A Constrained Sparse-Representation-Based Binary Hypothesis Model for Target Detection in Hyperspectral Imagery

Qiang Ling [ORCID], Yulan Guo [ORCID], Zaiping Lin, Li Liu, and Wei An

Abstract—In this paper, we propose a novel constrained sparse-representation-based binary hypothesis model for target detection in hyperspectral imagery. This model is based on the concept that a target pixel can only be linearly represented by the union dictionary combined by the background dictionary and target dictionary, while a background pixel can be linearly represented by both the background dictionary and the union dictionary. To be physically meaningful, the non-negativity constraint is imposed to the weight vector. To suppress the target signals in the background dictionary, the upper bound constraint is also imposed to the weight vector. These upper bounds are adaptively estimated by the similarities between the atoms in the background dictionary and target. Then, the weight vectors under different hypotheses are recovered by a fast coordinate descent method. Finally, both the residual difference and weight difference between the two hypotheses are used to perform the target detection. An important advantage of the proposed method is the robustness to varying target contamination. Extensive experiments conducted on real and synthetic hyperspectral datasets have demonstrated the superiority of the proposed detector in detection performance and computational cost. Specifically, for the Avon dataset, our method achieves the highest area under the receiver operating characteristic (ROC) curve of 99.91%, and achieves the shortest runtime of 109.76 s.

Index Terms—Binary hypothesis, constrained sparse representation (SR), hyperspectral imagery (HSI), target detection.

I. INTRODUCTION

HYPERSPECTRAL imagery (HSI) has opened up new opportunities for the target detection due to the abundant spatial and spectral information of distinct objects [1]. Unlike

Q. Ling, Z. Lin, and W. An are with the College of Electronic Science, National University of Defense Technology, Changsha 410073, China (e-mail: lq910131@163.com; linzaiping@sina.com; anwei@nudt.edu.cn).

Y. Guo is with the College of Electronic Science, National University of Defense Technology, Changsha 410073, China, and also with the School of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou 510275, China (e-mail: yulan.guo@nudt.edu.cn).

L. Liu is with the College of System Engineering, National University of Defense Technology, Changsha 410073, China, and also with the Center for Machine Vision and Signal analysis, University of Oulu, 90014 Oulu, Finland (e-mail: liuli_nudt@nudt.edu.cn).

the traditional spatial-information-based image processing techniques, the rich spectral information available in HSI data can be used to build more elaborate spatial–spectral models to achieve a more accurate target detection. Hyperspectral target detection aims to locate all instances of an target in a scene with prior spectral characteristics of target. It has been widely used in many civilian and military applications including agricultural estimation [2], mineral exploration [3], food safety and quality monitoring [4], search and rescue [5], and explosive detection [6].

With prior spectral characteristics of target, target detection can be considered as a binary classifier where pixels are labeled as target or background. Therefore, target detection algorithms are typically derived from the binary hypothesis model (i.e., target absent hypothesis and target present hypothesis). According to the model used for spectral variation, classical target detection approaches can be classified into three types: probabilistic model, subspace model, and hybrid model. Probabilistic models assume that the background or target can be modeled with a specific distribution, e.g., multivariate normal distribution [7], elliptically contoured distribution [8], and finite mixture distribution [9]. Well-known probabilistic models include spectral matched filter (SMF) [7] and adaptive coherence estimator (ACE) [10]. The SMF first estimates the background covariance matrix, and then, adopts the generalized likelihood ratio test to perform detection with a single target spectrum. Subspace models assume that the background or target lies in a low-dimensional subspace, but the distribution within the subspace does not matter. Well-known subspace models include orthogonal subspace projection (OSP) [11] and sparsity-based detectors [12], [13]. The OSP assumes that target has some components orthogonal to the background subspace, and detects target by maximizing the signal-to-noise ratio (SNR) in the subspace orthogonal to the background subspace. Hybrid models employ both the probabilistic model and the subspace model. The well-known hybrid models include matched subspace detector (MSD) [14] and target-constrained interference-minimized filter [15]. The MSD is modeled using the target and background subspaces constructed by significant eigenvectors. Meanwhile, it assumes that the binary hypothesis follows a multivariate normal distribution with the same scaled identity covariance matrix but different means [13]. Besides, many nonparametric methods have also been proposed. Examples include kernel-based detectors [16], [17], manifold-based detectors [18], [19], tensor-based

detectors [20], [21], and multitask-learning-based detectors [22], [23].

In recent years, sparse representation (SR) has been successfully applied to HSI target detection [12], [13], [22], [24], [25]. The basic sparsity-based detector [12] uses a sparsity model to represent the test pixel by a few training samples, and directly employs the residuals (i.e., reconstruction errors) to perform the target detection. The simultaneous joint sparsity detector [24] forces the pixels within a small neighborhood to be simultaneously represented by a few common training samples but with different weights. The sparse-representation-based binary hypothesis (SRBBH) detector [13] effectively combines SR and binary hypothesis, then sparsely and separately represents the test pixel with different training samples under different hypotheses. There are two advantages of these sparsity-based detectors, which are as follows: 1) the background data do not have to follow any specific statistical distribution and 2) compared to algorithms using a single target spectrum (such as SMF), the target spectral characteristics are better represented by sparsity-based detectors as the target subspace constructed by prior target spectra is used. These sparsity-based detectors usually use only the residuals to detect targets. Actually, the weight is also an important information for the target detection as it represents the target or background fraction of a test sample.

However, the detection performance of most existing detectors degenerates when the background is contaminated by outliers or pixels with target signal [26]. Due to the limited number of training samples selected in background dictionary, sparsity-based detectors are especially sensitive to target contamination. For probabilistic models, this contamination results in a corruption of the estimated background covariance matrix [27]. Therefore, many robust covariance matrix estimation methods, such as quasi-local estimation [28] and minimum covariance determinant [29], have been proposed. For subspace models, this contamination results in a corruption of the estimated background subspace. The background subspace, which employs the first few principal components to identify the large variance directions, can be robustly estimated by robust principal components methods [30], [31]. The hybrid sparsity and statistics (HSS) detector [25] incorporates the statistical distribution characteristics with the sparse representation theory. To handle target contamination, a purification process is also employed to obtain a pure background dictionary. Another example to alleviate the effect of target contamination is the support vector machine (SVM) [32], which introduces a regularization term with slack variables to avoid overfitting caused by outliers.

In this paper, a robust HSI target detection method is proposed based on constrained sparse representation and binary hypothesis model (CSRBBH). Similar to other sparsity-based detectors, it is a nonparametric model without requiring any background statistical information. Unlike other sparsity-based detectors, two constraints are imposed to the weight vector. First, the non-negativity constraint is imposed for physical meaning. Second, the upper bound constraint is imposed to suppress the target signals in the background dictionary. The motivation lies in the suppression of atoms that are similar to target in the background dictionary. That is, the weights of atoms similar to the target are given small upper bounds, and the weights of atoms different from the target are given large upper bounds. Therefore, adaptive estimation of the upper bound vector is achieved. Another difference between our method and the existing sparsity-based methods is, both the residual difference and weight difference between the two hypotheses are used to detect targets in our detector, resulting in an improved detection performance. Moreover, our sparsity model is solved by a fast dual coordinate descent (DCD) method. The contributions of this paper can be summarized as follows.

1) A CSRBBH model is proposed. Both non-negativity constraint and upper bound constraint are imposed to the weight vector. This model is nonparametric and physically meaningful.
2) Both the recovered residual and weight information are used in the proposed detector. This fully exploits the weight difference and residual difference between the two hypotheses to achieve a better detection performance.
3) The upper bound vector is adaptively estimated. This estimation is invariant to spectral intensity and can suppress target signals in the background dictionary. Consequently, the proposed detector is robust to target contamination.

The rest of this paper is organized as follows. The proposed CSRBBH algorithm and its implementation details are described in Section II. Experiments on real HSI datasets are presented in Section III, followed by the conclusions in Section IV.

II. CSRBBH FOR TARGET DETECTION

In this section, we propose a CSRBBH model for the target detection. First, the proposed CSRBBH model is derived from the SRBBH model with constraints on the weight vector. Then, the adaptive estimation of the upper bound vector is described in details. Finally, the scheme and some implementation tricks of the proposed detector are introduced.

A. SRBBH Model

In the SRBBH model [13], a pure background is assumed to lie in the background subspace, and a pure target or a subpixel target is assumed to lie in a low-dimensional union subspace (which is combined by the target and background subspaces). Therefore, the union of the target and background dictionaries can linearly and sparsely represent a pure target or a subpixel target, while the single background dictionary cannot. This binary hypothesis model is expressed as

$$H_0 : \boldsymbol{y} = \boldsymbol{A}_b \boldsymbol{\alpha} + \boldsymbol{n}_0, \text{ target absent}$$
$$H_1 : \boldsymbol{y} = \boldsymbol{A}_b \boldsymbol{\beta}_b + \boldsymbol{A}_t \boldsymbol{\beta}_t + \boldsymbol{n}_1 = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{n}_1, \text{ target present}$$

$$(1)$$

where $\boldsymbol{y}$ denotes the test pixel, $\boldsymbol{A} = [\boldsymbol{A}_b, \boldsymbol{A}_t] \in \mathbb{R}^{M \times (N_b + N_t)}$ denotes the union dictionary consisting of the background dictionary $\boldsymbol{A}_b$ and target dictionary $\boldsymbol{A}_t$, M denotes the number of bands, N_b and N_t denote the number of atoms in $\boldsymbol{A}_b$ and $\boldsymbol{A}_t$, respectively. $\boldsymbol{\alpha}$ is the weight vector corresponding to $\boldsymbol{A}_b$ under hypothesis H_0. $\boldsymbol{\beta}_b$ and $\boldsymbol{\beta}_t$ are the weight vectors corresponding

to $\boldsymbol{A}_b$ and $\boldsymbol{A}_t$ under hypothesis H_1, respectively. $\boldsymbol{\beta} = [\boldsymbol{\beta}_b^{\mathrm{T}}, \boldsymbol{\beta}_t^{\mathrm{T}}]^{\mathrm{T}}$ is a sparse vector concatenated by $\boldsymbol{\beta}_b$ and $\boldsymbol{\beta}_t$. $\boldsymbol{n}_0$ and $\boldsymbol{n}_1$ are noise terms.

With a specific sparsity level, the sparse vector $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be recovered by solving

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{y} - \boldsymbol{A}_b\boldsymbol{\alpha}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}\|_0 \leq K$$

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\beta}\|_0 \leq K \tag{2}$$

where $\|\cdot\|_0$ denotes the l_0 norm, and K is a given upper bound on the sparsity level. These problems can be approximately solved by the orthogonal matching pursuit (OMP) [33] method.

This recovery process implicitly leads to a competition between hypothesis H_0 and hypothesis H_1. The residuals recovered by these two hypotheses are

$$r_0(\boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{A}_b\hat{\boldsymbol{\alpha}}\|_2$$

$$r_1(\boldsymbol{y}) = |\,|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{\beta}}|\,|_2. \tag{3}$$

Then, the class label of the test pixel $\boldsymbol{y}$ can be determined by comparing these residuals

$$D_{\mathrm{SRBBH}}(\boldsymbol{y}) = r_0(\boldsymbol{y}) - r_1(\boldsymbol{y}). \tag{4}$$

B. CSRBBH Model

Similar to the SRBBH algorithm, the binary hypothesis model (1) is also used in our CSRBBH algorithm. According to the linear mixture model (LMM), each mixed pixel can be linearly represented by several endmembers (pure pixels) weighted by their corresponding abundances [34]. These abundances are non-negative and their sum is equal to one. Due to the variation of illumination and atmosphere condition, the spectra of an object exhibit significant intensity variations throughout all spectral bands [35], especially when the prior target spectra are collected from other HSIs or measured by ground experiments. This strong spectral variability is the main criticism about the abundance sum-to-one constraint [36]. Consequently, we only impose the non-negativity constraint on the weight vector. Then, the optimization problem of the hypothesis H_1 is changed to

$$\min_{\boldsymbol{\beta}} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta}\|_2^2$$

$$\text{s.t.} \quad \beta_i \geq 0, \ i = 1, \dots, N \tag{5}$$

where $N = N_b + N_t$ is the number of atoms in the union dictionary $\boldsymbol{A}$. The objective function can be replaced by

$$\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta}\|_2^2 = (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta})$$

$$= \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\beta} - 2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} \tag{6}$$

where $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}$ is a constant and can be removed from the objective function.

Note that, compared to the SRBBH model, the sparsity constraint $\|\boldsymbol{\beta}\|_0 \leq K$ is removed in (5). This is due to the Karush–Kuhn–Tucker (KKT) conditions [37] (which are applied to constrained convex optimization problems). According to the KKT conditions, the sparsity of optimal solution $\boldsymbol{\beta}^*$ can still be guaranteed after removing this sparsity constraint, more details can be found in [38].

In practice, the background dictionary $\boldsymbol{A}_b$ is usually generated locally by a sliding dual window centered at the test pixel $\boldsymbol{y}$ [39], while the target dictionary $\boldsymbol{A}_t$ is usually obtained by a spectrometer in ground experiments or endmember extraction from HSI. However, during the sliding process of the dual window, some target pixels may inevitably fall into $\boldsymbol{A}_b$, especially when targets are densely and evenly distributed in the scene. This will lead to an impure background dictionary for the test pixel and result in a weakened discriminative performance [25]. To avoid this problem, we introduce parameters C_i as the upper bound of β_i to limit the weights of target atoms presented in $\boldsymbol{A}_b$. Then, the optimization problem of hypothesis H_1 can be described as

$$\min_{\boldsymbol{\beta}} \quad \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\beta} - 2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\beta}$$

$$\text{s.t.} \quad 0 \leq \beta_i \leq \begin{cases} C_i & i = 1, \dots, N_b \\ +\infty & i = N_b + 1, \dots, N \end{cases} \tag{7}$$

where $C_i \geq 0$. If all atoms in $\boldsymbol{A}_t$ are set with small C_i, and the spectral intensities of these atoms are much smaller than the test target pixel, the test target pixel cannot be effectively recovered. Therefore, C_i corresponding to the atoms in $\boldsymbol{A}_t$ are set to $+\infty$. The rest C_i are used to construct an upper bound vector $\boldsymbol{C}(\boldsymbol{C} = [C_1, \dots, C_{N_b}]^{\mathrm{T}})$. Apparently, C_i corresponding to target pixels presented in $\boldsymbol{A}_b$ should be small, and C_i corresponding to background pixels should be large. The adaptive estimation of $\boldsymbol{C}$ will be described in Section II-C.

To maintain the comparability between the hypotheses H_0 and H_1, the same upper bound constraints for the atoms in $\boldsymbol{A}_b$ are also used in the hypothesis H_0. Similarly, we can obtain the optimization problem of the hypothesis H_0 as

$$\min_{\boldsymbol{\alpha}} \quad \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}_b^{\mathrm{T}}\boldsymbol{A}_b\boldsymbol{\alpha} - 2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}_b\boldsymbol{\alpha}$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C_i, \ i = 1, \dots, N_b. \tag{8}$$

The optimization problem of (8) can be replaced by

$$\min_{\tilde{\boldsymbol{\alpha}}} \quad \tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\tilde{\boldsymbol{\alpha}} - 2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\tilde{\boldsymbol{\alpha}}$$

$$\text{s.t.} \quad 0 \leq \tilde{\alpha}_i \leq \begin{cases} C_i, & i = 1, \dots, N_b \\ 0, & i = N_b + 1, \dots, N \end{cases} \tag{9}$$

where $\tilde{\boldsymbol{\alpha}}^T = [\boldsymbol{\alpha}^T, \boldsymbol{0}^T]$, $\boldsymbol{0}$ is an $N_t \times 1$ vector of all zeros.

Equations (7) and (9) are very similar to the Lagrange dual problem of the linear SVM [40]. They are quadratic programming problems and can be solved by a standard quadratic programming solver with a computational cost of $O(N^3)$, which is too expensive. Due to the extensive applications of the SVM over past decades, many fast methods have been proposed to solve these problems, such as the cutting plane method [41], the bundle methods [42], and the exponentiated gradient method [43]. In this paper, a DCD method [44] is employed to solve these problems.

Similar to the SRBBH, once the optimal solution $\tilde{\alpha}^*$ and β^* are obtained, the residuals can be calculated by

$$r_0(\boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{A}\tilde{\boldsymbol{\alpha}}^*\|_2 \qquad (10)$$

$$r_1(\boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\beta}^*\|_2 . \qquad (11)$$

Different from the SRBBH, both the residuals and weights are used to detect targets

$$D_{\text{CSRBBH}}(\boldsymbol{y}) = \|\tilde{\boldsymbol{\alpha}}^* - \boldsymbol{\beta}^*\|_1 \, (r_0(\boldsymbol{y}) - r_1(\boldsymbol{y})) \qquad (12)$$

where $\|\cdot\|_1$ denotes the l_1 norm. $\|\tilde{\boldsymbol{\alpha}}^* - \boldsymbol{\beta}^*\|_1$ denotes the weight difference between hypotheses H_0 and H_1, and $r_0(\boldsymbol{y}) - r_1(\boldsymbol{y})$ denotes their residual difference. If $D_{\text{CSRBBH}}(\boldsymbol{y})$ is larger than a given threshold, then $\boldsymbol{y}$ is labeled as target.

If $\boldsymbol{y}$ is a pure background pixel, $\tilde{\boldsymbol{\alpha}}^*$ and $\boldsymbol{\beta}^*$ are very similar, then $\|\tilde{\boldsymbol{\alpha}}^* - \boldsymbol{\beta}^*\|_1$ is close to 0. Meanwhile, both $r_0(\boldsymbol{y})$ and $r_1(\boldsymbol{y})$ are close to 0, hence, the detection value $D_{\text{CSRBBH}}(\boldsymbol{y})$ is close to 0. In contrast, if $\boldsymbol{y}$ is a pure target pixel or a mixed target pixel, $\tilde{\boldsymbol{\alpha}}^*$ and $\boldsymbol{\beta}^*$ are different, then $\|\tilde{\boldsymbol{\alpha}}^* - \boldsymbol{\beta}^*\|_1$ is large. Meanwhile, $r_0(\boldsymbol{y})$ is large, $r_1(\boldsymbol{y})$ is close to 0, hence, the detection value $D_{\text{CSRBBH}}(\boldsymbol{y})$ is much larger than 0. Therefore, the differences between the background pixels and target pixels are enlarged by multiplying the weight difference with the residual difference.

C. Estimation of the Upper Bound Vector

In this section, the upper bound vector $\boldsymbol{C}$ is adaptively estimated to suppress the target signals in the background dictionary $\boldsymbol{A}_b$. As aforementioned, C_i corresponding to target pixels presented in $\boldsymbol{A}_b$ should be small, and C_i corresponding to background pixels presented in $\boldsymbol{A}_b$ should be large. Therefore, a similarity measurement is required to determine which pixels are more likely to be a target. To achieve invariance to spectral intensities, the correlation coefficient [45] ρ is used in this paper

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \frac{(\boldsymbol{x} - \bar{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{y} - \bar{\boldsymbol{y}})}{\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2 \, \|\boldsymbol{y} - \bar{\boldsymbol{y}}\|_2} \qquad (13)$$

where $\rho \in [-1, 1]$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^M$ denotes the M-dimensional spectral vectors, $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ denotes the average value of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. An ρ close to 1 represents a high similarity between $\boldsymbol{x}$ and $\boldsymbol{y}$. This similarity measurement has been widely used in the hyperspectral classification [46]. One of the most important property of the correlation coefficient is its invariance to multiplicative scaling

$$\rho(k_1 \boldsymbol{x} + b_1, k_2 \boldsymbol{y} + b_2) = \rho(\boldsymbol{x}, \boldsymbol{y}) \qquad (14)$$

where $k_1, k_2 > 0$. This property means that the correlation coefficient between two pixels is invariant to spectral intensities, and sensitive to spectral shape.

For each pixel $\boldsymbol{a}_i$ in $\boldsymbol{A}_b$, its correlation coefficients to all prior target spectra are first calculated. Then, due to the multiple type of prior target spectra, the maximum of these correlation coefficients are considered as the similarity of $\boldsymbol{a}_i$ to target

$$s_i = \max\{\rho(\boldsymbol{a}_i, \boldsymbol{a}_j) \mid j = N_b + 1, \dots, N\} \qquad (15)$$

where $i = 1, \dots, N_b$.

Finally, we use a piece-wise function to map this similarity to the upper bound C_i as

$$C_i = \begin{cases} +\infty, & s_i < s_{\min} \\ \dfrac{1}{2\eta N_b} + \dfrac{1}{2\eta N_b[1 + e^{k(s_i - s_0)}]}, & s_{\min} \le s_i \le s_{\max} \\ \dfrac{1}{2\eta N_b}, & s_i > s_{\max} \end{cases} \qquad (16)$$

where $s_0 = (s_{\min} + s_{\max})/2$, $s_{\min}$, $s_{\max}$, and k are given parameters. η denotes the fraction of target atoms in the background dictionary. If $s_{\min} \le s_i \le s_{\max}$, the piece-wise function is a sigmoid function, which is centrosymmetric at point $(s_0, 3/(4\eta N_b))$ and whose bending degree is controlled by k. This sigmoid function is close to $1/(2\eta N_b)$ if s_i is close to $s_{\max}$, and close to $1/(\eta N_b)$ if s_i is close to $s_{\min}$.

If $s_i < s_{\min}$, $\boldsymbol{a}_i$ is supposed to be a background pixel, then C_i is set to $+\infty$. If $s_{\min} \le s_i \le s_{\max}$, $\boldsymbol{a}_i$ can be a target pixel or a background pixel, then the sigmoid function is used to estimate C_i. Usually, due to the similar collection conditions within a small region, the sum-to-one constraint is approximately satisfied when representing a test background pixel. Considering the background dictionary with contamination level η, there are ηN_b target atoms in $\boldsymbol{A}_b$. Apparently, $C_i < 1/(\eta N_b)$, so the sum of C_i corresponding to the target atoms in $\boldsymbol{A}_b$ is smaller than 1. That means a test target pixel cannot be effectively recovered by $\boldsymbol{A}_b$. If $s_i > s_{\max}$, $\boldsymbol{a}_i$ is supposed to be a target pixel, then C_i is set to $1/(2\eta N_b)$ rather than 0. That is because, some background regions in a scene can be very similar to target. If C_i is set to 0, C_i corresponding to the background atoms in these regions are likely all set to 0. Consequently, the test background pixel cannot be effectively represented, resulting in false alarms. If C_i is set to $1/(2\eta N_b)$, the sum of C_i corresponding to the background atoms in these regions can be larger than 1 due to the redundant background information in $\boldsymbol{A}_b$. Consequently, the test background pixel can be recovered by $\boldsymbol{A}_b$. Therefore, the target signals in the background dictionary are suppressed by this estimation.

Parameters $s_{\min}$, $s_{\max}$, and k in the sigmoid function need to be estimated. We have computed the correlation coefficients between same materials and different materials under different data collection platforms and instruments. It can be observed from these data that, samples with correlation coefficients smaller than 0.5 are very likely to be different materials, samples with correlation coefficients larger than 0.9 are very likely to be same materials, and samples with correlation coefficients between 0.5 and 0.9 can be same materials, different materials, or mixed materials. Therefore, $s_{\min}$ and $s_{\max}$ are set to 0.5 and 0.9, respectively. k is set to 20 to ensure the sigmoid function is close to $1/(2\eta N_b)$ at $s_i = 0.9$. Under these parameters, the sigmoid function is shown in Fig. 1.

To detect a target pixel with its spatial neighbors contaminated by target signals, the adaptive estimation of $\boldsymbol{C}$ can give small upper bounds on the weights corresponding to target atoms. Therefore, the test target pixel cannot be effectively represented by the background dictionary. In contrast, the test background pixel can be effectively represented due to the large upper bounds on
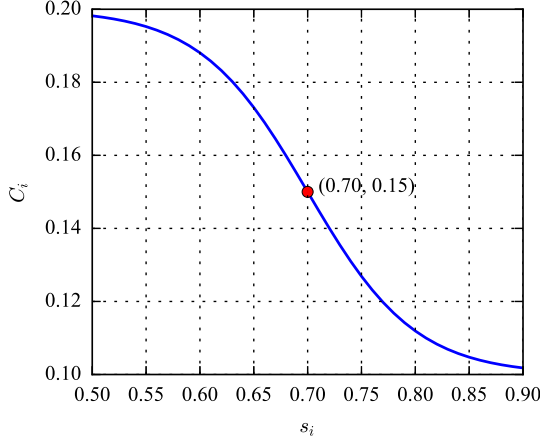
Fig. 1. Sigmoid function with $N_b = 100$, $\eta = 5\%$, $s_{\min} = 0.5$, $s_{\max} = 0.9$, and $k = 20$.

the weights corresponding to background atoms and the redundant background information in the spatial neighborhood. Consequently, although the background dictionary is contaminated by target signals, targets can be detected due to the adaptive estimation of C.

D. Final Scheme for the CSRBBH

The final scheme for the proposed CSRBBH algorithm is described in this section. Some implementation tricks are used to accelerate the CSRBBH. To be more suitable for the HSI target detection, we have slightly modified the DCD method, details are described in the Appendix. The gradient of the objective function is introduced as an additional both input and output parameter in the modified DCD method. It will be used in the following tricks. The semigradients of (7) and (9) can be expressed as

$$\boldsymbol{G}_b = \boldsymbol{Q}\tilde{\boldsymbol{\alpha}} + \boldsymbol{p}, \ \boldsymbol{G} = \boldsymbol{Q}\boldsymbol{\beta} + \boldsymbol{p} \tag{17}$$

where $\boldsymbol{Q} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{p}^{\mathrm{T}} = -\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}$. The computational complexity of $\boldsymbol{Q}$ and $\boldsymbol{p}$ is $O(MN^2)$ and $O(MN)$, respectively.

1) Trick 1: The calculation of C can be simplified. The correlation coefficient between $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ can be replaced as

$$
\begin{aligned}
\rho(\boldsymbol{a}_i, \boldsymbol{a}_j) &= \frac{(\boldsymbol{a}_i - \bar{\boldsymbol{a}}_i)^{\mathrm{T}}(\boldsymbol{a}_j - \bar{\boldsymbol{a}}_j)}{\|\boldsymbol{a}_i - \bar{\boldsymbol{a}}_i\|_2 \|\boldsymbol{a}_j - \bar{\boldsymbol{a}}_j\|_2} \\
&= \frac{Q_{ij} - M\bar{\boldsymbol{a}}_i\bar{\boldsymbol{a}}_j}{\sqrt{(Q_{ii} - M\bar{\boldsymbol{a}}_i\bar{\boldsymbol{a}}_i)(Q_{jj} - M\bar{\boldsymbol{a}}_j\bar{\boldsymbol{a}}_j)}}
\end{aligned}
\tag{18}
$$

where Q_{ij}, Q_{ii}, and Q_{jj} can be extracted from $\boldsymbol{Q}$. $\bar{\boldsymbol{a}}$ denotes the average value of $\boldsymbol{a}$. The computational complexity of all $\bar{\boldsymbol{a}}$ is $O(MN)$. Therefore, the computational complexity of the upper bound vector C is $O(MN + N_t N_b)$.

2) Trick 2: The modified DCD method can be accelerated by a proper initial point. When solving (7) and (9), the DCD requires an initial feasible solution $(\tilde{\boldsymbol{\alpha}}, \boldsymbol{G}_b)$ and $(\boldsymbol{\beta}, \boldsymbol{G})$. As mentioned in the Appendix, the initial feasible solution is usually set to $(\boldsymbol{0}, \boldsymbol{p})$. This initialization strategy is adopted when solving (9)

$$\tilde{\boldsymbol{\alpha}} = \boldsymbol{0}, \ \boldsymbol{G}_b = \boldsymbol{p}. \tag{19}$$

The same initialization strategy can also be adopted when solving (7). In this case, solving (7) and solving (9) have similar computational cost. Using the optimal solution information of (9), we propose a more effective initialization strategy when solving (7)

$$\boldsymbol{\beta} = \tilde{\boldsymbol{\alpha}}^*, \ \boldsymbol{G} = \boldsymbol{G}_b^*. \tag{20}$$

If $\boldsymbol{y}$ is a pure background pixel without any target components, it can and only can be effectively represented by the atoms in $\boldsymbol{A}_b$. Therefore, the optimal solution of (7) and (9) will be the same. In this case, the number of iterations is zero when solving (7). Fortunately, targets are usually small and rarely present in the scene, and most of the pixels in the scene are background pixels. Therefore, the computational cost of solving (7) in the whole scene can be ignored.

3) Trick 3: The computational cost of calculating the residuals can also be reduced. Combining (6), (11), and (17), $r_1(\boldsymbol{y})$ can be calculated in a more efficient way as

$$
\begin{aligned}
r_1(\boldsymbol{y}) &= \sqrt{\boldsymbol{\beta}^{*\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\beta}^* - 2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\beta}^* + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}} \\
&= \sqrt{\boldsymbol{\beta}^{*\mathrm{T}}(\boldsymbol{G}^* + \boldsymbol{p}) + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}.
\end{aligned}
\tag{21}
$$

The computational complexity of (11) is $O(MN)$, such operations are expensive. While the computational complexity of (21) is $O(M + N)$, which is much smaller than $O(MN)$. Similarly, $r_0(\boldsymbol{y})$ can be calculated by

$$r_0(\boldsymbol{y}) = \sqrt{\tilde{\boldsymbol{\alpha}}^{*\mathrm{T}}(\boldsymbol{G}_b^* + \boldsymbol{p}) + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}. \tag{22}$$

As described in the Appendix, the modified DCD method is terminated when the decrease of objective value reaches a tolerance ϵ. Usually, a large tolerance ϵ should be set for a large absolute optimal objective value. To easily set ϵ for different HSI datasets with different ranges of pixel values, we linearly normalize an HSI to interval $(0, 1)$. In this paper, ϵ is set to 10^{-6}. Combining these tricks, the implementation details of the proposed CSRBBH method are presented in Algorithm 1. The modified DCD method has a computational cost of $O(\log(1/\epsilon)N)$. Therefore, the total computational complexity of our CSRBBH algorithm is about $O(MN^2 + \log(1/\epsilon)N)$ for each test pixel. As described in Algorithms 1 and 2, the maximum matrices to save are $\boldsymbol{A}$ and $\boldsymbol{Q}$, which have the space complexity of $O(MN)$ and $O(N^2)$, respectively. Therefore, the space complexity of our CSRBBH algorithm is about $O(MN + N^2)$. Using the kernel trick, the proposed CSRBBH algorithm can be easily extended to a kernel version.

III. EXPERIMENTAL RESULTS

In this section, three widely used HSI datasets are first introduced. Then, the capability of the proposed method to represent a target pixel, a mixed pixel, and a background pixel is analyzed in details. The robustness of the proposed method with respect to target contamination is also demonstrated. Finally, the

---

Algorithm 1: The CSRBBH Algorithm for Target Detection.

Input: Three-dimensional hyperspectral cube, target dictionary A_t, dual window size $(\omega_{in}, \omega_{out})$, and parameter η.

Initialize: Linearly normalize the whole HSI.

for each test pixel y in the scene **do**

  1) Collect A_b based on $(\omega_{in}, \omega_{out})$, $A = [A_b, A_t]$;

  2) $Q = A^T A$, $p^T = -y^T A$. Calculate C by (16);

  3) Initialize the solution of DCD by (19);

  4) Solve (9) and obtain the optimal solution $(\tilde{\alpha}^*, G_b^*)$;

  5) Initialize the solution of DCD by (20);

  6) Solve (7) and obtain the optimal solution (β^*, G^*);

  7) Calculate $r_0(y)$ and $r_1(y)$ by (22) and (21);

  8) Compute the detection result via (12).

end for

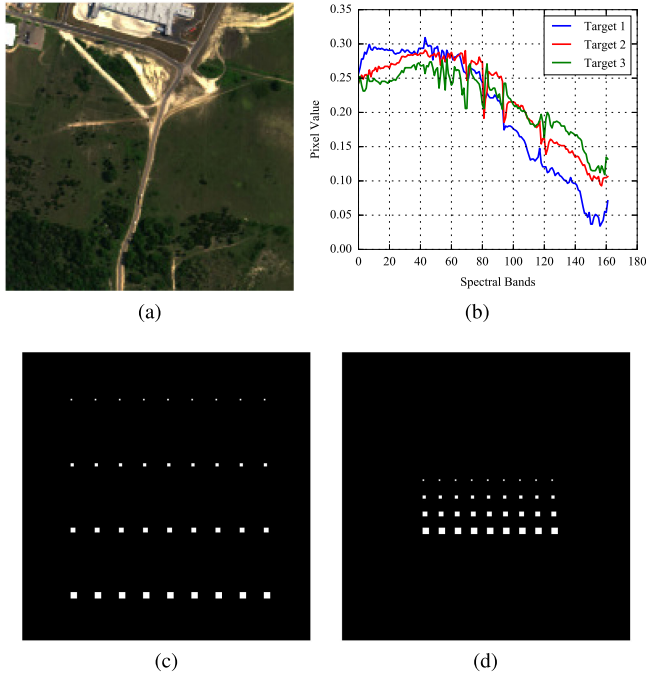**Output:** Target detection map.

Fig. 2. Synthetic Urban dataset. (a) False color image. (b) Prior target spectra. (c) Ground-truth map of Urban-I. (d) Ground-truth map of Urban-II.

detection performance and execution time of the proposed method are compared to several existing methods.

A. Dataset Description and Evaluation Metrics

The first dataset was collected by the Hyperspectral Digital Image Collection Experiment (HYDICE) sensor. The whole dataset has a size of 307×307, while a region in the lower right with a size of 180×180 is selected for experiments, as shown in Fig. 2(a). It covers an urban area, with a spatial resolution of 1 m. This dataset has 210 bands covering the spectral range of 400–2500 nm, with a spectral resolution of 10 nm. After removing the water absorption bands and low SNR bands (1–4, 76, 87, 101–111, 136–153, and 198–210), 162 available bands are remained.
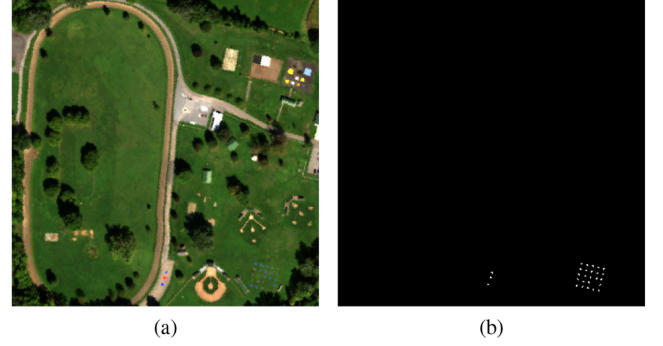

Fig. 3. Avon dataset. (a) False color image. (b) Ground-truth map.

Based on the LMM, 36 target panels constructed by three target spectra are implanted as a grid, which has in four rows and nine columns. These three target spectra are selected from the vehicles in the upper right of the whole scene, as shown in Fig. 2(b). These 36 target panels are implanted into the scene with two spatial distribution densities, namely, Urban-I and Urban-II. In Urban-I, the target panels are sparsely distributed, the distances between neighboring panels within each row and each column are 40 pixels and 15 pixels, respectively, as shown in Fig. 2(c). In Urban-II, the target panels are densely distributed, the distances between neighboring panels within each row and each column are both 10 pixels, as shown in Fig. 2(d). The sizes of the panels in the first, second, third, and fourth rows are 1×1, 2×2, 3×3, and 4×4 pixels, respectively. The first–third columns of these panels are mixed by target 1 and the background, the fourth–sixth columns of these panels are mixed by target 2 and the background, the seventh–ninth columns of these panels are mixed by target 3 and the background. The target abundances of these panels in each column (from left to right) are 10%, 30%, 50%, 10%, 30%, 50%, 10%, 30%, and 50%, respectively. For this dataset, the target dictionary A_t is constructed by these three vehicle spectra.

The second dataset was collected by a pushbroom hyperspectral ProSpecTIR-VS sensor during the "SpecTIR Hyperspectral Airborne Experiment 2012" (SHARE 2012) data collection campaign [47]. From the Avon-morning reflectance data, a region with a size of 330×330 is selected for experiments, as shown in Fig. 3(a). It covers a driving park in Avon, south of Rochester, NY, USA, with a spatial resolution of 1 m. This dataset has 360 bands covering the spectral range of 400–2450 nm, with a spectral resolution of 5 nm. There are 24 tarps and three red or blue felts with 67 target pixels to be detected in the scene, as shown in Fig. 3(b). For this dataset, we select 5 pixels from different targets in the scene to construct the target dictionary A_t.

The third dataset was collected by the Airborne Visible/Infrared Imaging Spectrometer sensor. The whole dataset has a size of 400×400, while a region with a size of 300×300 is selected for experiments, as shown in Fig. 4(a). It covers a naval air station in San Diego, CA, USA, with a spatial resolution of 3.5 m. This dataset has 224 bands covering the spectral range of 370–2510 nm, with a spectral resolution of 10 nm.
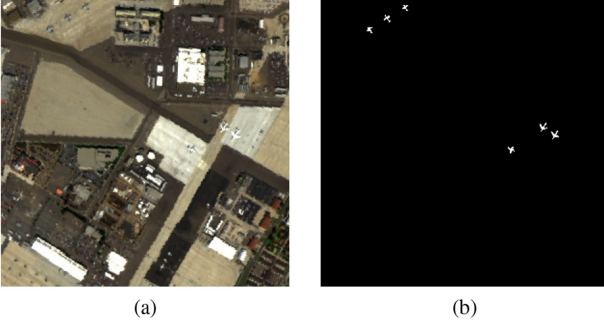
Fig. 4. San Diego dataset. (a) False color image. (b) Ground-truth map.

(a)

(b)

Fig. 5. Background dictionary $\boldsymbol{A}_b$ and target dictionary $\boldsymbol{A}_t$. (a) $\boldsymbol{A}_b$ with 95 background atoms and five target atoms. (b) $\boldsymbol{A}_t$ with five pure target atoms.

After removing the water absorption bands and low SNR bands (1–6, 33–35, 97, 107–113, 153–166, and 221–224), 189 available bands are remained. There are six airplanes with 132 target pixels to be detected in the scene, as shown in Fig. 4(b). For this dataset, we select 6 pixels from different targets in the scene to construct the target dictionary $\boldsymbol{A}_t$.

The proposed CSRBBH algorithm was compared to several existing methods including SMF [48], MSD [14], SVM with composite kernel (SVM-CK) [49], SRBBH [13] integrated with a purified background dictionary (SRBBH-PBD), and HSS integrated with a purified background dictionary (HSS-PBD) [25]. Pixels falling between the inner and outer window regions are used to construct the background dictionary $\boldsymbol{A}_b$. For all detectors, we use the same target dictionary $\boldsymbol{A}_t$ and background dictionary $\boldsymbol{A}_b$ as the prior training samples. For the SMF, the prior target spectrum is determined as the mean of the target atoms in $\boldsymbol{A}_t$. For the MSD, the significant eigenvectors of the target and background covariance matrices are used to generate the target and background subspaces [50]. For the SVM-CK, the composite kernel combines the spectral and spatial features via a weighted summation, where the commonly used Gaussian radial basis function kernel is adopted [12]. For the SRBBH and HSS, a purification process proposed in [25] is applied to the background dictionary to handle target contamination.

The receiver operating characteristic (ROC) curve with confidence intervals of the false alarm rate [51] is used for performance evaluation. According to the detection map, a specific threshold is applied to calculate the detection probability (P_d) and false alarm rate (P_f), which can be used to plot the ROC curve. The ROC curve of a better detector lies close to the upper left corner [52]. For each point on the ROC curve, the confidence interval of false alarm rate can be computed using an appropriate statistical model under a specific significance level. Therefore, each ROC curve is surrounded by these confidence intervals. If one ROC curve of a detector is on the upper left of the other and their confidence intervals do not overlap, it can be considered that this detector statistically performs better. Otherwise, if the confidence intervals are overlapped, there is no statistical difference. In this paper, the significance level is set to 0.05.

However, if the detection probabilities of one detector are higher than that of the other at some false alarm rates, but lower than that of the other at other false alarm rates, the ROC curve
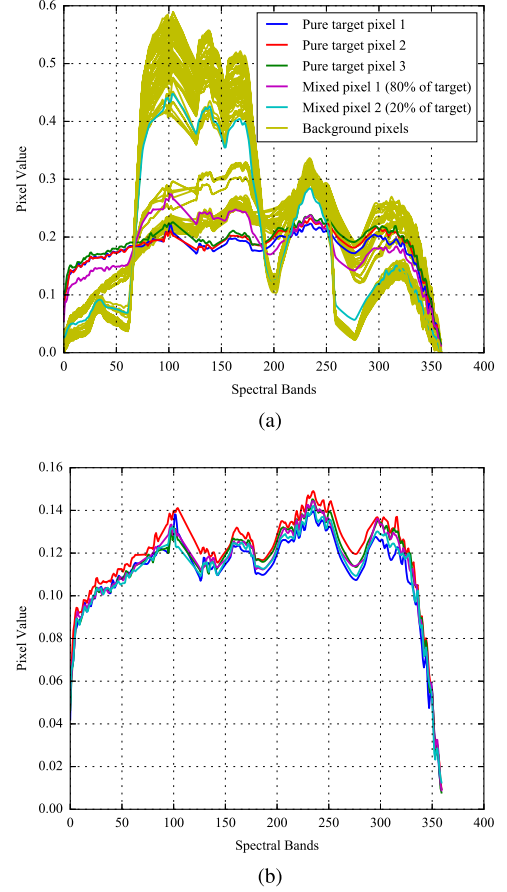
cannot clearly distinguish which detector is better. In this case, it is more proper to use the area under the ROC curve (AUC) [53] to evaluate the comprehensive performance. The AUC value is defined as the area between the ROC curve and the axis of false alarm rate, it can be calculated by a number of trapezoids

$$\text{AUC} = \frac{1}{2} \sum_{i=1}^{n-1} (P_f^{i+1} - P_f^i)(P_d^{i+1} + P_d^i) \quad (23)$$

where $(P_f^i, P_d^i)(i = 1, \ldots, n)$ denotes the ith point on the ROC curve. n is the total number of these points. A larger AUC value indicates a better detection performance.

B. Representation Ability

To illustrate the superiority of the proposed CSRBBH algorithm, an experiment is conducted to reconstruct a target pixel, a mixed target pixel, and a background pixel. First, we select 95 background samples from the boundary area between grass and soil, and three target samples from the sports court area in the Avon dataset. Two mixed target samples are simulated based on the LMM, one with 80% of target and the other with 20% of target. All samples are used to construct the background dictionary $\boldsymbol{A}_b$, as shown in Fig. 5(a). The target samples are located at the 48th, 49th, 50th, 51st, and 52nd
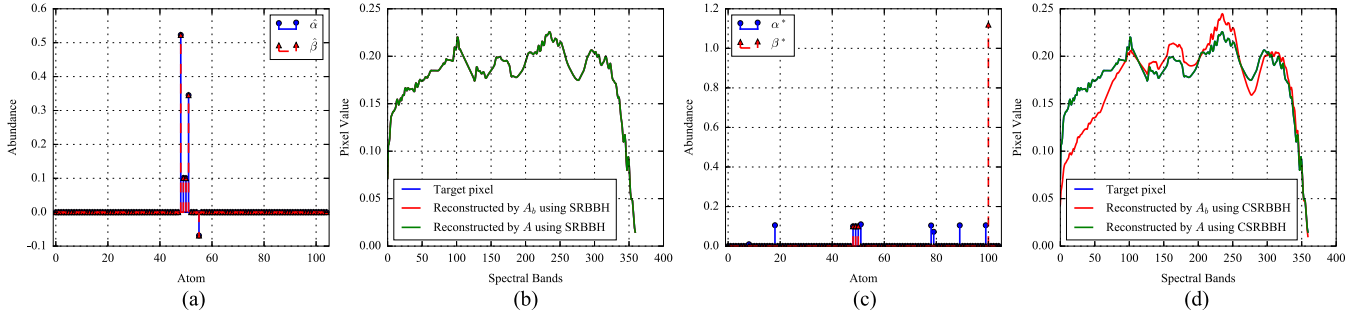
Fig. 6. Example of reconstructing a target pixel by the background dictionary $\boldsymbol{A}_b$ and the union dictionary $\boldsymbol{A}$. (a) Weight vector $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ recovered by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the SRBBH. (b) Target pixel reconstructed by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the SRBBH. (c) Weight vector $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ recovered by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the CSRBBH. (d) Target pixel reconstructed by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the CSRBBH.
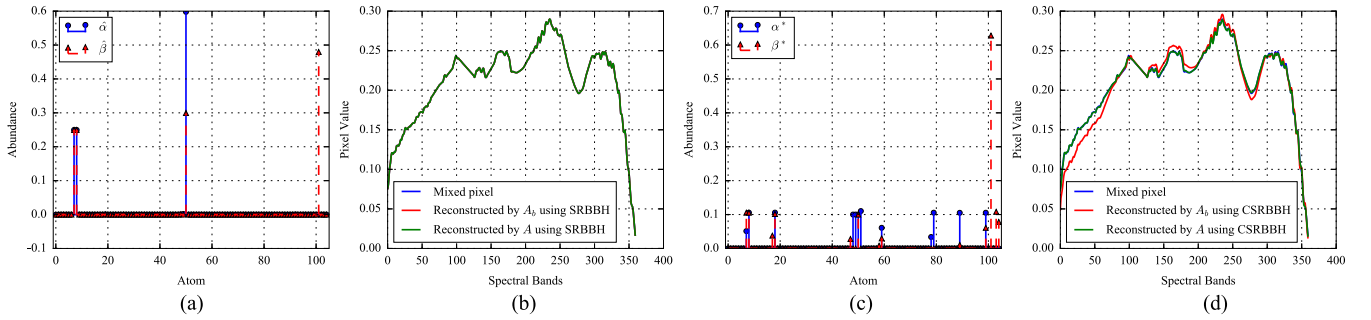
Fig. 7. Example of reconstructing a mixed target pixel by the background dictionary $\boldsymbol{A}_b$ and the union dictionary $\boldsymbol{A}$. (a) Weight vector $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ recovered by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the SRBBH. (b) Mixed pixel reconstructed by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the SRBBH. (c) Weight vector $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ recovered by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the CSRBBH. (d) Mixed pixel reconstructed by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the CSRBBH.
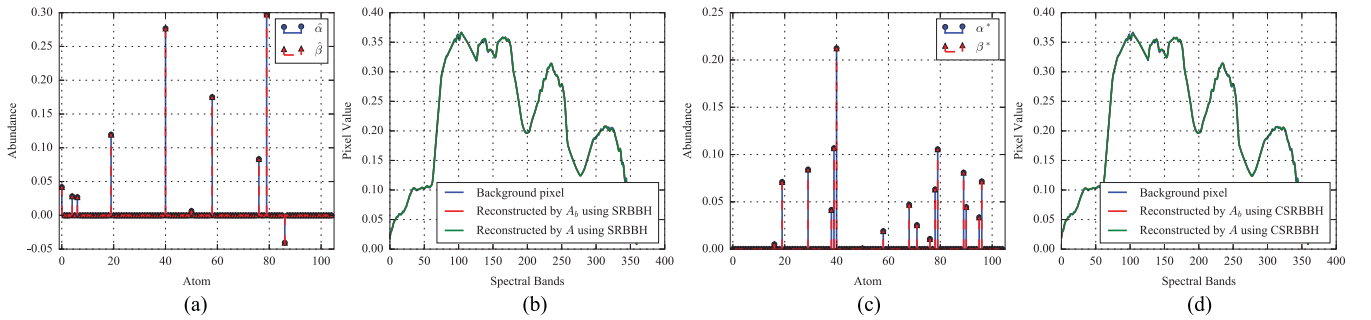
Fig. 8. Example of reconstructing a background pixel by the background dictionary $\boldsymbol{A}_b$ and the union dictionary $\boldsymbol{A}$. (a) Weight vector $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ recovered by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the SRBBH. (b) Background pixel reconstructed by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the SRBBH. (c) Weight vector $\boldsymbol{\alpha}^*$ and $\boldsymbol{\beta}^*$ recovered by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the CSRBBH. (d) Background pixel reconstructed by $\boldsymbol{A}_b$ and $\boldsymbol{A}$ using the CSRBBH.

columns. Then, we select five target samples from the shadow area in the sports court to construct the target dictionary $\boldsymbol{A}_t$, as shown in Fig. 5(b). The test pixels contain a target pixel, a mixed target pixel, and a background pixel, the mixed target pixel consists of 60% of target (the 50th column of $\boldsymbol{A}_b$) and 50% of soil (the eighth column of $\boldsymbol{A}_b$). Finally, the same dictionaries are employed to reconstruct the test samples for the CSRBBH or SRBBH, and their reconstructions are calculated by $\boldsymbol{A}_b \boldsymbol{\alpha}^*$ (or $\boldsymbol{A}_b \hat{\boldsymbol{\alpha}}$) and $\boldsymbol{A} \boldsymbol{\beta}^*$ (or $\boldsymbol{A} \hat{\boldsymbol{\beta}}$). In this experiment, η is set to 5%. The examples of reconstructing a target pixel, a mixed target pixel, and a background pixel are shown in Figs. 6–8, respectively.

First, we demonstrate the effectiveness of the estimation of the upper bound vector $\boldsymbol{C}$. For the 95 background atoms in $\boldsymbol{A}_b$, the estimated upper bounds C_i are mostly $+\infty$. While for the five target atoms in $\boldsymbol{A}_b$, C_i are 0.1, 0.1, 0.1, 0.1102, and 0.1981, respectively. Consequently, a target pixel or a mixed target pixel is represented by $\boldsymbol{A}_b$ using the CSRBBH, the recovered weights of the five target atoms in $\boldsymbol{A}_b$ are smaller than C_i, as shown in Figs. 6(c) and 7(c). In this case, the target pixel and mixed pixel are not well reconstructed by $\boldsymbol{A}_b$, as shown in Figs. 6(d) and 7(d). In contrast, the weights of the SRBBH corresponding to the atoms in $\boldsymbol{A}_b$ are not constrained, thus the target pixel and mixed pixel are well reconstructed by $\boldsymbol{A}_b$, as shown in Figs. 6(b)

and 7(b). This clearly demonstrates the effectiveness of the estimation of C. The upper bounds of the five target atoms in A_t are $+\infty$. Therefore, even the spectral intensities of the target pixel and mixed target pixel are significantly different from A_t, they are well recovered by A using the CSRBBH. In summary, by adaptively estimating C, target signals in the background dictionary are suppressed, so targets can be detected.

Then, we demonstrate the superiority of using both residuals and weights. From Figs. 6(c), 7(c), and 8(c), we can find that, for the example of reconstructing a target pixel or a mixed target pixel using the CSRBBH, the weights recovered by A_b and A have a large difference. While for the example of reconstructing a background pixel, the weights recovered by A_b and A are exactly the same. Specifically, the weight difference $\|\tilde{\alpha}^* - \beta^*\|_1$ for the target pixel and the mixed target pixel is 1.73 and 1.56, respectively. While for the background pixel, the weight difference is zero. Therefore, compared to using (4), the detection value differences between target pixels and background pixels are increased by using (12), resulting in an improved detection performance. As shown in Fig. 8(c), the similar weights recovered by A_b and A also mean that the time consumption of recovering the background pixel by the union dictionary is very small and can be ignored according to Trick 2 (see Section II-D). Due to the huge number of background pixels presented in the scene, the computational cost of the CSRBBH is greatly reduced.

C. Robustness to Target Contamination

In this section, we test the robustness of the proposed CSRBBH algorithm with respect to different target contamination levels. In this experiment, the synthetic Urban-I dataset, Urban-II dataset, and the Avon dataset are used. The AUC performance of our CSRBBH algorithm is further compared to several algorithms dealing with target contamination: SVM-CK, SRBBH-PBD, and HSS-PBD. For the synthetic Urban-I dataset and Urban-II dataset, the inner window is varied from 3 to 7, and the outer window is varied from 11 to 17. According to the ground-truth map, the contamination level for these windows ranges from 0 to 0.1071 for both the Urban-I dataset and Urban-II dataset. For the Avon dataset, the inner window size is varied from 1 to 5, and the outer window size is varied from 7 to 13. According to the ground-truth map, the contamination level for these windows ranges from 0 to 0.0764. In this experiment, η is set to 8% as some inappropriate dual window sizes are adopted. All parameters of other detectors are empirically set to obtain the optimal performance.

The AUC performance achieved on the synthetic Urban-I dataset, Urban-II dataset, and the Avon dataset with different window sizes ($\omega_{\text{in}}, \omega_{\text{out}}$) is shown in Tables I–III, respectively. As expected, the AUC performance of the proposed CSRBBH algorithm is robust to different window sizes. That is attributed to the adaptive estimation of the upper bound vector, which limits the weights of the target atoms in the background dictionary. For the synthetic Urban-I dataset, the AUC performance of the SVM-CK is robust to different window sizes. That is attributed

TABLE I
AUC (IN PERCENT) PERFORMANCE ACHIEVED ON THE SYNTHETIC URBAN-I
DATASET WITH DIFFERENT WINDOW SIZES

Algorithms	ω_{out} / ω_{in}	11	13	15	17
SVM-CK	3	85.09	86.79	87.48	87.75
	5	94.41	95.25	95.52	95.57
	7	96.96	97.54	**97.71**	97.64
SRBBH-PBD	3	74.44	73.00	70.70	71.39
	5	87.56	87.85	84.74	86.19
	7	98.10	**98.38**	96.69	97.50
HSS-PBD	3	77.89	75.60	74.61	74.81
	5	92.17	92.26	89.89	89.96
	7	**99.57**	98.80	98.08	98.40
CSRBBH	3	91.49	92.84	93.25	93.09
	5	98.04	98.32	98.24	98.44
	7	**99.05**	99.01	98.99	99.02

TABLE II
AUC (IN PERCENT) PERFORMANCE ACHIEVED ON THE SYNTHETIC URBAN-II
DATASET WITH DIFFERENT WINDOW SIZES

Algorithms	ω_{out} / ω_{in}	11	13	15	17
SVM-CK	3	85.36	86.63	81.42	69.84
	5	92.94	93.40	87.23	73.81
	7	97.65	**98.00**	90.12	75.70
SRBBH-PBD	3	79.81	78.05	75.73	67.21
	5	94.96	94.11	91.48	80.92
	7	98.85	**98.87**	96.29	84.01
HSS-PBD	3	71.85	73.98	73.06	68.27
	5	91.31	93.32	90.83	78.77
	7	**99.49**	99.17	96.02	83.59
CSRBBH	3	88.52	89.64	87.90	84.37
	5	97.21	97.52	94.99	89.86
	7	99.70	**99.77**	96.76	91.62

TABLE III
AUC (IN PERCENT) PERFORMANCE ACHIEVED ON THE AVON DATASET WITH
DIFFERENT WINDOW SIZES

Algorithms	ω_{out} / ω_{in}	7	9	11	13
SVM-CK	1	96.71	93.76	82.40	71.46
	3	**99.88**	99.74	91.71	76.38
	5	99.73	99.56	91.24	75.36
SRBBH-PBD	1	63.87	63.78	69.81	68.99
	3	77.79	70.53	60.85	66.32
	5	**93.17**	78.60	61.72	66.82
HSS-PBD	1	80.01	82.14	79.10	74.27
	3	99.15	99.16	86.09	74.99
	5	**99.87**	98.75	85.73	74.47
CSRBBH	1	99.60	99.80	99.65	98.82
	3	**99.92**	99.87	99.75	99.36
	5	99.85	99.78	99.59	98.84

to the introduction of slack variables, which can avoid over-fitting caused by outliers. However, for the Avon dataset, the AUC performance of the SVM-CK degenerates rapidly if the outer window size is increased. That is because, the scene of the Avon dataset has lots of small structures, making it difficult for the SVM-CK to accurately model the background when the outer window size is increased. The AUC performance of the SRBBH-PBD and HSS-PBD is not robust to different window sizes due to the ineffectiveness of the purification process when detecting subpixel targets. This purification process first finds all the doubtful target pixels in the whole scene using spectral angle mapper, and then, removes these doubtful target pixels falling into the background dictionary. However, the spectral angle between pure pixel target and subpixel target with low target abundance is usually very large. Consequently, the pixels with low target abundance in the scene are not considered as doubtful targets and cannot be removed from the background dictionary, resulting in a degeneration of the detection performance. In general, compared to the SVM-CK, SRBBH-PBD, and HSS-PBD, the proposed CSRBBH algorithm exhibits a stronger robustness to different window sizes, and achieves the best AUC performance among these detectors. These results have clearly demonstrated the robustness of CSRBBH to various levels of contamination in the background dictionary.

D. Detection Performance Under Different Settings of Parameter η

In this section, we test the detection performance of the proposed CSRBBH algorithm with different values of parameter η. In this experiment, the Avon dataset is used. The proposed CSRBBH algorithm is further compared to several detectors including SMF, MSD, SVM-CK, SRBBH-PBD, and HSS-PBD. The dual window size is set to (3, 9) and (1, 9) to obtain a pure and a contaminated background dictionary. The contamination level for window size (1, 9) is 0.05. η is set to 0% ($C_i = +\infty$), 2%, and 5%, respectively. All parameters of other detectors are empirically set to obtain the optimal performance.

First, we analyze the effects of parameter η on a pure background dictionary. The ROC curves achieved on the Avon dataset with dual window size (3, 9) is shown in Fig. 9. It can be seen that, if η is set to 5%, the detection probability achieved by the CSRBBH is lower than HSS-PBD and MSD when the false alarm rate is lower than 10^{-3} and 2×10^{-4}. However, if η is set to 0% and 2%, the detection probability achieved by the CSRBBH is comparable to HSS-PBD and higher than other detectors when the false alarm rate is lower than 5×10^{-4}. That is because, η denotes the fraction of target atoms in the background dictionary. According to (16), if η is too large, the upper bounds of some background atoms in the background dictionary will be set with small values. When representing a test background pixel, this can result in a larger recovery residual if the fraction of background atoms in the background dictionary is small. Consequently, some background pixels can have large detection values, resulting in a weakened detection performance.
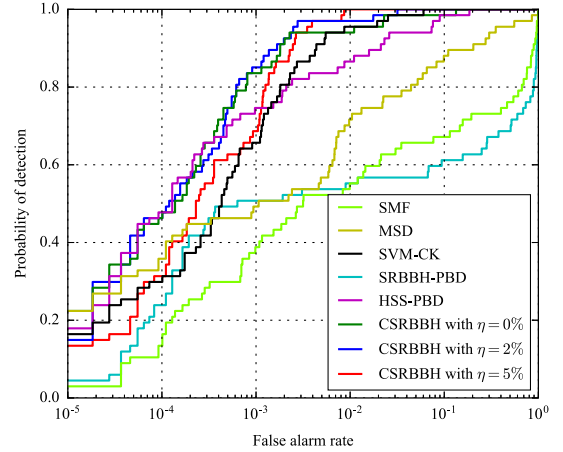

Fig. 9. ROC curves achieved on the Avon dataset with dual window size (3, 9).
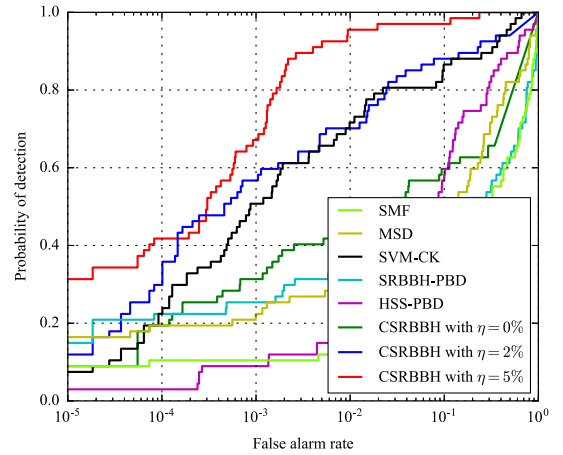

Fig. 10. ROC curves achieved on the Avon dataset with dual window size (1, 9).

Second, we analyze the effects of parameter η on a contaminated background dictionary. The ROC curves achieved on the Avon dataset with dual window size (1, 9) is shown in Fig. 10. It can be seen that, if η is set to 0%, the ROC performance achieved by the CSRBBH is not superior to other detectors. If η is set to 2%, the ROC performance achieved by the CSRBBH is comparable to the SVM-CK and superior to other detectors. However, if η is set to 5%, the detection probability achieved by the CSRBBH is higher than other detectors at all false alarm rates. That is attributed to the adaptive estimation of the upper bound vector. According to (16), if η is set with a large value, the upper bounds of target atoms in the background dictionary will be set with small values. When representing a test target pixel with the background dictionary, this can result in a larger recovery residual. Consequently, the target pixels can have large detection values, resulting in a better detection performance.

E. Comparison to the State-of-the-Art

In this experiment, the proposed CSRBBH algorithm is compared to several detectors including SMF, MSD, SVM-CK, SRBBH-PBD, HSS-PBD, and CSRBBH-NA [CSRBBH with
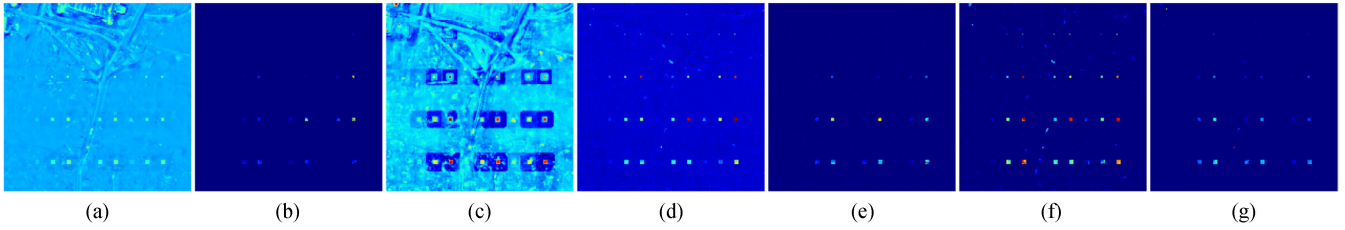
Fig. 11. Detection map achieved on the Urban-I dataset. (a) SMF. (b) MSD. (c) SVM-CK. (d) SRBBH-PBD. (e) HSS-PBD. (f) CSRBBH-NA. (g) CSRBBH.
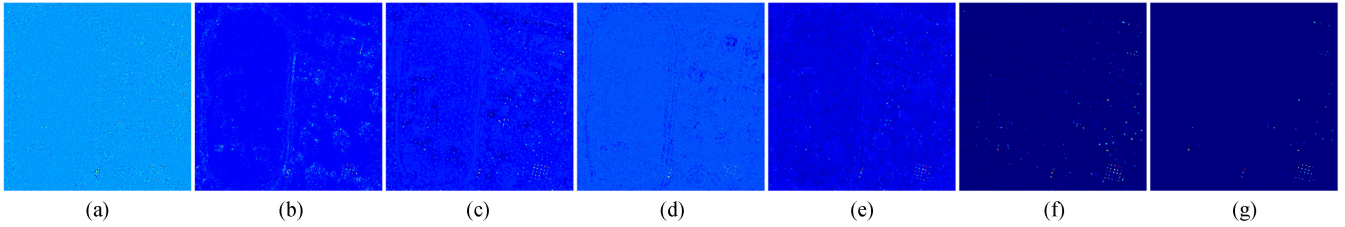

Fig. 12. Detection map achieved on the Avon dataset. (a) SMF. (b) MSD. (c) SVM-CK. (d) SRBBH-PBD. (e) HSS-PBD. (f) CSRBBH-NA. (g) CSRBBH.
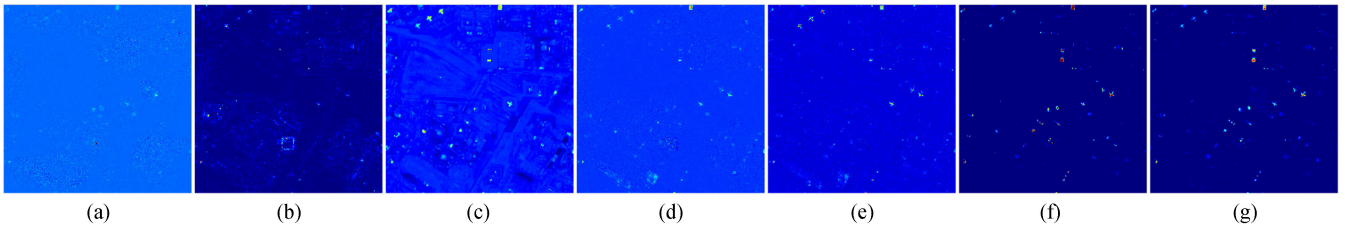

Fig. 13. Detection map achieved on the San Diego dataset. (a) SMF. (b) MSD. (c) SVM-CK. (d) SRBBH-PBD. (e) HSS-PBD. (f) CSRBBH-NA. (g) CSRBBH.

detection discriminant defined in (4)]. For a fair comparison, the inner window size should be set large enough to generate a relatively pure local background. Therefore, the dual window size is set to (7, 15) for the Urban-I dataset, (3, 9) for the Avon dataset, and (15, 21) for the San Diego dataset, resulting in 176, 72, and 216 background training samples in A_b, respectively. Usually, the contamination level in the background dictionary is lower than 5% when a suitable dual window is adopted. Therefore, η is set to 5% in this experiment. All parameters of other detectors are empirically set to obtain the optimal performance.

The detection map achieved on the synthetic Urban-I dataset, the Avon dataset, and the San Diego dataset are shown in Figs. 11(a)–(f), 12(a)–(f), and 13(a)–(f), respectively. From these figures, we can see that all methods except SMF and MSD can successfully distinguish all targets from the background. The detection map of the SMF has more noisy pixels as compared to other detectors, resulting in higher false alarm rates. The MSD produces few background false alarm pixels and achieves a promising background suppression. However, the target pixels are more darker as compared to other detectors, resulting in lower detection probabilities. The SVM-CK achieves a promising target enhancement performance, but still remains a lot of background edges. Some target pixels are not enhanced by the SRBBH-PBD, and some background pixels are

also retained. Compared to SRBBH-PBD, HSS-PBD achieves a better background suppression. As shown in Figs. 11(a), 12(g), and 13(a)–(f), the target pixels of the proposed CSRBBH algorithm are more obvious. Meanwhile, only a few background pixels have high detection values. In summary, our CSRBBH algorithm has achieved good performance in both background suppression and target enhancement. Comparing the detection map of CSRBBH-NA and CSRBBH, we can see that CSRBBH achieves better background suppression. This demonstrates the superiority of using both residuals and weights.

The ROC curves with 95% confidence intervals of the false alarm rate achieved on the three datasets are shown in Fig. 14(a)–(c), and the AUC results are shown in Table IV. We mainly illustrate the superiority of the proposed CSRBBH algorithm by the ROC performance achieved on the Avon dataset. As shown in Fig. 14(b), when the false alarm rate is higher than 2×10^{-3}, the proposed CSRBBH algorithm achieves a much higher detection probability than other traditional detectors. Meanwhile, its confidence intervals are distinctly separated from other traditional detectors. Consequently, the ROC performance of the CSRBBH outperforms other traditional detector when the false alarm rate is higher than 2×10^{-3}. When the false alarm rate is lower than 2×10^{-3}, the ROC performance of MSD, SVM-CK, SRBBH-PBD, HSS-PBD, CSRBBH-NA, and CSRBBH outperform the SMF, but their confidence intervals of the false
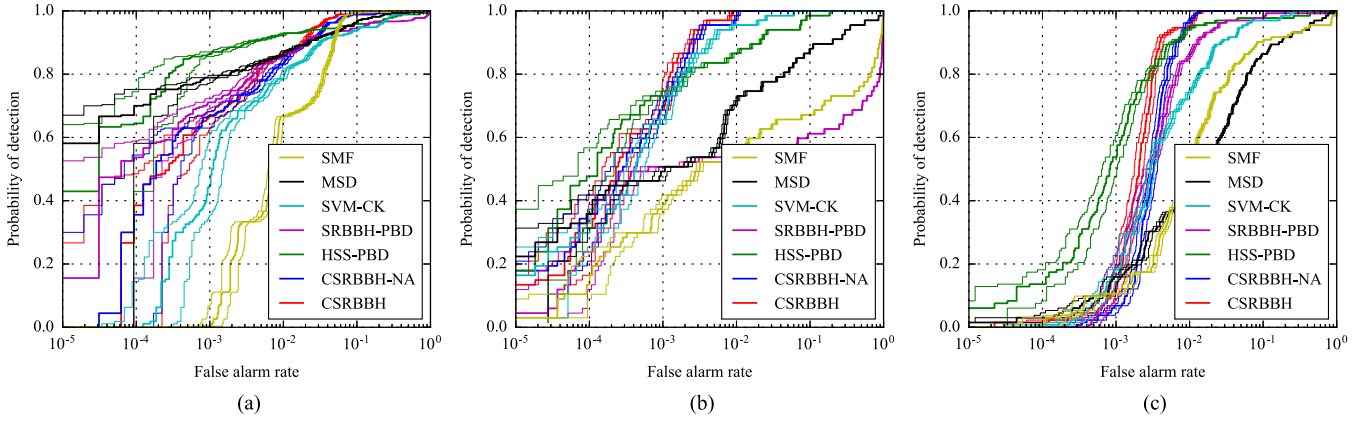
Fig. 14.   ROC curves with 95% confidence intervals of the false alarm rate achieved on the three datasets. (a) Urban-I dataset. (b) Avon dataset. (c) San Diego dataset.

TABLE IV
AUC (IN PERCENT) PERFORMANCE ACHIEVED ON THE THREE DATASETS

Detectors	SMF	MSD	SVM-CK	SRBBH-PBD	HSS-PBD	CSRBBH-NA	CSRBBH
Urban-I Dataset	98.23	98.60	97.71	96.69	98.09	98.90	**99.03**
Avon Dataset	77.63	93.66	99.74	70.53	99.16	99.88	**99.91**
San Diego Dataset	93.23	93.15	98.57	98.62	98.66	99.63	**99.73**

TABLE V
RUNTIMES (IN SECONDS) ACHIEVED ON THREE DATASETS

Detectors	SMF	MSD	SVM-CK	SRBBH-PBD	HSS-PBD	CSRBBH-NA	CSRBBH
Urban-I Dataset	39.40	105.00	94.66	85.01	84.70	**28.67**	28.69
Avon Dataset	501.07	1114.26	227.89	173.09	173.18	**108.38**	109.76
San Diego Dataset	152.08	378.03	358.06	247.69	249.38	**60.94**	61.05

alarm rate are overlapped with each other. That is, there is no statistical difference among these algorithms when the false alarm rate is lower than 2×10^{-3}. Moreover, our CSRBBH algorithm achieves the best AUC performance among all detectors. Specifically, the AUC value achieved by the CSRBBH-NA is 99.88%, which is 29.35% higher than the AUC value achieved by the SRBBH-PBD. This clearly demonstrates the effectiveness of the non-negativity constraint and the estimation of the upper bound vector. Comparing the ROC performance of the CSRBBH-NA and CSRBBH, we can see that CSRBBH achieves a higher detection probability at most false alarm rates. Meanwhile, the AUC value of the CSRBBH is slightly larger than the CSRBBH-NA. This also clearly demonstrates the superiority of using both residuals and weights. Similar conclusions can also be obtained from the experiments conducted on the Urban-I dataset and the San Diego dataset. Note that, for the San Diego dataset, HSS-PBD achieves the best ROC performance when the false alarm rate is lower than 3×10^{-3}. However, when the false alarm rate is higher than 3×10^{-3}, our CSRBBH achieves the best ROC performance. Considering the whole range of the false alarm rate, the AUC value achieved by our CSRBBH algorithm is 1.07% higher than that of the HSS-PBD. In summary, compared to other traditional detectors, our CSRBBH algorithm has achieved the best overall detection performance.

F. Execution Time

The runtime consumed by the aforementioned detectors on three datasets is shown in Table V. All experiments were conducted in Python 3.5 on an PC with an Intel Core i7-3770 CPU and a 8 GB of RAM. Note that, SVM and DCD were implemented in C++, and wrapped into a pyd extension module by SWIG. The inverse operation and eigenvalue decomposition operation were computed by scipy functions, which uses the high-performance LAPACK library. Therefore, the runtime of these algorithms is comparable. On these three datasets, our CSRBBH algorithm costs the shortest execution time. It is faster than the SMF, MSD, and SVM-CK by a large margin. In summary, compared to existing detectors, the proposed CSRBBH algorithm has a relatively lower computational complexity.

IV. CONCLUSION

In this paper, we have proposed a novel target detection algorithm based on cCSRBBH model. The non-negativity constraint is imposed to ensure the physical meaning. The upper bound constraint is imposed to suppress the target signals in the background dictionary. The upper bound vector is adaptively estimated by the similarities between the atoms in the background dictionary and the target dictionary. This CSRBBH model is

solved by a fast DCD algorithm. It can handle target subpixels with different levels of contamination in their surrounding background. The proposed method have been tested on both real and synthetic HSI datasets. Experimental results have clearly demonstrated the superiority of the proposed detector in robustness and detection performance. Moreover, our detector also has a relatively low computational complexity. Specifically, for the Avon dataset, our detector achieves the highest AUC value of 99.91%, and achieves the shortest runtime of 109.76 s.

APPENDIX
DCD METHOD

In this Appendix, we describe a DCD method [44] to solve constrained quadratic programming problems. Consider the following problem:

$$\min_{\boldsymbol{\alpha}} \quad f(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\alpha} + \boldsymbol{p}^{\mathrm{T}}\boldsymbol{\alpha}$$
$$\text{s.t.} \quad 0 \le \alpha_i \le C_i, \ i = 1, \dots, N \qquad (24)$$

where $\boldsymbol{Q}$ denotes an $N \times N$ positive semi-definite matrix, and $\boldsymbol{p}$ denotes the linear term. C_i is the upper bound of α_i, and $\boldsymbol{C} = [C_1, \dots, C_N]$ denotes the upper bound vector. The gradient $\boldsymbol{G}$ of $f(\boldsymbol{\alpha})$ is

$$\boldsymbol{G} = \nabla f(\boldsymbol{\alpha}) = \boldsymbol{Q}\boldsymbol{\alpha} + \boldsymbol{p}. \qquad (25)$$

The DCD method updates only one variable α_i each time by solving the following sub-problem:

$$\min_{d_i} f(\boldsymbol{\alpha} + d_i \boldsymbol{e}_i) \quad \text{s.t.} \quad 0 \le \alpha_i + d_i \le C_i \qquad (26)$$

where $\boldsymbol{e}_i = [0, \dots, 0, 1, 0, \dots, 0]^{\mathrm{T}}$. Therefore, the decrease of the objective function can be derived as

$$Df_i = f(\boldsymbol{\alpha} + d_i \boldsymbol{e}_i) - f(\boldsymbol{\alpha}) = \frac{1}{2}Q_{ii}d_i^2 + G_i d_i \qquad (27)$$

where $f(\boldsymbol{\alpha})$ is a constant and G_i is the ith component of $\boldsymbol{G}$. Equation (26) has an optimum at $d_i = 0$ if and only if $PG_i = 0$, where $\boldsymbol{PG}$ means the projected gradient

$$PG_i = \begin{cases} \min(G_i, 0), & \text{if } \alpha_i = 0 \\ \max(G_i, 0), & \text{if } \alpha_i = C_i \\ G_i, & \text{if } 0 < \alpha_i < C_i. \end{cases} \qquad (28)$$

If $|PG_i| \ne 0$, (26) has an optimum at

$$d_i = \min(\max(-G_i/Q_{ii}, -\alpha_i), C_i - \alpha_i). \qquad (29)$$

The original DCD method is designed for the large-scale sparse dataset. The implementation details can be found in [44, Algorithm 3]. In experiments, we found that it is very slow for the processing of HSI dataset. Therefore, we modify this method as follows.

1) *Working Set Selection*. Instead of permuting $k(k = 1, \dots, N)$ randomly and updating all α_k in each iteration t, we only update α_i where i is the index that makes the objective function decrease fastest.
2) *Update Method*. Instead of updating G_k and PG_k before the update of α_k, we update vector $\boldsymbol{G}$ and vector $\boldsymbol{PG}$ after the update of α_i.

Algorithm 2: The Modified DCD Method.

Input: $\boldsymbol{Q}, \boldsymbol{p}, \boldsymbol{C}, \boldsymbol{\alpha}, \boldsymbol{G}, \epsilon$, and $t_{\max}$.
Initialize: $t = 0, \boldsymbol{d} = \boldsymbol{0}, \boldsymbol{Df} = \boldsymbol{0}$, initialize $\boldsymbol{PG}$ by (28).
while $t < t_{max}$ **do**
  1)   **for** $k = 1, \dots, N$ **do**
        if $|PG_k| \ne 0$ **and** $C_k > 0$,
          $d_k = \min(\max(-G_k/Q_{kk}, -\alpha_k), C_k - \alpha_k)$
          $Df_k = Q_{kk}d_k^2 + 2G_k d_k$
    end for
  2) $i = \arg \min_{k}\{Df_k \mid k = 1, \dots, N\}$
  3)   **if** $|Df_i| < \epsilon$, **break**
  4) $\alpha_i = \alpha_i + d_i$
  5)   **for** $k = 1, \dots, N$ **do**
    a) $G_k = G_k + Q_{ik}d_i$
    b) $PG_k = \begin{cases} \min(G_k, 0), & \text{if } \alpha_k = 0 \\ \max(G_k, 0), & \text{if } \alpha_k = C_k \\ G_k, & \text{if } 0 < \alpha_k < C_k \end{cases}$
    end for
  6) $t = t + 1$
end while
Output: $\boldsymbol{\alpha}$ and $\boldsymbol{G}$.

---

3) *Stop Strategy*. Instead of stopping the method using the difference between the maximum and minimum value of $\boldsymbol{PG}$, the method is stopped when the decrease of the objective value reaches a tolerance ϵ.

The objective value of the modified DCD method decreases with the number of iterations. Meanwhile, the decrease of objective value is maximized at each iteration. The implementation details of the modified DCD method is shown in Algorithm 2. Usually, the optimization process starts from an initial point $\boldsymbol{\alpha} = \boldsymbol{0}$, so $\boldsymbol{G} = \boldsymbol{p}$.

For the global optimal solution $\boldsymbol{\alpha}^*$, an ϵ-accurate solution $\boldsymbol{\alpha}^t$ is defined as $f(\boldsymbol{\alpha}^t) \le f(\boldsymbol{\alpha}^*) + \epsilon$. The modified DCD method reaches an ϵ-accurate solution in $O(\log(1/\epsilon))$ iterations. The proof can be found in [44, Th. 1 and Appendix 7.1]. Clearly, our stop strategy $|Df_i| < \epsilon$ satisfies

$$f(\boldsymbol{\alpha}^t) \le f(\boldsymbol{\alpha}^*) + \epsilon \le f(\boldsymbol{\alpha}^{t+1}) + \epsilon. \qquad (30)$$

Therefore, (30) means the modified DCD method can be terminated within $O(\log(1/\epsilon))$ iterations. Algorithm 2 spends most time in steps 1), 2), and 5), and their computational cost is $O(N)$. Therefore, the computational cost of the modified DCD method is about $O(\log(1/\epsilon)N)$.
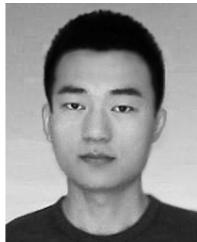
REFERENCES

[1] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing—Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2008.
[2] B. Datt, T. R. McVicar, T. G. V. Niel, D. L. B. Jupp, and J. S. Pearlman, "Preprocessing EO-1 hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1246–1259, Jun. 2003.
[3] B. Hörig, F. Kühn, F. Oschütz, and F. Lehmann, "HyMap hyperspectral remote sensing to detect hydrocarbons," *Int. J. Remote Sens.*, vol. 22, no. 8, pp. 1413–1422, 2001.

[4] P. M. Mehl, Y.-R. Chen, M. S. Kim, and D. E. Chan, "Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations," *J. Food Eng.*, vol. 61, no. 1, pp. 67–81, 2004.

[5] M. T. Eismann, A. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, Jun. 2009.

[6] A. Zare, J. Bolton, P. Gader, and M. Schatten, "Vegetation mapping for landmine detection using long-wave hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 172–178, Jan. 2008.

[7] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[8] J. Theiler and B. R. Foy, "EC-GLRT: Detecting weak plumes in non-Gaussian hyperspectral clutter using an elliptically-contoured generalized likelihood ratio test," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, pp. I–221–I–224.

[9] 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.

[10] S. Kraut, L. L. Scharf, and R. W. Butler, "The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 427–438, Feb. 2005.

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

[12] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Jun. 2011.

[13] Y. Zhang, B. Du, and L. Zhang, "A sparse representation-based binary hypothesis model for target detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1346–1354, Mar. 2015.

[14] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[15] H. Ren and C.-I. Chang, "Target-constrained interference-minimized approach to subpixel target detection for hyperspectral images," *Opt. Eng.*, vol. 39, pp. 3138–3145, 2000.

[16] T. Wang, B. Du, and L. Zhang, "A kernel-based target-constrained interference-minimized filter for hyperspectral sub-pixel target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 626–637, Apr. 2013.

[17] H. Kwon and N. M. Nasrabadi, "A comparative analysis of kernel subspace target detectors for hyperspectral imagery," *Eurasip J. Adv. Signal Process.*, vol. 2007, no. 1, pp. 1–13, Jan. 2006.

[18] L. Zang, L. Zang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.

[19] A. K. Ziemann, J. Theiler, and D. W. Messinger, "Hyperspectral target detection using manifold learning and multiple target spectra," in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop*, 2015, pp. 1–7.

[20] J. Tan, J. Zhang, and Y. Zhang, "Target detection for polarized hyperspectral images based on tensor decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 674–678, May 2017.

[21] Y. Liu, G. Gao, and Y. Gu, "Tensor matched subspace detector for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 1967–1974, Apr. 2017.

[22] Y. Zhang, B. Du, L. Zhang, and T. Liu, "Joint sparse representation and multitask learning for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 894–906, Feb. 2017.

[23] Y. Zhang, K. Wu, B. Du, and X. Hu, "Multitask learning-based reliability analysis for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, to be published, doi: 10.1109/JSTARS.2019.2894802.

[24] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Simultaneous joint sparsity model for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 676–680, Jul. 2011.

[25] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, Nov. 2016.

[26] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, Jun. 2014.

[27] J. Theiler and B. R. Foy, "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 98–102, Jan. 2006.

[28] C. E. Caefer, J. Silvermana, O. Orthalb, D. Antonellib, Y. Sharonib, and S. R. Rotman, "Improved covariance matrices for point target detection in hyperspectral data," *Opt. Eng.*, vol. 47, no. 7, pp. 076 402-1–076 402-13, Jul. 2008.

[29] S. Matteoli, M. Diani, and G. Corsini, "Hyperspectral anomaly detection with kurtosis-driven local covariance matrix corruption mitigation," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 532–536, May 2011.

[30] E. J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal components?" *J. ACM*, vol. 58, no. 3, pp. 11-1–11-37, May 2011.

[31] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Neural Inf. Process. Syst.*, 2009, vol. 22, pp. 2080–2088.

[32] B. Demir and S. Ertürk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4071–4084, Nov. 2010.

[33] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[34] J. Liu and J. Zhang, "Spectral unmixing via compressive sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7099–7110, Nov. 2014.

[35] M. Cui and S. Prasad, "Spectral-angle-based discriminant analysis of hyperspectral data for robustness to varying illumination," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4203–4214, Sep. 2017.

[36] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[37] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[38] 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. 2538–2371, Apr. 2019, doi: 10.1109/TGRS.2018.2872900.

[39] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Dec. 2013.

[40] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, 2008.

[41] T. Joachims, "Training linear SVMs in linear time," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2006, pp. 217–226.

[42] A. J. Smola, S. V. N. Vishwanathan, and Q. V. Le, "Bundle methods for machine learning," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 1377–1384.

[43] M. Collins, A. Globerson, T. Koo, X. Carreras, and P. Bartlett, "Exponentiated gradient algorithms for conditional random fields and maxmargin Markov networks," *J. Mach. Learn. Res.*, vol. 9, pp. 1775–1822, 2008.

[44] C.-J. Hsieh, K.-W. Chang, C.-J. Lin, S. S. Keerthi, and S. Sundararajan, "A dual coordinate descent method for large-scale linear SVM," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 408–415.

[45] O. A. Carvalho, Junior, and P. R. Meneses, "Spectral correlation mapper (SCM): An improvement on the spectral angle mapper (SAM)," presented at the 9th Airborne Earth Science Workshop, Pasadena, CA, USA, 2000, Paper JPL Pub. 00-18.

[46] B. Tu, X. Zhang, X. Kang, G. Zhang, J. Wang, and J. Wu, "Hyperspectral image classification via fusing correlation coefficient and joint sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 340–344, Mar. 2018.

[47] A. Giannandrea *et al.*, "The SHARE 2012 data campaign," *Proc. SPIE*, vol. 8743, pp. 87 430F–1–87 430F–15, 2013.

[48] N. M. Nasrabadi, "Regularized spectral matched filter for target recognition in hyperspectral imagery," *IEEE Signal Process. Lett.*, vol. 15, no. 3, pp. 317–320, Mar. 2008.

[49] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[50] 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.
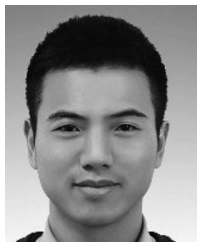
[51] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, Apr. 2008.

[52] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[53] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, Pittsburgh, PA, USA, 2006, pp. 233–240.

Zaiping Lin received the B.Eng. and Ph.D. degrees from the National University of Defense Technology (NUDT), Changsha, China, in 2007 and 2012, respectively.

He is currently a Lecturer with the College of Electronic Science and Engineering, NUDT. His research interests include infrared image processing and signal processing.

Qiang Ling received the B.Eng. degree in measurement engineering and the M.Eng. degree in control science and engineering from Air Force Engineering University, Shaanxi, China, in 2013 and 2016, respectively. He is currently working toward the Ph.D. degree with the College of Electronic Science, National University of Defense Technology, Changsha, China.

His research interests include pattern recognition and hyperspectral image processing.

Li Liu received the B.Sc. degree in communication engineering, the M.Sc. degree in photogrammetry and remote sensing, and the Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2003, 2005, and 2012, respectively.

In 2012, she joined the faculty at NUDT, where she is currently an Associate Professor with the College of System Engineering. During her Ph.D., she spent more than two years as a Visiting Student with the University of Waterloo, Waterloo, ON, Canada, from 2008 to 2010. From 2015 to 2016, she spent ten months visiting the Multimedia Laboratory, Chinese University of Hong Kong, Hong Kong. From 2016 to 2018, she was a Senior Researcher with the Machine Vision Group, University of Oulu, Finland. Her current research interests include facial behavior analysis, texture analysis, image classification, object detection, and recognition.

Dr. Liu was a Guest Editor of Special Issues for IEEE Transactions on Pattern Analysis and Machine Intelligence and International Journal of Computer Vision. Her papers have currently more than 1800 citations in Google Scholar. She is currently an Associate Editor for the *Visual Computer Journal*. She was a Co-Chair of seven international Workshops at Conference on Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision, and European Conference on Computer Vision. She is going to lecture a tutorial at CVPR'19.

Yulan Guo received the B.Eng. and Ph.D. degrees from the National University of Defense Technology (NUDT), Changsha, China, in 2008 and 2015, respectively.

He was a Visiting Ph.D. Student with the University of Western Australia, Perth, WA, Australia, from 2011 to 2014. He is currently an Assistant Professor with the College of Electronic Science, NUDT. He has authored more than 70 articles in journals and conferences, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and *International Journal of Computer Vision*. His research interests include three-dimensional (3-D) vision, particularly on 3-D feature learning, 3-D modeling, 3-D object recognition, and 3-D biometrics.

Dr. Guo was a Program Committee Member for several international conferences (e.g., ACM International Conference on Multimedia, International Joint Conferences on Artificial Intelligence, and Association for the Advancement of Artificial Intelligence conference). He was the recipient of the NUDT Outstanding Doctoral Dissertation Award in 2015 and the Chinese Association for Artificial Intelligence (CAAI) Outstanding Doctoral Dissertation Award in 2016. He served as an Associate Editor for the *IET Computer Vision*, a Guest Editor for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and a Reviewer for more than 30 international journals and conferences.

Wei An received the Ph.D. degree from the National University of Defense Technology (NUDT), Changsha, China, in 1999.

In 2016, she visited the University of Southampton, Southampton, U.K., as a Senior Visiting Scholar. She is currently a Professor with the College of Electronic Science, NUDT. She has authored or co-authored more than 100 journal and conference publications. Her research interests include signal processing and image processing.