDFs in (45) and (46), the PFA and PD can be calculated as
\[ \text{PFA} = \int_{\lambda_{Wald}}^{+\infty} f(T_{Wald}|H_0)dx, \quad (47) \]
\[ \text{PD} = \int_{\lambda_{Wald}}^{+\infty} f(T_{Wald}|H_1)dx, \quad (48) \]
respectively. Note that the expression of the PFA in (47) is independent of noise parameters, thus the proposed detectors bear the CFAR property.

Interestingly, we can derive the PFA and PD in finite-sum forms for particular cases, by using the results in [28]. Specifically, when \( M \) is even, the PFA can be obtained as
\[
\text{PFA} = \frac{\Gamma(\frac{N+K}{2} - i)}{\Gamma\left(\frac{M}{2} - i\right) \Gamma\left(\frac{N+K-M}{2}\right)} \lambda_{Wald}^{\frac{M+K}{2} - 1},
\]
(49)

When \( N + K - M \) is even, the PD can be derived as
\[
\text{PD} = 1 - \lambda_{Wald}^{\frac{M-K}{2}} \sum_{j=1}^{\frac{N+K-M}{2}} \frac{\Gamma\left(\frac{N+K}{2} - j\right)}{\Gamma\left(\frac{M}{2} + j\right) \Gamma\left(\frac{N+K-M}{2} - j + 1\right)} \lambda_{Wald}^{j - 1} \exp\left[\frac{-\sigma}{2(1+\lambda_{Wald})}\right] \sum_{m=0}^{j-1} \frac{\sigma}{2(1+\lambda_{Wald})} m!. \quad (50)
\]

Obviously, these finite-sum expressions are computationally more efficient than the integral forms for performance evaluation.

V. SIMULATIONS AND PERFORMANCE EVALUATIONS

In this section, numerical simulations are conducted to show the performance of the proposed decision schemes.

A. Simulated Data

The simulations are conducted on Gaussian vectors (the noise vectors) with dimension \( M = 6 \). The number of pixels under test \( N \) is assumed to be 9, i.e., \( N = 9 \). The noise covariance matrix \( \Sigma \) is chosen to be \( \Sigma_{i,j} = \rho^{|i-j|} \), where \( \rho = 0.8 \). Note that the GSNR is defined in (44). Without loss of generality, the spectral signature vector \( s \) is selected to be \( s = [1, 1, 1, 1, 1, 1]^T \), and signal pattern vector is set to be \( \alpha = \alpha[1, 1, \ldots, 1]^T \), where the term \( \alpha \) is a positive scalar adopted to control the GSNR.

In the following figures, the lines show the results obtained from the derived theoretical expressions, while the markers show the results from Monte Carlo (MC) counting techniques. To evaluate the detection probabilities and set the detection threshold (for the desired PFA), we resort to \( 10^4 \) and \( 10^7 \) MC trials, respectively.

In Fig. 1, we illustrate the false alarm regulation for the proposed detectors when \( K = 3, 9, 31, 123 \). The agreement between the lines and markers confirms the theoretical result in (47). Furthermore, in order to compare the performance of the proposed detectors and the RXD (i.e., \( K = 0 \)), we plot the detection probability as a function of the GSNR in Fig. 2 for \( K = 0, 3, 9, 31, 123 \). The false alarm probability is set to be \( 10^{-5} \). We can observe that the theoretical results (denoted by the lines) match the MC ones (denoted by the markers) very well. In addition, we can observe that for different values of GSNR, the detection probabilities of the proposed detectors are always higher than those of the RXD.

In Fig. 3, we plot the receiver operating characteristic (ROC) curves of the RXD (i.e., \( K = 0 \)) and the proposed detectors for the cases of \( K = 0, 3, 9, 31, 123 \). The GSNR is fixed to be 15 dB. It can be seen that the proposed detectors outperform the RXD when the training data are employed. Moreover, the
Fig. 3. ROC curves for the fixed GSNR = 15 dB (M = 6, N = 9). The lines denote the theoretical expressions, and the markers denote the results obtained from Monte Carlo simulations.

Fig. 4. PD for PFA = 10^{-5} corresponding to different values of K = 3, 9, 17, 35, 71, 123, 139 when GSNR = 15 dB and 18 dB (M = 6, N = 9). The lines denote the theoretical expressions, and the symbols denote the results obtained from Monte Carlo simulations.

detection performance of the proposed detectors becomes better as the number of training data K increases, since the estimation accuracy of the background covariance matrix improves.

The detection probability as a function of the training data number K for the different GSNR is presented in Fig. 4, where the PFA is fixed to be 10^{-5}. We can observe that as the number of training data increases, the performance of the proposed detector becomes better.

B. Synthetic Hyperspectral Image

The synthetic hyperspectral data are applied to conduct experiments for the performance assessment of the proposed detectors. The hyperspectral data we used are the images of the Pavia City of Italy\(^2\) collected by ROSIS-03, displayed in Fig. 5. These images are constituted of 610 × 340 pixels and 103 continuous spectral bands. We extract a small part of the images to perform the experiments. The location is marked out by a white rectangle in Fig. 5. The extracted part consists of 32 × 32 pixels. In order to avoid the well-known problem caused by high dimensionality [15], we choose the first continuous 6 bands, i.e., M = 6.

Since raw hyperspectral data is often non-Gaussian [18], [29], [30], one simple method to address the problem is to perform a local demeaning using a sliding window [8], [23], [31]. The classical Gaussianity test “Q-Q plot” for the 32 × 32 image over the first band indicates that the residual images are approximately Gaussian.

The training data are generally obtained by using a guard window method [8], [23]. We demonstrate the procedure in Fig. 6. The covariance estimation window (the outer solid rectangle in Fig. 6) is adopted to collect the training data in a small neighborhood of the PUTs, and the guard window (the dashed rectangle in Fig. 6) excludes potential target pixels near the PUTs. As a consequence, the pixels between the guard window and the covariance estimation window are adopted as the training

\(^2\)The PaviaU hyperspectral data can be downloaded at the website: www.ehu.eus/ccwintco/uploads/ee/PaviaU.mat
Fig. 7. Experiment results of the RXD and the proposed detectors for different values of GSNR ($M = 6, N = 9, K = 32, PFA = 10^{-5}$). The pixels with values below the threshold are set to zero. (a) anomaly location. (b) and (c) the detection results of the RXD and the proposed detectors when GSNR = 16.19 dB. (d) and (e) the detection results of the RXD and the proposed detectors when GSNR = 22.21 dB. (f) and (g) the detection results of the RXD and the proposed detectors when GSNR = 24.15 dB.

Fig. 8. (a) The false color image of the Nuance Cri hyperspectral data. (b) The test result of the RXD. (c) The test result of the proposed detector.

That is, in our experiments, the signal pattern vector $\mathbf{a}$ is $\mathbf{a} = \alpha \mathbf{a}_0$, where $\mathbf{a}_0 = [0, 1, 0, 1, 1, 0, 1, 0]^T$, the scalar $\alpha$ is used to control the GSNR. It should be mentioned that the background covariance matrix $\mathbf{M}$ of the residue subimages is unknown in advance. Without loss of generality, we replace the unknown covariance matrix $\mathbf{M}$ by its ML estimate $\hat{\mathbf{M}}$. Thus, the GSNR can be written as

$$\text{GSNR} = \alpha^2 |\mathbf{a}_0|^2 (\mathbf{s}^T \hat{\mathbf{M}}^{-1} \mathbf{s}).$$

(51)

We consider three different GSNR values: 1) GSNR = 16.19 dB; 2) GSNR = 22.21 dB; 3) GSNR = 24.15 dB.

For comparison purposes, the RXD and the proposed detectors are assessed in our experiments. First, we compute by (47) to determine the theoretical threshold for a given $PFA = 10^{-5}$. Then, after the test statistic is calculated for each pixel, we set the pixels with values below the threshold to be zero. The results are shown in Fig. 7, where Fig. 7(a) illustrates the location of the implanted anomaly. Figs. 7(c), (e) and (g) demonstrate that the anomaly can be successfully detected by the proposed detectors when the GSNR is set to be 16.19 dB, 22.21 dB and 24.15 dB, respectively. In contrast, in Figs. 7(b), (d) and (f), we can observe that the anomaly target can only be properly detected by the RXD when the GSNR is relatively high, i.e., 24.15 dB, in the three given GSNR values. Hence, we can conclude that the proposed detectors outperform the RXD, due to the use of training data.

C. Real Hyperspectral Data

We next conduct experiments on the Nuance Cri hyperspectral data which comprise of $400 \times 400$ pixels and 46 continuous spectral bands in the wavelengths from 650 to 1100 nm [6], [34]. The background of the Nuance Cri hyperspectral image is grass and 10 rocks in the image are considered as anomalies to be detected.
First, we implement the local demeaning procedure on the Nuance Cri hyperspectral data. Next, we determine the spatial pattern vector $a$ according to the ground-truth of the anomalies. For example, for the anomaly marked by the white rectangle in Fig. 8(a), the ground-truth can be described by a matrix with 0 and 1 elements, shown in Fig. 9. Thus, the spatial pattern vector $a$ can be acquired by reshaping the matrix in Fig. 9 into a column vector (in this case, $N = 108$ and $M = 46$). Note that the spatial pattern vectors $a$ of the 10 anomalies are different, and each of them can be obtained by using this approach. In our experiments, training data are selected by using the guard window method mentioned in Section V-B. We choose the size of the guard window and of the covariance estimation window to be $24 \times 24$ and $25 \times 25$, respectively. Hence, the training data size is $K = 96$. For the Nuance Cri hyperspectral data, we set the PFA to be $10^{-4}$.

For comparison purposes, the performance of the RXD and the proposed detectors are evaluated in Fig. 8(b) and Fig. 8(c), respectively. It can be seen that the proposed detector can successfully detect all the anomalies, while the RXD without training data cannot detect any of them. This result can be explained by the fact that the former exploits the training data for covariance matrix estimation.

VI. CONCLUSION

We have considered the anomaly detection problem for widespread targets (anomalies) with a given signal pattern in hyperspectral images when training data are available. The GLRT, Rao test and Wald test have been proposed, and we have proved that they coincide with each other. In addition, we have analyzed the statistical properties of the proposed detectors, and obtained the theoretical expressions for the PFA and PD of the proposed detectors. These expressions reveal that the proposed detectors bear the CFAR against the noise covariance matrix. Numerical examples based on synthetic and real data show that, when the training data are employed, the performance of the proposed detectors, compared to the conventional RXD, can be improved.

APPENDIX A

EQUIVALENCE OF THE GLRT, ROA TEST AND WALD TEST

In this appendix, we prove that the GLRT, the Rao and Wald tests derived for the detection problem (4) coincide with each other.

A. Equivalence of the GLRT and Rao test

First, we prove that the GLRT and the Rao test are statistically equivalent. Define

$$ YY^T = (RR^T + R_a R_a^T) \in \mathbb{R}^{M \times M}. $$

Thus, the GLRT and the Rao test in (15) and (26) can be rewritten as

$$ T_{GLRT} = \frac{|YY^T|}{|YY^T| I_M - (YY^T)^{-1} \text{Ra}(\text{Ra})^T}{a^T a} H_1 \geq H_0 \lambda_{GLRT}, $$

and

$$ T_{Rao} = \frac{(\text{Ra})^T (YY^T)^{-1} (\text{Ra})}{a^T a} H_1 \geq H_0 \lambda_{Rao}, $$

respectively.

To further simplify (53), we can factor out the determinant of the $M \times M$-dimension matrix $YY^T$ in the denominator. This yields

$$ T_{GLRT} = \frac{|YY^T|}{|YY^T| I_M - (YY^T)^{-1} \text{Ra}(\text{Ra})^T}{a^T a} H_1 \geq H_0 \lambda_{GLRT}, $$

$$ = \frac{1}{1 - (\text{Ra})^T (YY^T)^{-1} (\text{Ra})}{a^T a} H_1 \geq H_0 \lambda_{GLRT}, $$

$$ = \frac{1}{1 - T_{Rao}}. $$

Evidently, the GLRT is equivalent to the Rao test.

B. Equivalence of the Wald and Rao test

Next, we show that the Wald test is equivalent to the Rao test. Similar to the derivations in (34)-(38), we perform a whitening procedure on $R$ and $X_L$. After the whitening procedure, the Wald test can be written as (38). Notice that $a^T a$ is a positive scalar, thus we can normalize the vector $a$ by letting

$$ \tilde{a} = \frac{a}{(a^T a)^{\frac{1}{2}}}. $$

Evidently the Wald test becomes

$$ T_{Wald} = \tilde{a}^T Z \left( Z \tilde{a}^T + Z \tilde{a} \tilde{a}^T \right)^{-1} Z \tilde{a}. $$

Define an $N \times N$ orthonormal matrix $U = [\tilde{a}, G^T]^T$, where $G$ is an $(N - 1) \times N$ matrix with orthonormal row vectors such that $\tilde{a}^T G^T = 0_{1,N-1}$. Using the transformation $U$, we obtain

$$ \tilde{a}^T U T = [1, 0, \ldots, 0], $$

$$ V \triangleq Z U T = [v_1, v_2, \ldots, v_N], $$

$$ V_L \triangleq Z U L T = [\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_K], $$

$$ Q \triangleq U P_{\tilde{a}} U^T = \left[ \begin{array}{cc} 0 & 0_{1,N-1} \\ 0_{N-1,1} & I_{N-1} \end{array} \right]. $$

As a consequence, the Wald test reduces to

$$ T_{Wald} = v_1^T (V Q V^T + V_L V_L^T)^{-1} v_1. $$

For a further simplification of the test in (59), we separate matrix $V$ into two parts $V = [v_1, V]$, where $V = [v_2, \ldots, v_N]$. Thus we have the following identity

$$ \Psi \triangleq V Q V^T + V_L V_L^T = V V^T + V_L V_L^T, $$

and the Wald test can be expressed as (39).
Applying similar approach, one can readily verify that the Rao test in (26) can be formulated as
\[ T_{Rao} = v_1^T (v_1 \psi_1 + \psi_1^{-1}) v_1. \]  
(61)
According to the matrix inversion lemma [35, p. 18], we have
\[ (v_1 v_1^T + \psi_1)^{-1} = \left( I_M - \frac{\psi_1^{-1} v_1 v_1^T}{1 + v_1^T \psi_1^{-1} v_1} \right) \psi_1^{-1}. \]  
(62)
Finally, a substitution of (62) into the Rao test in (61) yields the expression
\[ T_{Rao} = \frac{v_1^T \psi_1^{-1} v_1}{1 + v_1^T \psi_1^{-1} v_1} = T_{Wald}. \]  
(63)
Thus, the Rao test is equivalent to the Wald test.

In conclusion, we prove that the GLRT, Rao test and Wald test coincide with each other because monotonous transformations do not change the test statistical properties.

**APPENDIX B**

**EQUIVALENT TRANSFORMATION OF THE WALT TEST**

In this appendix, we derive an equivalent form of the Wald test in (39) to facilitate analysis of its statistical properties.

Rewrite the matrix \( \psi \) as \( \psi = \psi W W^T \), where
\[ W = [v_2, \ldots, v_N, \tilde{v}_1, \ldots, \tilde{v}_K] \in \mathbb{R}^{M \times (N+K-1)}. \]  
(64)
Thus, we can rewrite the Wald test in (39) in the following form
\[ T_{Wald} = v_1^T (\psi W W^T)^{-1} v_1 = |v_1|^2 \Delta, \]  
(65)
where \( \gamma = \frac{\psi_1}{|v_1|} \in \mathbb{R}^{M \times 1} \) is a normalized vector, and \( \Delta = \gamma^T (\psi W W^T)^{-1} \gamma \).

Now by conditioning on the vector \( v_1 \), we can treat \( \gamma \) as a constant normalized vector. Thus, we can define an orthonormal matrix \( U_1 \in \mathbb{R}^{M \times M} \) such that [9], [18]
\[ U_1 \gamma = [1, 0, \ldots, 0]^T \in \mathbb{R}^{M \times 1}. \]  
(66)
Applying the transformation \( U_1 \) to matrix \( W \), we define
\[ D = U_1 W = U_1 [v_2, \ldots, v_N, \tilde{v}_1, \ldots, \tilde{v}_K] \in \mathbb{R}^{M \times (N+K-1)}. \]  
(67)
Hence, we can rewrite the test term \( \Delta \) as follows
\[ \Delta = [1, 0, \ldots, 0] (DD^T)^{-1} [1, 0, \ldots, 0]^T. \]  
(68)
We partition the matrix \( D \) into two parts \( D = [d_1, \tilde{D}^T]^T \), where \( d_1 \in \mathbb{R}^{(N+K-1) \times 1} \) and \( D \in \mathbb{R}^{(M-1) \times (N+K-1)} \). Thus the term \( (DD^T)^{-1} \) can be written as
\[ (DD^T)^{-1} = \begin{bmatrix} d_1^T d_1 & d_1^T \tilde{D}^T \\ \tilde{D} d_1 & \tilde{D} \tilde{D}^T \end{bmatrix}^{-1} \notag \]  
\[ \equiv \begin{bmatrix} b_{11} & b_{21} \\ b_{21} & B_{22} \end{bmatrix}, \]  
(69)
where \( b_1 = d_1^T d_1 \) is a scalar, \( b_{21} \) is a column vector of dimension \( M - 1 \), and \( B_{22} \) is an \( (M - 1) \times (M - 1) \)-dimensional matrix.

Applying the partitioned matrix inversion theorem [35, p. 17], the term \( \Delta \) reduces to
\[ \Delta = b_{11} = \frac{d_1^T d_1 - d_1^T \tilde{D}^T (D \tilde{D}^T)^{-1} \tilde{D} d_1}{d_1^T \tilde{D}^T d_1}, \]  
(70)
where \( P^\perp \triangleq I_{N+K-1} - \tilde{D}^T (D \tilde{D}^T)^{-1} \tilde{D} \) is a projection operator such that \( \text{tr}(P^\perp) = N+K-M \). According to the properties of the projection matrix, one can verify that \( P^\perp \) has \( N+K-M \) unity eigenvalues and \( M-1 \) zero eigenvalues. Thus, \( P^\perp \) can be diagonalized as
\[ \Lambda = U_2^T P^\perp U_2 \]  
\[ = \begin{bmatrix} I_{N+K-M} & 0_{N+K-M, M-1} \\ 0_{M-1, N+K-M} & 0_{M-1, M-1} \end{bmatrix}, \]  
(71)
where the orthonormal matrix \( U_2 \) is the modal matrix of \( P^\perp \) [26].

We now proceed by defining
\[ \xi \triangleq \Lambda^\frac{1}{2} U_2^T d_1 = [\xi_1, \ldots, \xi_{N+K-1}]^T \in \mathbb{R}^{(N+K-1) \times 1}. \]  
(72)
It is obvious that the last \( M-1 \) elements in \( \xi \) equal to zero. By fixing \( P^\perp \) temporarily, we have
\[ 1_{\Delta} = d_1^T P^\perp d_1 = \xi^T \xi = \sum_{j=1}^{N+K-M} \xi_j^2. \]  
(73)
Thus, after a substitution of (73) into (65), the Wald test becomes
\[ T_{Wald} = \frac{v_1^T v_1}{\xi^T \xi}. \]  
(74)
In the following, we show that the Wald test in (74) can be expressed as a ratio of two independent Chi-squared distributed random variables. First, note that the distributions of \( v_1 \) under hypotheses \( H_0 \) and \( H_1 \) are
\[ H_0 : v_1 \sim N(0, M \cdot I_M), \]  
\[ H_1 : v_1 \sim N(M^{1/2} (s(a^T a)^{1/2}), I_M). \]  
(75)
Hence, the numerator \( v_1^T v_1 \) is distributed as
\[ \nu \triangleq v_1^T v_1 ~\sim ~ \begin{cases} \chi_M^2, & \text{under } H_0, \\ \chi_M^2(\sigma), & \text{under } H_1, \end{cases} \]  
(76)
where the term \( \sigma \) is the non-centrality parameter, given by
\[ \sigma = E[v_1^T H_1^1 E[v_1 H_1] = (s^T M^{-1} s)|a|^2]. \]  
(77)
Next, notice that the component vectors \( (i.e., v_2, \ldots, v_N, \tilde{v}_1, \ldots, \tilde{v}_K) \) of the matrix \( W \) are independent of each other, and the mean vectors and covariance matrices of each component are shown in (41) and (40), respectively. Thus one can readily verify that the term \( d_1 \) has a zero mean vector and a identity covariance matrix under both hypotheses \( H_0 \) and \( H_1 \), that is \( E[d_1 | H_1] = 0_{N+K-1, 1}, i = 0, 1 \), and \( \text{cov}[d_1 | H_1] = I_{N+K-1, 1}, i = 0, 1 \). Hence, it is easy to find that, conditioned on \( v_1 \) and \( P^\perp \), \( \xi_j, j = 1, \ldots, N+K-M \) are independent zero-mean Gaussian random variables under both hypotheses \( H_0 \) and \( H_1 \), i.e., the conditioned joint PDF is
\[ f_\xi(\xi_1, \ldots, \xi_{N+K-M} | v_1, P^\perp) = N(0_{N+K-M, 1}, I_{N+K-M}). \]  
(78)
Also, due to the whitening procedure, it is clear that \( \xi \) is statistically independent to \( v_1 \) and \( P^\perp \). Thus, the denominator in (74) is Chi-squared distributed with \( N+K-M \) degrees of freedom under both hypotheses \( H_0 \) and \( H_1 \), that is
\[ \tau \triangleq \xi^T \xi \sim \chi_{N+K-M}^2. \]  
(79)
Finally, we can equivalently express the test statistic of the Wald test in (74) as (42).
Weijian Liu (Senior Member, IEEE) received the B.S. degree in information engineering and M.S. degree in signal and information processing, both from Wuhan Radar Academy, Wuhan, China, and the Ph.D. degree in information and communication engineering from National University of Defense Technology, Changsha, China, in 2006, 2009, and 2014, respectively. He is currently a Lecturer at Wuhan Electronic Information Institute. His current research interests include multichannel signal detection, statistical and array signal processing. He is currently an Associate Editor for the Circuits, Systems, and Signal Processing.

Danilo Orlando (Senior Member, IEEE) was born in Gagliano del Capo, Italy, on August 9, 1978. He received the Dr. Eng. Degree (with honors) in computer engineering and the Ph.D. degree (with maximum score) in information engineering, both from the University of Salento (formerly University of Lecce), Italy, in 2004 and 2008, respectively. From July 2007 to July 2010, he worked with the University of Cassino (Italy), engaged in a research project on algorithms for track-before-detect of multiple targets in uncertain scenarios. From September to November 2009, he has been Visiting Scientist at the NATO Undersea Research Centre (NURC), La Spezia (Italy). From September 2011 to April 2015, he has worked at Elettronica SpA and was engaged as System Analyst in the field of Electronic Warfare. In May 2015, he joined Università degli Studi “Niccolò Cusano”, where he is currently Associate Professor. His main research interests are in the field of statistical signal processing and image processing with more emphasis on adaptive detection and tracking of multiple targets in multisensor scenarios. He has held visiting positions at the Department of Avionics and Systems of ENSICA (now Institut Supérieur de l’Aéronautique et de l’Espace, ISAE), Toulouse (France) in 2007 and at Chinese Academy of Science, Beijing (China) in 2017-2019. He is Senior Member of IEEE; he has served IEEE TRANSACTIONS ON SIGNAL PROCESSING as Senior Area Editor and currently is Associate Editor for IEEE Open Journal on Signal Processing, EURASIP Journal on Advances in Signal Processing, and MDPI Remote Sensing. He is also author or co-author of about 110 scientific publications in international journals, conferences, and books.

Hongbin Li (Fellow, IEEE) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, in 1999, all in electrical engineering. From July 1996 to May 1999, he was a Research Assistant in the Department of Electrical and Computer Engineering at the University of Florida. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he is currently the Charles and Rosanna Batchelor Memorial Chair Professor. He was a Summer Visiting Faculty Member at the Air Force Research Laboratory in the summers of 2003, 2004 and 2009. His general research interests include statistical signal processing, wireless communications, and radars. Dr. Li received the IEEE Jack Neubauer Memorial Award in 2013 from the IEEE Vehicular Technology Society, Outstanding Paper Award from the IEEE AFICON Conference in 2011, Provost’s Award for Research Excellence in 2019, Harvey N. Davis Teaching Award in 2003, and Jess H. Davis Memorial Research Award in 2001 from Stevens Institute of Technology, and Sigma Xi Graduate Research Award from the University of Florida in 1999. He has been a member of the IEEE SPS Signal Processing Theory and Methods Technical Committee (TC) and the IEEE SPS Sensor Array and Multichannel TC, an Associate Editor for Signal Processing (Elsevier), IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, as well as a Guest Editor for IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and EURASIP Journal on Applied Signal Processing. He has been involved in various conference organization activities, including serving as a General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing (SAM) Workshop, Hoboken, NJ, June 17-20, 2012. Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi.