

Joint Sparse Representation and Multitask Learning for Hyperspectral Target Detection

Yuxiang Zhang, *Member, IEEE*, Bo Du, *Senior Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, and Tongliang Liu, *Senior Member, IEEE*

Abstract—With the high spectral resolution, hyperspectral images (HSIs) provide great potential for target detection, which is playing an increasingly important role in HSI processing. Many target detection methods uniformly utilize all the spectral information or employ reduced spectral information to distinguish the targets and background. Simultaneously reducing spectral redundancy and preserving the discriminative information is a challenging problem in hyperspectral target detection. The multitask learning (MTL) technique may have the potential to solve the above problem, since it can explore the redundancy knowledge to construct multiple sub-HSIs and integrate them without any information loss. This paper proposes the joint sparse representation and MTL (JSR-MTL) method for hyperspectral target detection. This approach: 1) explores the HSIs similarity by a band cross-grouping strategy to construct multiple sub-HSIs; 2) takes full advantage of the MTL technique to integrate the sparse representation models for the multiple related sub-HSIs; and 3) applies the total reconstruction error difference accumulated over all the tasks to detect the targets. Extensive experiments were carried out on three HSIs, and it was founded that JSR-MTL generally shows a better detection performance than the other target detection methods.

Index Terms—Hyperspectral imagery, multitask learning (MTL), sparse representation, target detection.

I. INTRODUCTION

TARGET detection focuses on distinguishing specific targets from various backgrounds, with prior knowledge of the targets [1], [2], which is essentially a binary classification problem. With the characteristic of the high spectral resolution, hyperspectral images (HSIs) with hundreds or even thousands of spectral bands can distinguish subtle spectral differences,

even between very similar materials, providing a unique advantage for target detection [3]–[5]. Target detection has, therefore, attracted much attention in many HSI applications, and it has been successfully used in real-world applications, such as detecting rare minerals in geology, oil pollution in environmental research, landmines in the public safety and defense domain, and man-made objects in reconnaissance and surveillance applications [6]–[10].

The current target detection methods mainly utilize the detailed spectral information from HSI data and use different techniques to distinguish the targets and background. One approach typically derives from the statistical hypothesis testing theory [11]. This kind of method implicitly or explicitly assumes that the test pixels in the HSI obey a certain statistical distribution, such as a Gaussian distribution. The generalized likelihood ratio test and the maximum likelihood estimation [11] methods can be conveniently used to construct a detector. According to the different hypothesis models, a large number of target detection methods have been proposed in recent decades. Well-known algorithms include the adaptive coherence/cosine estimator (ACE) [12], the spectral matched filter (SMF) [13], and the matched subspace detector (MSD) [14]. Some target detection methods, such as the constrained energy minimization (CEM) [15], are conducted via filtering or projection techniques building a finite impulse response filter and imposing it on the targets and backgrounds in order to highlight the targets and suppress the background [16], [17]. Another kind of target detection method is based on machine learning techniques, which have recently been the subject of great attention. Typical examples of these methods are sparsity-based detectors [17]–[21]. For example, Chen *et al.* [20] introduced sparse representation into hyperspectral target detection and proposed the sparsity-based target detector (STD), where a test sample is approximately represented by very few training samples from both target and background dictionaries, and the recovered sparse representation is directly used for the detection decision. The sparse representation-based binary hypothesis detector (SRBBHD) was further proposed for mixed-target pixels [21] through building a sparse representation-based binary hypothesis model. The other common methods are the kernel-based target detectors [22]–[25]. Kwon and Nasrabadi [22] combined the kernel trick and kernel technique with the target detection problem, mapping the original feature space into a high-dimensional kernel space in order to deal with the linearly inseparable problem in the original space. Typical methods

Manuscript received June 16, 2016; revised September 2, 2016; accepted October 3, 2016. Date of publication December 5, 2016; date of current version December 29, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61471274 and Grant 41431175, in part by the Natural Science Foundation of Hubei Province under Grant 2014CFB193, and in part by the Fundamental Research Funds for the Central Universities under Grant 2042014kf0239 (*Corresponding author: Bo Du.*)

Y. Zhang is with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430079, China (e-mail: zhangyx@cug.edu.cn).

B. Du is with the School of Computer, Wuhan University, Wuhan 430079, China (e-mail: gunsapace@163.com).

L. Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn).

T. Liu is with the Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: tliang.liu@gmail.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2016.2616649

0196-2892 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

include the kernel SMF [23], the kernel orthogonal subspace projection detector [24], and kernel MSD [25].

The above existing target detection methods usually treat each pixel in the image as a vector format. The vector corresponding to each pixel is an input to a detector, resulting in a response value. For an ideal detector, the response values for the target pixels are large, while the response values for the nontarget pixels are small. Thus, the difference between the target and nontarget pixels can be represented by the output 2-D image with the response values. It can be observed that, most of the current target detection methods uniformly utilize the discriminative information within all single-band images, without considering the inherent similarity between the adjacent single-band images. However, the spectral resolution of HSIs is so high that even the adjacent single-band images present a great spectral similarity or redundancy. In this case, the adjacent single-band images inherently contain both redundancy and discriminative information.

This phenomenon deserves further consideration. On the one hand, it is proved that, spectral redundancy provides an obstacle for effective target detection [26]–[28]. Many methods on dimension reduction and metric learning for hyperspectral target detection have been proposed to relieve this problem [29]–[31]. However, none of them can guarantee that all the valuable discriminative spectral information underlying the HSI data is preserved, since the HSI data dimension is greatly reduced after the process. On the other hand, those methods using a uniform vector as an input can fully explore the discriminative information underlining the spectrum of each pixel. This is because all the original bands are directly used to both construct the model and perform the detection. Typical methods include filter-based target detectors [16] and sparsity representation-based detectors [20]. However, the spectral redundancy behind the HSI data is not taken into consideration at all, which may also hinder the accuracy and efficiency of target detection procedure [26]. In a word, there exists a dilemma to reduce spectral redundancy and preserve the discriminative information. Can we find an effective way to solve this problem?

The recent development of the multitask learning (MTL) technique has attracted much interests, and MTL has been successfully used in many real-world applications, such as handwritten digit recognition [32], face recognition [33], spam filtering [34], and Web search ranking [35]. In machine learning, single-task learning (STL) is to learn one task at a time. Large problems are divided into small and independent subproblems that are learned separately and then recombined [36]. This methodology is sometimes counterproductive, because it may ignore the potential information contained in the training signals of other tasks. Compared with STL, MTL is an approach that learns a problem together with other related problems at the same time [36]. This often leads to a better model for the main task, as it allows the learner to use the commonality among the tasks, and thus, MTL can work particularly well if the tasks have some commonality.

Based on the above analysis, MTL, which requires commonality among the tasks, is innovatively introduced to explore the inherent spectral similarity between the single-band images

for hyperspectral target detection. Based on the consistency of the MTL technique and the HSI characteristic, this paper proposes the joint sparse representation and MTL (JSR-MTL) method for hyperspectral target detection, which may provide a compromise between reducing the redundancy in HSI data and preserving the discriminative information in target detection procedure. The proposed method explores the spectral similarity between the adjacent single-band images to construct multiple sub-HSIs with the band cross-grouping strategy, which leads to multiple related detection tasks. In this way, the redundancy in each detection task can be effectively avoided. The proposed method further explores the similarity among the sub-HSIs to analyze the latent sparse representation of each task, and the multiple sparse representation model is then integrated via a unified MTL approach to complement each other for target detection. In this way, the elaborate and fine spectral information behind the high-dimension original HSI data set is fully used, so that the discriminative information is not lost. Finally, the proposed method applies the total reconstruction error difference between the target and the background subdictionaries accumulated over all the tasks to detect the probable targets. To the best of our knowledge, it is the first attempt to effectively compromise data redundancy reduction and discriminative information preservation with the MTL technique in hyperspectral target detection, which may provide a fresh eye idea in hyperspectral target detection field.

The rest of this paper is organized as follows. Section II briefly introduces the classical STD and the MTL theory. The proposed JSR-MTL method is then presented in detail in Section III. The experimental results of the proposed method with several HSIs are presented in Section IV. Finally, the conclusions are drawn in Section V.

II. BRIEF INTRODUCTION TO PREVIOUS METHODS

A. Classical Sparsity-Based Target Detector

The STD is conducted based on the assumption that, the spectrum of a background pixel approximately lies in a low-dimensional subspace spanned by the background training samples, and the spectrum of a target pixel approximately lies in a low-dimensional subspace spanned by the target training samples. Thus, the unknown test pixel approximately lies in the union of the background and target subspace. The recovery process implicitly leads to a competition between the background subspace and the target subspace [20].

Considering an HSI with N pixels and B bands, then a test pixel \mathbf{x} can be represented by a sparse and linear combination of the background training samples $\{\mathbf{a}_i^b\}_{i=1,2,\dots,N_b}$ and the target training samples $\{\mathbf{a}_i^t\}_{i=1,2,\dots,N_t}$

$$\begin{aligned}\mathbf{x} &\approx (\alpha_1^b \mathbf{a}_1^b + \alpha_2^b \mathbf{a}_2^b + \dots + \alpha_{N_b}^b \mathbf{a}_{N_b}^b) \\ &\quad + (\alpha_1^t \mathbf{a}_1^t + \alpha_2^t \mathbf{a}_2^t + \dots + \alpha_{N_t}^t \mathbf{a}_{N_t}^t) \\ &= \mathbf{A}^b \boldsymbol{\alpha}^b + \mathbf{A}^t \boldsymbol{\alpha}^t \\ &= \mathbf{A} \boldsymbol{\alpha}\end{aligned}\tag{1}$$

where N_b and N_t are the number of background and target training samples, respectively. $\mathbf{A}^b \in \mathbb{R}^{B \times N_b}$ and $\mathbf{A}^t \in \mathbb{R}^{B \times N_t}$ are the background dictionary and target dictionary, whose

columns are the background and target training samples, respectively. \mathbf{A} is the union dictionary consisting of both the background and target training samples. $\boldsymbol{\alpha}$ is a concatenation of $\boldsymbol{\alpha}^b$ and $\boldsymbol{\alpha}^t$, which are sparse vectors with only a few nonzero entries. The sparse vector $\boldsymbol{\alpha}$ can be recovered by solving

$$\hat{\boldsymbol{\alpha}} = \arg \min \|\mathbf{A}\boldsymbol{\alpha} - \mathbf{x}\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}\|_0 \leq K_0 \quad (2)$$

where $\|\cdot\|_0$ denotes the ℓ_0 -norm, which is defined as the number of nonzero entries in the vector. K_0 is a given upper bound on the sparsity level [37]. The above problem can be solved by the orthogonal matching pursuit [38] algorithm.

Partially reconstructed pixels, using only the background or target dictionary, can be obtained by decomposing the sparse vector $\hat{\boldsymbol{\alpha}}$ into $\hat{\boldsymbol{\alpha}}^b$ and $\hat{\boldsymbol{\alpha}}^t$, which are the vectors, respectively, corresponding to the background and target training samples. The residuals of recovery by the background and target subspaces are

$$\begin{aligned} r_b(\mathbf{x}) &= \|\mathbf{x} - \mathbf{A}^b \hat{\boldsymbol{\alpha}}^b\|_2 \\ r_t(\mathbf{x}) &= \|\mathbf{x} - \mathbf{A}^t \hat{\boldsymbol{\alpha}}^t\|_2. \end{aligned} \quad (3)$$

The class label of the test pixel can be determined by these residuals, and the output of the detector is calculated by

$$D_{\text{STD}}(\mathbf{x}) = r_b(\mathbf{x}) - r_t(\mathbf{x}). \quad (4)$$

B. Multitask Learning Theory

The MTL methodology was proposed by Caruana [36]. It is an inductive transfer method that uses the domain-specific information contained in the training signals of related tasks [36], which can guarantee that the related tasks can learn from each other and make the inductive transfer method work [39], [40].

There are two key techniques of MTL. One is the construction of multiple tasks with commonality. The multiple tasks can be constructed in various ways, which may depend on the specific application. The above-mentioned handwritten digit recognition problem involves constructing ten related digit recognition tasks, according to the number of digits [32]. Some of the MTL-based classification methods construct related classification tasks according to the different features captured from the training samples of the same class [33], [41].

The other key technique is the relevance analysis of multiple tasks. Tasks can be related in various ways. There are two commonly used approaches: 1) tasks may be related through assuming that all the learned functions are close to each other in some norm and 2) tasks may also be related in that they all share a common underlying representation [42], [43]. For the first type of task relatedness, the typical approach is the linear regression function. For example, the handwritten digit recognition problem assumes that the ten digit recognition tasks share a common linear regression model [32]. For the second type of task relatedness, there are some widely used representations, such as sparsity, a manifold constraint, or a graphical model structure [32]. For example, the classification method proposed in [34] assumes that different features for different tasks share a common sparse representation with the $\ell_{1,2}$ mixed-norm. These two types of task relatedness can

also be combined, for example, we study the problem of estimation multiple regression equations, where multiple tasks have the same functions and simultaneously share the sparsity assumptions on the regression coefficients [44]. The model can be specified as

$$\begin{aligned} \mathbf{y}^1 &= \mathbf{X}^1 \boldsymbol{\beta}^1 + \boldsymbol{\eta}^1 \\ \mathbf{y}^2 &= \mathbf{X}^2 \boldsymbol{\beta}^2 + \boldsymbol{\eta}^2 \\ &\vdots \\ \mathbf{y}^T &= \mathbf{X}^T \boldsymbol{\beta}^T + \boldsymbol{\eta}^T \end{aligned} \quad (5)$$

where for each $t = 1, 2, \dots, T$, \mathbf{y}^t is an n -dimensional vector of the observations, \mathbf{X}^t is a prescribed $n \times M$ design matrix, and $\boldsymbol{\beta}^t$ is the unknown vector of the regression coefficients. The equations can be incorporated into the following model:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}^t} \sum_{t=1}^T \|\mathbf{y}^t - \mathbf{X}^t \boldsymbol{\beta}^t\|_2^2 + \Omega(\boldsymbol{\beta}) \quad (6)$$

where $\Omega(\boldsymbol{\beta})$ further encodes the task relatedness, and different assumptions on the task relatedness lead to different regularization terms. It can be assumed that the regression vectors share the same ℓ_0 -norm sparsity pattern, which means that the set of relevant predictor variables is the same across the different equations.

III. JOINT SPARSE REPRESENTATION WITH MULTITASK LEARNING FOR HYPERSPECTRAL TARGET DETECTION

In this section, we further analyze the spectral similarity behind the HSI. Based on the analysis, multiple related detection tasks are constructed through constructing multiple sub-HSIs for each task.

A. Construction of Multiple Detection Tasks With the Band Cross-Grouping Strategy

To better illustrate the spectral similarity behind the HSI, we took the Hyperspectral Digital Image Collection Experiment (HYDICE) data set as an example, and the similarity between each band image was calculated via correlation coefficients. The HYDICE data set is introduced in detail in the following experimental part. The result is provided through a correlation coefficient matrix, as shown in Fig. 1, where the axes represent the band order of the HSI. It can be seen that the correlation coefficient values near to the diagonal (corresponding to the adjacent bands) are much higher.

To utilize the spectral similarity, MTL is introduced for the hyperspectral target detection. One key problem is how to construct the multiple related detection tasks. Considering that each data set naturally corresponds to a detection task, multiple related detection tasks can be constructed through constructing multiple related sub-HSIs from the original HSI. The multiple related sub-HSIs can be obtained with the band cross-grouping strategy. That is to say, in accordance with the band order of the original HSI, we cross-group the multiple adjacent single-band images into different groups. Finally, each group forms

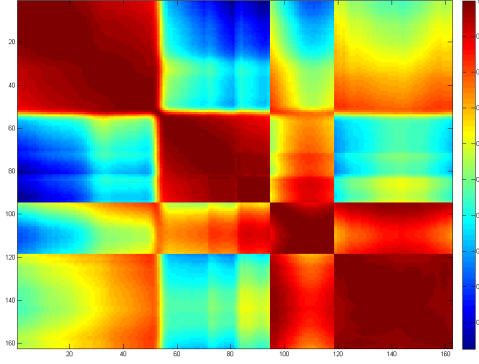


Fig. 1. Spectral similarity between different bands of an HSI.

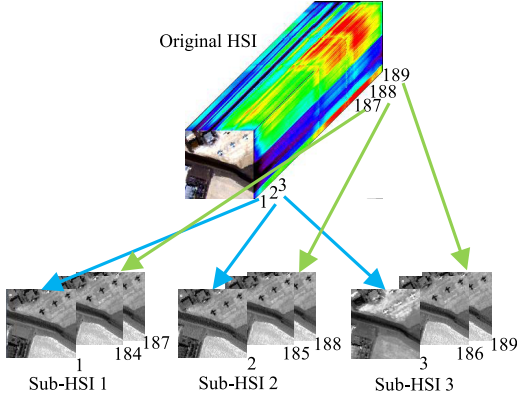


Fig. 2. Band cross-grouping strategy for the multiple detection tasks.

a sub-HSI, as shown in Fig. 2. In this way, the corresponding single-band images in the same position in each sub-HSI are usually highly related to each other. For example, the first single-band images (labeled as 1, 2, and 3) in three sub-HSIs are similar to each other, which guarantee the relatedness among the multiple sub-HSIs. Therefore, these multiple related sub-HSIs correspond to multiple related detection tasks.

For the original HSI with the size of $h \times w \times B$, which can be labeled as $\mathbf{X} \in \mathbb{R}^{h \times w \times B}$, the number of detection tasks is set as K , and the size of each sub-HSI is $\{h \times w \times B^k\}_{k=1}^K$. The band number of the sub-HSI $\{B^k\}_{k=1}^K$ can be calculated as follows:

$$B^k = \begin{cases} B/K, & a = 0 \\ B/K + 1, & 0 < k \leq a \\ B/K, & a < k \leq K \end{cases} \quad (7)$$

where B divided by K is equal to B/K , and a is the remainder, labeled as $\text{mod}(B, K) = a$. It can be seen that the band numbers of the sub-HSIs may not be the same, which is permitted in this paper, and is illustrated in the following part.

B. Joint Sparse Representation and Multitask Learning Model

Another key problem of MTL is the relevance analysis of the multiple tasks. Fortunately, the spectral similarity of the

multiple sub-HSIs naturally guarantees the relevance of the multiple detection tasks. Furthermore, since the multiple sub-HSIs are highly related to each other, the multiple detection tasks are likely to share a common underlying representation. In recent years, sparse representation presented improved detection performance at some cases [18]–[21]. Therefore, it is reasonable to assume that the multiple detection tasks share the sparse representation.

Considering the k th sub-HSI, a test sample $\{\mathbf{x}^k \in \mathbb{R}^{B^k}\}_{k=1}^K$ can be modeled to lie in the union of the background and target subspaces, respectively, spanned by the background training samples $\{\mathbf{d}_i^{kb} \in \mathbb{R}^{B^k \times N_b}\}_{i=1}^{N_b}$ and the target training samples $\{\mathbf{d}_i^{kt} \in \mathbb{R}^{B^k \times N_t}\}_{i=1}^{N_t}$. Therefore, \mathbf{x}^k can be represented by a sparse linear combination of all the training samples

$$\begin{aligned} \mathbf{x}^k &= (w_1^{kb} \mathbf{d}_1^{kb} + w_2^{kb} \mathbf{d}_2^{kb} + \dots + w_{N_b}^{kb} \mathbf{d}_{N_b}^{kb}) \\ &\quad + (w_1^{kt} \mathbf{d}_1^{kt} + w_2^{kt} \mathbf{d}_2^{kt} + \dots + w_{N_t}^{kt} \mathbf{d}_{N_t}^{kt}) + \boldsymbol{\varsigma}^k \\ &= \mathbf{D}^{kb} \mathbf{w}^{kb} + \mathbf{D}^{kt} \mathbf{w}^{kt} + \boldsymbol{\varsigma}^k = \mathbf{D}^k \mathbf{w}^k + \boldsymbol{\varsigma}^k \end{aligned} \quad (8)$$

where $\boldsymbol{\varsigma}^k$ is the random noise. N_b and N_t are the number of background and target training samples, respectively. \mathbf{D}^{kb} and \mathbf{D}^{kt} are the $B^k \times N_b$ background dictionary and $B^k \times N_t$ target dictionary, respectively. $\mathbf{w}^k \in \mathbb{R}^{N_b+N_t}$ is a concatenation of \mathbf{w}^{kb} and \mathbf{w}^{kt} , which are the coefficient subvectors over the k th subdictionary \mathbf{D}^{kb} and \mathbf{D}^{kt} . It can be seen that the size of these coefficient vectors is not related to the band number of the sub-HSI. Thus, the fact that the band numbers of the sub-HSIs may not be the same does not affect the representation model.

Specifically, the background dictionary is generated locally for each test pixel through a dual concentric window, which separates the local area around each pixel into two regions, a small inner window region (IWR) centered within a larger outer window region (OWR) [20]. $\mathbf{D}^k \in \mathbb{R}^{B^k \times (N_b+N_t)}$ is the union dictionary consisting of both the background and target training samples for the k th sub-HSI.

Since the K groups of subimages are highly related to each other, they are likely to share the same sparse representation. The sparse representation for a single-task case can be generalized to a multiple-task case. Thus, for the multiple detection tasks, the original test pixel $\mathbf{x} \in \mathbb{R}^B$ decomposed into K subvectors can be represented as

$$\begin{aligned} \mathbf{x}^1 &= \mathbf{D}^{1b} \mathbf{w}^{1b} + \mathbf{D}^{1t} \mathbf{w}^{1t} + \boldsymbol{\varsigma}^1 = \mathbf{D}^1 \mathbf{w}^1 + \boldsymbol{\varsigma}^1 \\ &\vdots \\ \mathbf{x}^K &= \mathbf{D}^{Kb} \mathbf{w}^{Kb} + \mathbf{D}^{Kt} \mathbf{w}^{Kt} + \boldsymbol{\varsigma}^K = \mathbf{D}^K \mathbf{w}^K + \boldsymbol{\varsigma}^K. \end{aligned} \quad (9)$$

These models can be incorporated into the multitask sparse representation model

$$\widehat{\mathbf{W}} = \arg \min_{\mathbf{W}} \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^k \mathbf{w}^k\|_2^2 + \rho \|\mathbf{W}\|_{2,1} \quad (10)$$

where $\mathbf{W} \in \mathbb{R}^{(N_b+N_t) \times K}$ is the matrix formed by stacking the vectors $\mathbf{w}^k \in \mathbb{R}^{N_b+N_t}$. ρ , which penalizes the $\ell_{2,1}$ -norm of the matrix \mathbf{W} , is the regularization parameter to trade off the data fidelity term and the regularization term. It is obtained by first computing the ℓ_2 -norm of the rows $\{\mathbf{w}_i\}_{i=1}^{N_b+N_t}$ (across

Algorithm 1 Coefficients Optimization Algorithm**Input:**

- 1) Data $\{\mathbf{D}^k, \mathbf{x}^k\}_{k=1}^K$, regularization parameter ρ ,
- 2) initial variables and parameters $\hat{\mathbf{w}}^{k,0} = (\mathbf{D}^k)^T \mathbf{x}^k$, $\hat{\mathbf{v}}^{k,0} = \hat{\mathbf{w}}^{k,0}$, $\alpha^0 = -1$, $t := 0$.

repeat

- 1) $\hat{\mathbf{w}}^{k,t+1} = \hat{\mathbf{v}}^{k,t} - \frac{1}{2^t} \left[-(\mathbf{D}^k)^T \mathbf{x}^k + (\mathbf{D}^k)^T \mathbf{D}^k \hat{\mathbf{v}}^{k,t} \right]$,
- 2) $\hat{\mathbf{w}}_i^{t+1} = \left[1 - \rho \eta^t / \|\hat{\mathbf{w}}_i^{t+1}\|_2 \right]_+ \hat{\mathbf{w}}_i^{t+1}$,
- 3) $\alpha^t = 2(t-1) / (1 + \sqrt{1 + 4t^2})$,
- 4) $\hat{\mathbf{v}}^{k,t+1} = (1 + \alpha^t) \hat{\mathbf{w}}^{k,t+1} - \alpha^t \hat{\mathbf{w}}^{k,t}$,

- 5) $t := t + 1$

until convergence is attained

Output: Coefficient vectors $\{\hat{\mathbf{w}}^k\}_{k=1}^K$.

the tasks) of the matrix \mathbf{W} , and then computing the ℓ_1 -norm of the vector $b(\mathbf{W}) = (\|\mathbf{w}_1\|_2, \dots, \|\mathbf{w}_{N_b+N_t}\|_2)^T$. This norm combines the tasks and encourages shared sparsity among the columns of the matrix \mathbf{W} . Moreover, this norm simultaneously encourages the sparsity of each column of the matrix \mathbf{W} .

For the model optimization, we use the popular accelerated proximal gradient (APG) algorithm [45]–[47] to efficiently solve the problem in (10), which has the convergence speed of $O(1/k^2)$. The APG algorithm alternately updates a matrix sequence $\hat{\mathbf{W}}^t = [\mathbf{w}_i^{k,t}]$ and an aggregation matrix sequence $\hat{\mathbf{V}}^t = [\mathbf{v}_i^{k,t}]$.

Given the current matrix aggregation matrix $\hat{\mathbf{V}}^t$, a generalized gradient mapping step is employed to update matrix $\hat{\mathbf{W}}^{t+1}$ as follows:

$$\begin{aligned} \hat{\mathbf{w}}^{k,t+1} &= \hat{\mathbf{v}}^{k,t} - \eta^t \nabla^{k,t} \\ \hat{\mathbf{w}}_i^{t+1} &= \left[1 - \rho \eta^t / \|\hat{\mathbf{w}}_i^{t+1}\|_2 \right]_+ \hat{\mathbf{w}}_i^{t+1} \end{aligned} \quad (11)$$

where $t \geq 1$, $k = 1, 2, \dots, K$, and $i = 1, 2, \dots, N_b + N_t$. $\nabla^{k,t} = -(\mathbf{D}^k)^T \mathbf{x}^k + (\mathbf{D}^k)^T \mathbf{D}^k \hat{\mathbf{v}}^{k,t}$, $\eta^t = 1/2^t$ is the step size, and $[\cdot]_+ = \max(\cdot, 0)$.

An aggregation forward step is then employed to update $\hat{\mathbf{V}}^{t+1}$ by linearly combining $\hat{\mathbf{W}}^{t+1}$ and $\hat{\mathbf{W}}^t$ as follows:

$$\hat{\mathbf{V}}^{t+1} = (1 + \alpha^t) \hat{\mathbf{W}}^{t+1} - \alpha^t \hat{\mathbf{W}}^t \quad (12)$$

where the sequence α^t is conventionally set to $\alpha^t = 2(t-1)/(1 + (1 + 4t^2)^{1/2})$, as applied in our implementation. The optimization method is summarized as Algorithm 1.

C. Residuals-Based Target Detection

After the recovery of the coefficient vectors $\hat{\mathbf{w}}^k$ corresponding to the union dictionary for each task, the subvectors according to the background subdictionary and the target subdictionary captures the discriminative information, which

can be used to obtain the detection decision via the recovery residuals with the two subdictionaries \mathbf{D}^{kb} and \mathbf{D}^{kt} .

We can then calculate the residual errors for the background and target between the multiple signals in the sub-HSIs $\{\mathbf{x}^k\}_{k=1}^K$ and the approximations recovered via their corresponding subdictionaries $\{\mathbf{D}^{kb}\}_{k=1}^K$ and $\{\mathbf{D}^{kt}\}_{k=1}^K$

$$\begin{aligned} r^b &= \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^{kb} \hat{\mathbf{w}}^{kb}\|_2 \\ r^t &= \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^{kt} \hat{\mathbf{w}}^{kt}\|_2 \end{aligned} \quad (13)$$

where $\hat{\mathbf{w}}^{kb}$ and $\hat{\mathbf{w}}^{kt}$ are the subsets of the coefficient vector $\hat{\mathbf{w}}^k$ associated with the background and target. The output of the test pixel \mathbf{x} is then calculated by

$$D(\mathbf{x}) = r^b - r^t. \quad (14)$$

Finally, a visual illustration of the proposed JSR-MTL algorithm for HSIs is shown in Fig. 3. Given an HSI, multiple sub-HSIs are extracted via the band cross-grouping strategy. We construct the multiple signals for each pixel $\{\mathbf{x}^k\}_{k=1}^K$, and multiple union dictionaries $\{\mathbf{D}^k\}_{k=1}^K$ with the local dual window. Each pixel is jointly and sparsely represented as a linear combination of the union dictionary corresponding to each task. Finally, the detection decision rules in favor of the target class or the background class with the lowest total reconstruction error accumulated over all the tasks.

IV. HYPERSPECTRAL DATA EXPERIMENTS**A. Data Set Description**

Three hyperspectral data sets were used in this paper to evaluate the effectiveness of the proposed detector introduced in Section III. The first data set consists of synthetic data, as used in [17] and [48], and is referred as the Synthetic data set. The 15 end member signatures selected from the U.S. Geological Survey [49] digital spectral library were used to generate the synthetic data, including Labradorite HS17.3B, Rhodochrosite HS67, and so on. The spectra were collected in 224 bands uniformly spanning 0.4–2.5 μm . Labradorite HS17.3B was used as the target spectrum. The target implantation method introduced by Chang *et al.* [50] was employed to generate the synthetic data. First, we divided the synthetic map, whose size was $s^2 \times s^2$ ($s = 8$), into $s \times s$ regions. Each region was initialized with the same type of ground cover, which was randomly selected from the aforementioned 15 kinds of spectra. Second, we implanted clean targets into the backgrounds by replacing their corresponding pixels. Finally, we mixed the synthetic map through a $(s+1) \times (s+1)$ spatial low-pass filter to generate the mixed pixels, and we added a Gaussian white noise with 20-dB SNR into the synthetic map at the same time. Fig. 4(a) shows the first band of the Synthetic data set, and Fig. 4(b) shows the ground truth. We selected two pixels labeled in red as the target atoms, so $N_t = 2$.

The second data set was collected by the HYDICE sensor [20], [51]. This data set has 210 spectral

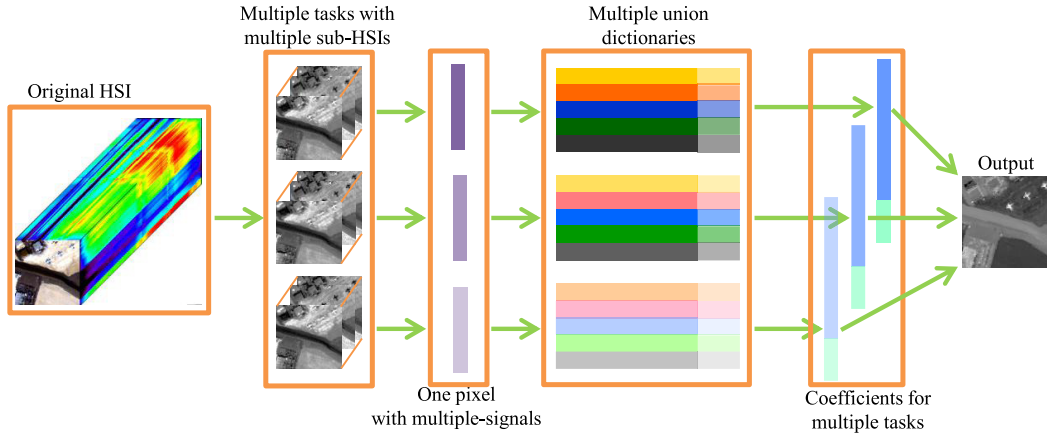


Fig. 3. Schematic of the JSR-MTL algorithm.

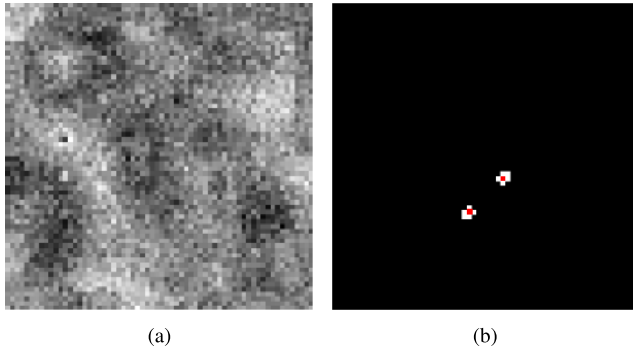


Fig. 4. Synthetic data set. (a) Image scene. (b) Ground truth.

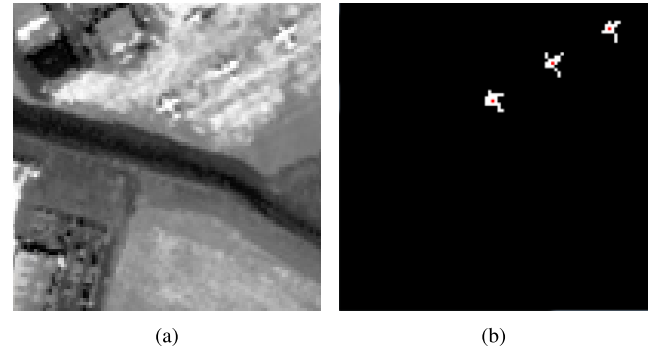


Fig. 6. AVIRIS data set. (a) Image scene. (b) Ground truth.

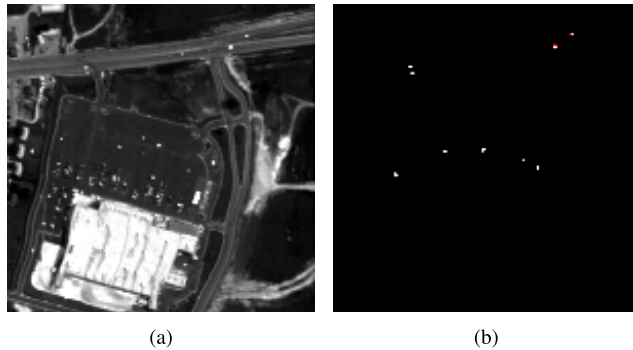


Fig. 5. HYDICE data set. (a) Image scene. (b) Ground truth.

bands in the visible CNIR range. The spatial resolution is 2 m and the spectral resolution is 10 nm. After removing the water absorption, low-SNR, and bad bands (1–4, 76, 87, 101–111, 136–153, and 198–210), 162 bands were retained. The image scene is 150×150 pixels, as shown in Fig. 5(a). This data set is an urban scene in which there are nine vehicles, including 21 target pixels as targets to be detected, as shown in Fig. 5(b). The spectral signatures of the target training samples were collected from the $N_t = 3$ pixels labeled in red.

The third data set was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor from San Diego, CA, USA. The spatial resolution of this image is

3.5 m per pixel [52], [53]. The image has 224 spectral channels in wavelengths ranging from 370 to 2510 nm. After removing the bands that correspond to the water absorption regions, low-SNR, and bad bands (1–6, 33–35, 97, 107–113, 153–166, and 221–224), 189 bands were retained in the experiments. An area of 100×100 pixels was used for the experiments. The image scene is shown in Fig. 6(a). There are three planes in the image, which consist of 58 pixels, as shown in Fig. 6(b). We selected one pixel labeled in red from each plane as the target atoms, $N_t = 3$.

B. Evaluation of the Multiple Detection Tasks Combination

First, the detection performance of the multiple detection tasks combination in the proposed JSR-MTL algorithm was investigated. Three cases needed to be compared: 1) the detection performance for the original HSI; 2) the detection performance for each sub-HSI; and 3) the detection performance for the multiple sub-HSI combination. We took three detection tasks via the band cross-grouping strategy as an example ($K = 3$), and the detection performance for the three cases is provided through the area under the curve (AUC) values, as shown in Table I. The best results for each data set are labeled in bold.

The sizes of the OWR for the three data sets were set as 17×17 . The sizes of the IWR are related to the size of the target, and were set as 7×7 , 5×5 , and 7×7 for the Synthetic, HYDICE, and AVIRIS data sets, respectively. The numbers of

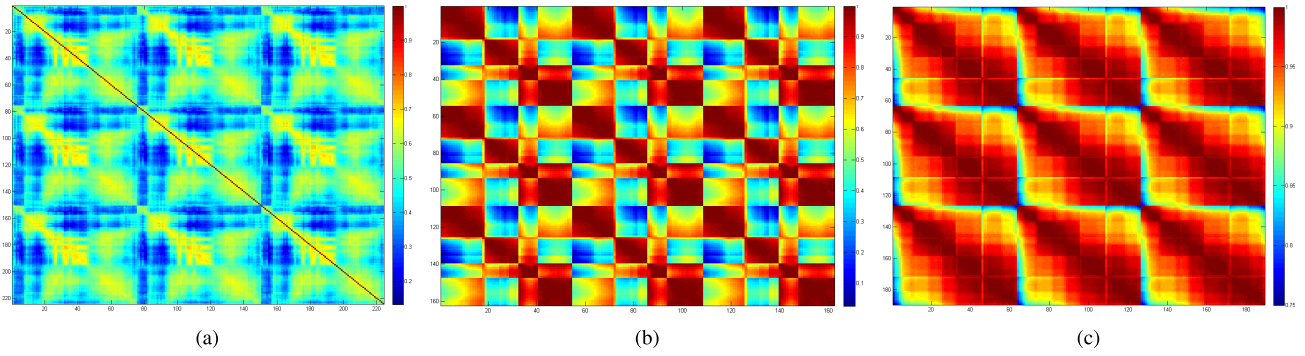


Fig. 7. Correlation coefficient matrix across different sub-HSIs for the three data sets. (a) Synthetic data set. (b) HYDICE data set. (c) AVIRIS data set.

TABLE I

AUC VALUES OBTAINED IN THE EVALUATION OF MULTIPLE DETECTION TASKS COMBINATION WITH TWO BAND GROUPING STRATEGIES

Strategy	Cases	Datasets		
		Synthetic	HYDICE	AVIRIS
Cross-grouping	Original HSI	0.88298	0.99366	0.9782
	Sub-HSI 1	0.86534	0.99309	0.98039
	Sub-HSI 2	0.87897	0.99564	0.98111
	Sub-HSI 3	0.87948	0.99229	0.98261
	Sub-HSIs combination	0.89081	0.99604	0.98059
Sequence-grouping	Sub-HSI 1	0.88808	0.90152	0.97404
	Sub-HSI 2	0.84739	0.83363	0.93349
	Sub-HSI 3	0.81244	0.66709	0.91058
	Sub-HSIs combination	0.8669	0.93991	0.96767

the background training samples for the three data sets were, therefore, $N_b = 240$, $N_b = 264$, and $N_b = 240$, respectively.

The results show that the performance of the multiple sub-HSI combination is better than that of the original HSI for all three data sets, and is generally better than that of the individual sub-HSIs, especially for the Synthetic and HYDICE data sets. For the AVIRIS data set, compared with the performance of the sub-HSIs, the detection performance of the sub-HSI combination is not always the best. We further analyze the reason for this in the following.

Second, the effectiveness of the band cross-grouping strategy in the JSR-MTL algorithm was evaluated and compared with the band sequence-grouping strategy in order to find out which way can better reveal the internal spectral similarity information across adjacent bands and better fit the MTL theory. The band sequence-grouping strategy means that we sequence-group the bands of the original HSI into multiple groups, and then each group forms a sub-HSI. For example, for the AVIRIS data set, when $K = 3$, we sequence-group the first 63 bands as the first sub-HSI, the second 63 bands as the second sub-HSI, and the last 63 bands as the third sub-HSI. For the two strategies, the detection performances of three cases are provided through the AUC values, as shown in Table I.

The results show that the performance of the multiple sub-HSIs combination with the band sequence-grouping strategy is worse than that of the original HSI for all three data sets. What is more, under each sub-HSI and multiple sub-HSIs combination cases, the performance of the band cross-grouping strategy is generally better than that of the band sequence-grouping strategy for all three data sets. The results

demonstrate that the spectral similarity between the multiple sub-HSIs with the band cross-grouping strategy leads to much obvious relatedness in the multiple detection tasks, which is more efficient for hyperspectral target detection.

Third, to further analyze the superiority of JSR-MTL with the band cross-grouping strategy, the commonality and complementary nature of the multiple sub-HSIs was investigated. Accordingly, the correlation coefficient matrices across three sub-HSIs were calculated for the three data sets, as shown in Fig. 7. A statistical analysis was further undertaken for the correlation coefficient matrices, and boxes were drawn to enclose the main parts of the values, excluding the biggest 10% and the smallest 10%, as shown in Fig. 8. It can be seen that the correlation coefficient matrix for the HYDICE data set shown in Fig. 7(b) is different from that shown in Fig. 1. The difference lies in the meaning of the band number. Fig. 1 shows the correlation coefficient matrix of the original HYDICE data set with band numbers $[1, 2, \dots, 161, 162]$. Fig. 7(b) shows the correlation coefficient matrix of the new data set, which was obtained by combining the three sub-HSIs together. The band numbers of the new combined data set $[1, 2, \dots, 161, 162]$ are actually $[1, 4, \dots, 160, 2, 5, \dots, 161, 3, 6, \dots, 162]$ of the original HSI. The correlation coefficient matrices of the combined sub-HSIs for the three data sets are, respectively, shown in Fig. 7.

For the Synthetic data set, as shown in Fig. 7(a), the first 75 dimensions correspond to the first sub-HSI, the next 75 dimensions correspond to the second sub-HSI, and the last 74 dimensions correspond to the third sub-HSI. For the HYDICE and AVIRIS data sets, as shown in Fig. 7(b) and (c), the dimensions, respectively, corresponding to the three sub-HSIs are 54 and 63. It can be seen that the correlation coefficient values between the sub-HSIs vary from 0.15 to 0.75, 0.05 to 1, and 0.75 to 1 for the Synthetic data set, the HYDICE data set, and the AVIRIS data set, respectively. In addition, as shown in Fig. 8, the middle 80% of the correlation coefficient values between sub-HSIs focus on the ranges of 0.32–0.62, 0.35–0.99, and 0.89–0.99 for the Synthetic data set, the HYDICE data set, and the AVIRIS data set, respectively. These results reflect two points. First, the correlation coefficient values between sub-HSIs are relatively large, especially for the Synthetic data set and the HYDICE data set. That is to say, on the one hand, different sub-HSIs are similar to each other. In addition, on the other hand, the different

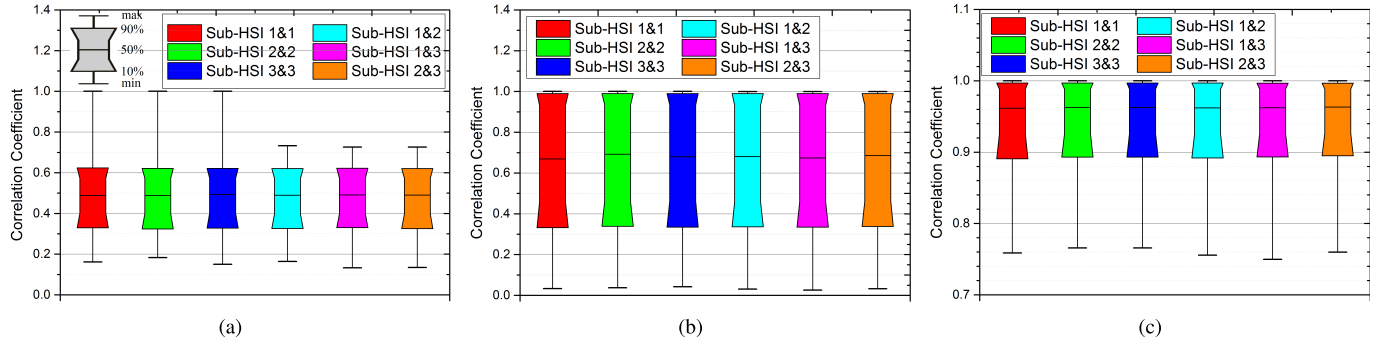


Fig. 8. Statistical analysis of the correlation coefficient matrix across different sub-HSIs for the three data sets. (a) Synthetic data set. (b) HYDICE data set. (c) AVIRIS data set.

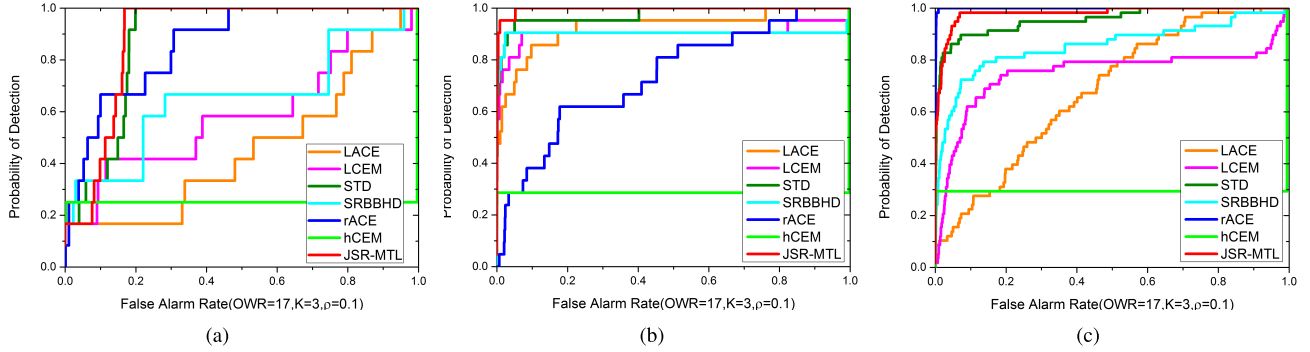


Fig. 9. ROC curves of the seven detectors for the three data sets (where $K = 3$, $\rho = 0.1$, and the size of the $OWR = 17$ for JSR-MTL). (a) Synthetic data set. (b) HYDICE data set. (c) AVIRIS data set.

sub-HSIs can reflect different aspects of the discriminative information of the HSI and are complementary to a certain degree. This result suggests that the multiple sub-HSIs have the commonality and complementary nature, it is logical to combine the sub-HSIs via MTL for target detection, and it has been verified that the multiple sub-HSI combination can obtain a better performance than the individual sub-HSIs for the Synthetic and HYDICE data sets. It can also be seen that the correlation coefficient values within sub-HSIs are similar to those between sub-HSIs. However, too large correlation coefficient values within sub-HSIs are not efficient for target detection, as a sub-HSI with very similar single-band images may contain less useful discriminative information. From the results of the AVIRIS data set, we can see that the correlation coefficient values both within sub-HSIs and between sub-HSIs are too large, especially the middle 80% of the correlation coefficient values, which can affect the performance of the multiple sub-HSI combination. In addition, it has been verified that the multiple sub-HSI combination cannot always obtain a best performance compared with all the individual sub-HSIs.

The detection performance of the proposed JSR-MTL algorithm was further analyzed and compared with the traditional detectors of local ACE (LACE), local CEM (LCEM), reweighted ACE (rACE) [48], hierarchical CEM (hCEM) [54], STD, and SRBBHD. We took $K = 3$ and the regularization parameter $\rho = 0.1$ for the JSR-MTL algorithm as an example. For all the detectors, we used the same given target spectra as *a priori* target spectra. For the local detectors of LACE, LCEM, STD, SRBBHD, and JSR-MTL, the sizes of the OWR for the three data sets were set as 17×17 . The sizes of the IWR were fixed as fixed as above-mentioned 7×7 ,

5×5 , and 7×7 for the Synthetic, HYDICE, and AVIRIS data sets, respectively. In the case of hCEM and LCEM, the mean of the target atoms was used as the target signature. We adopted the pixels falling in the OWR to estimate the background covariance matrix for LACE, to estimate the background correlation matrix for LCEM, and to construct the background dictionary for STD, SRBBHD, and JSR-MTL. The detection performance of the seven detectors is provided through the receiver operation characteristics (ROC) curves, as shown in Fig. 9.

For the Synthetic data set, as shown in Fig. 9(a), the ROC curves of JSR-MTL and the other detectors are not separated as much as for the other data sets. However, the ROC curve of JSR-MTL is above that of LACE, and it is above those of LCEM, hCEM, STD, and SRBBHD when the false alarm rate (FAR) is greater than 0.116, and it is above that of rACE when the FAR is greater than 0.161. For the HYDICE data set, as shown in Fig. 9(b), the ROC curve of JSR-MTL is always above those of the other detectors. For the AVIRIS data set, as shown in Fig. 9(c), rACE obtains the best result and the ROC curve of JSR-MTL is above those of the rest detectors.

Overall, compared with other detectors, hCEM obtains a low probability of detection at a large range of FAR. The detection performance of rACE is very different for the three data sets and it does not show a robust detection performance, especially for the HYDICE and AVIRIS data sets. JSR-MTL obtains a good performance compared with the other detectors for all three data sets. In addition, we can further adjust the number of detection tasks K , the regularization parameter ρ , and the window size to obtain an even better performance.

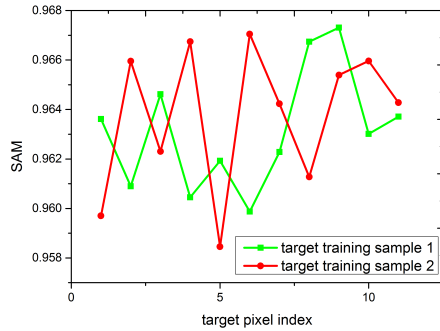


Fig. 10. Spectral similarity of the target spectra for the Synthetic data set.

In addition, the low detection probability of JSR-MTL for the Synthetic data set in a low FAR is analyzed as follows. With the spatial low-pass filter and Gaussian white noise, all the target pixels in the Synthetic data set are mixed pixels. The spectral similarity between the target spectra was further analyzed via the spectral angel mapper (SAM), and the SAM between the target training samples with the other target pixels is shown in Fig. 10. It can be seen that the target spectra are much different from each other and show a great spectral variability. Therefore, this data set is relatively hard to detect the targets when compared with the other two data sets. Ideally, for the target test pixel in JSR-MTL, the total reconstruction error via the subtarget dictionary is zero. However, considering the target spectral variability, for the test target pixel, the subtarget dictionary may be hard to well represent the pixel spectrum. Therefore, the total reconstruction error via the subtarget dictionary is nonzero, and the output of the test target pixel is decreased compared with the ideal case. That is to say, the output of the target training samples is large, while the output of the other target training samples is relatively low and the detection output range of the target pixels is large, as shown in Fig. 15(c). Thus, in the low FAR, the number of the detected target pixels is small, and leads to a low detection probability. However, it can also be noticed that, the target box and the background box for JSR-MTL are slightly less than the other cases. In addition, in the larger FAR range, the detection probability of JSR-MTL is larger than that of the other detectors, which still demonstrates the effectiveness of the JSR-MTL.

The reasons for the detection performance of hCEM and rACE are further analyzed as follows. The reason for the detection performance of hCEM may be that the hierarchical suppress process decreases the separability between the weaker target and the background. In detail, in hCEM, ideally the undesired background spectra are gradually suppressed after each layer's detection, while the target spectra will remain unchanged. However, it should be noticed that, the original CEM output of the weak target (e.g., the edge pixels of the targets) is not large and may be not well separated from the background. Thus, the output score of the weak target will be continuously suppressed during each iteration in hCEM, which will lead to the unsatisfactory detection performance. The reason for the unstable detection performance of rACE may lie in the algorithm itself. The rACE aims at iteratively generating a newly target signature with *a priori* target signature via the weighted average spectrum of all pixels spectra. The

weights are set as the ACE detection result during each iteration. Then, the newly target signature is put into the ACE detector resulting the final detection result. It can be noticed that, the finally detection result depends on the original ACE detection result, and the detection error of each iteration will be accumulated. However, the ACE detector cannot guarantee well detection performance for each data set.

C. Parameter Analysis for the JSR-MTL Algorithm

In this section, we examine the effect of the parameters on the detection performance of the JSR-MTL algorithm with the three data sets. We fixed the other parameters and focused on one specific parameter at a time. There are three key parameters in the JSR-MTL algorithm: the detection task number parameter K , the regularization parameter ρ , and the size of the dual window. The range of K was set as [1, 2, 3, 4, 5, 6, 7, 8, 9] and the range of ρ was set as [1, 0.5, 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5}]. For the size of the dual window, the size of the IWR is related to the size of the target. When the size of the IWR is set too large, the background training samples in the OWR will not effectively represent the local background characteristic. Thus, the sizes of the IWR were fixed as above-mentioned 7×7 , 5×5 , and 7×7 for the Synthetic, HYDICE, and AVIRIS data sets, respectively. The range of the size of the OWR was set as [17, 19, 21, 23, 25].

The experimental results are provided through the AUC values, as shown in Figs. 11–13. The x -axes and the y -axes, respectively, represent the value range of the corresponding parameter and the AUC values.

For the Synthetic data set in Fig. 11(a), the AUC value of the JSR-MTL algorithm improves, as the detection task number parameter K increases to 3. After that, the detection performance generally decreases as K increases to 9. For the HYDICE data set in Fig. 11(b), the AUC value improves, as K increases to 5 and then generally decreases as K increases to 9. For the AVIRIS data set in Fig. 11(c), the AUC value improves as K increases to 3, and slowly decreases as K increases to 9; however, the AUC value when K equals to 7 or 8 has anomalous increase. Based on these results, it can be generally concluded that the performance of the JSR-MTL algorithm improves as the detection task number parameter K increases and then begins to decrease after the maximum value. A large detection task number K can affect the detection performance of JSR-MTL. The reason for this may be as follows. A large detection task number K results in too many detection tasks, which will lead to too many unknown coefficients in $\mathbf{W} \in \mathbb{R}^{(N_b+N_t) \times K}$ according to (9); however, the rows of the dictionary $\mathbf{D}^k \in \mathbb{R}^{B^k \times (N_b+N_t)}$ for the multiple representation models in (9) will be significantly decreased according to (7). This can lead to a weakened estimation for the unknown coefficient matrix \mathbf{W} , which will affect the detection performance.

In Fig. 12, for the three data sets, it can be seen that the AUC values of the JSR-MTL algorithm increase when the regularization parameter ρ decreases from 1 to 10^{-1} , and the AUC values gradually decrease as ρ decreases from 10^{-2} to 10^{-5} . These results show that when the regularization parameter ρ is too small or too large, the detection

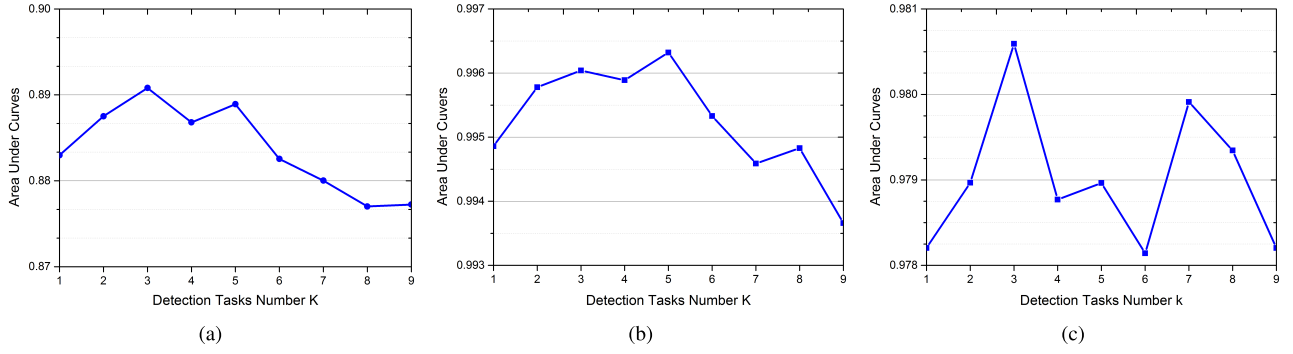


Fig. 11. Detection performance of JSR-MTL versus the detection task number K . (a) Synthetic data set. (b) HYDICE data set. (c) AVIRIS data set.

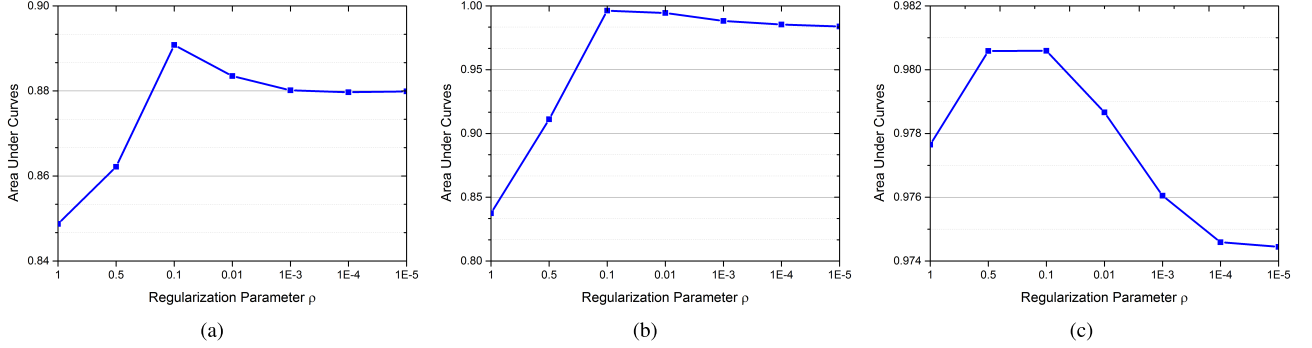


Fig. 12. Detection performance of JSR-MTL versus the regularization parameter ρ . (a) Synthetic data set. (b) HYDICE data set. (c) AVIRIS data set.

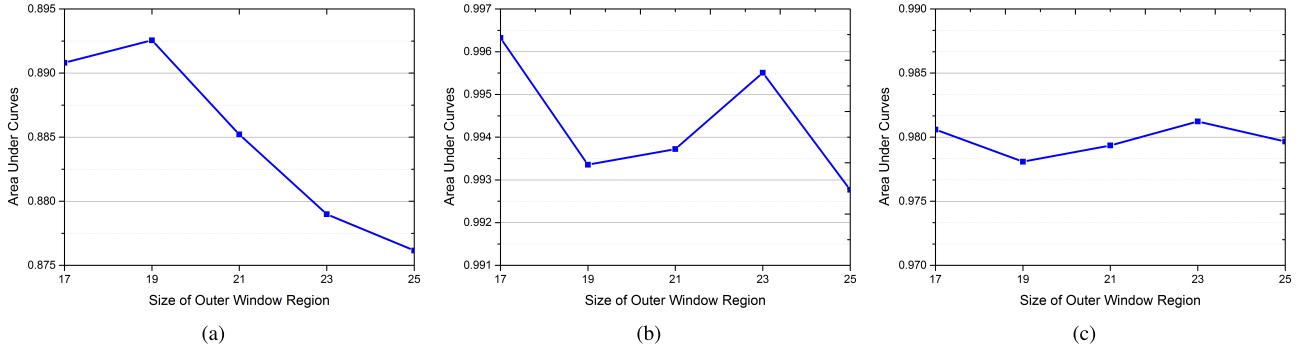


Fig. 13. Detection performance of JSR-MTL versus the size of the OWR. (a) Synthetic data set. (b) HYDICE data set. (c) AVIRIS data set.

performance of JSR-MTL decreases. The reason for this may be as follows. A too small regularization parameter ρ makes the dominant part of (10) become the first term $\|\mathbf{x}^k - \mathbf{D}^k \mathbf{w}^k\|_2^2$, which will weaken the effect of the multiple detection task combination, and will affect the final detection performance of JSR-MTL. A too large regularization parameter ρ makes the dominant part of (10) become the second term, which will weaken the effect of the data representation, and again affect the final detection performance of JSR-MTL.

For the Synthetic data set in Fig. 13(a), the AUC value of the JSR-MTL algorithm improves as the size of the OWR increases to 19, and then decreases. For the HYDICE data set in Fig. 13(b), the best AUC value is obtained when the size of the OWR is 17. The AUC value then decreases when the size of the OWR is 19. After that, the AUC value improves as the size of the OWR increases to 23 and then decreases as the size of OWR increases to 25. For the AVIRIS data set in Fig. 13(c), a good AUC value is obtained when the size of the OWR is 17, and then, the AUC value decreases

when the size of the OWR is 19. After that, the AUC value improves as the size of the OWR increases to 23 and then decreases as the size of the OWR increases to 25. Based on these results, it can be seen that the regular pattern of the size of OWR is not obvious; however, it can still generally concluded that a too large or too small size of OWR can affect the detection performance of JSR-MTL. The reason for this may be as follows. For a too large size of OWR, the background training samples in the OWR will not effectively represent the local background characteristic, which may include some other background materials. For a too small size of OWR, the background training samples in the OWR are not sufficient to represent the local background characteristic.

D. Detection Performance

In this section, we further analyze the detection performance of the proposed JSR-MTL algorithm with the optimal parameter values for the three data sets. The detection task number parameter K was, respectively, set as 3, 5, and 3 for

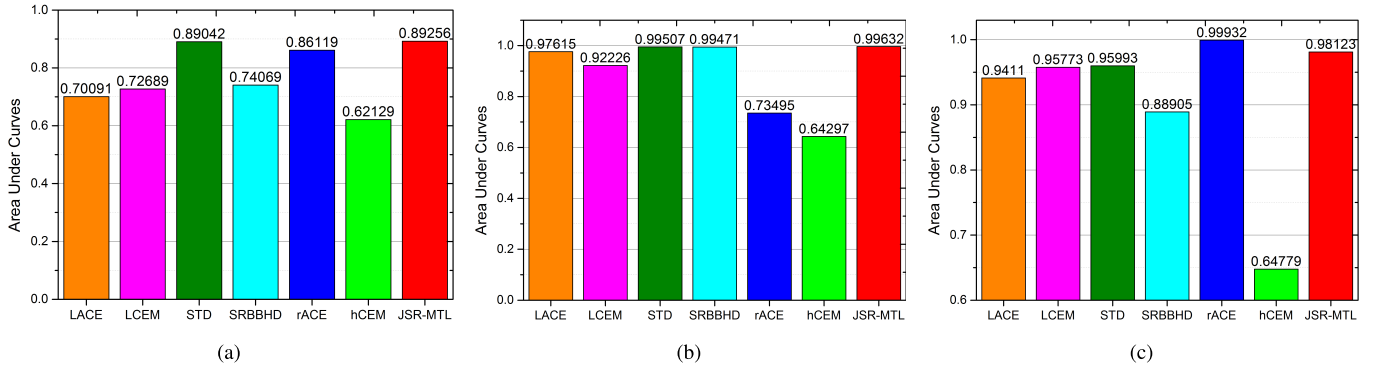


Fig. 14. AUC values of the seven detectors with the optimal values for the three data sets. (a) Synthetic data set. (b) HYDICE data set. (c) AVIRIS data set.

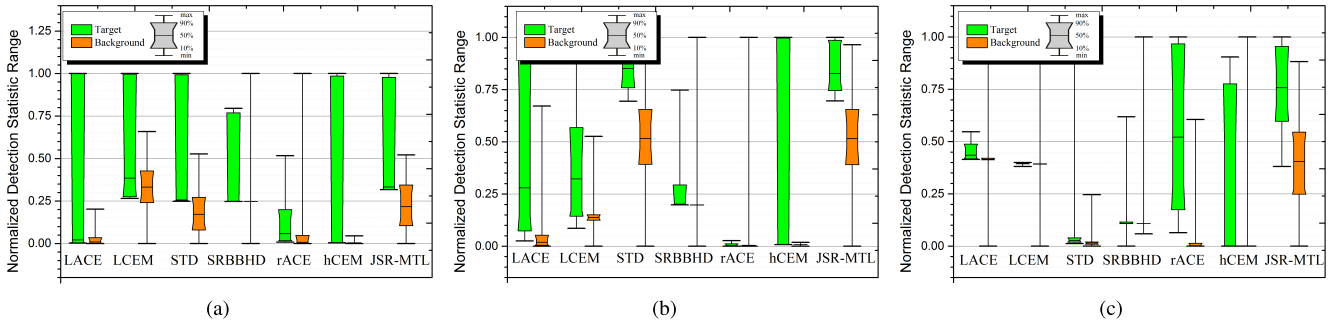


Fig. 15. Separability maps of the seven detectors for the three data sets. (a) Synthetic data set. (b) HYDICE data set. (c) AVIRIS data set.

the three data sets. The regularization parameter ρ was set as 0.1 for the three data sets. The size of the OWR was set as 19, 17, and 23 for the three data sets. For the comparison methods, the parameters were also tested, for example, the dual window size for the local detectors (LACE, LCEM, STD, and SRBBHD), the sparsity level for the sparsity-based detectors (STD, SRBBHD), and so on. The optimal parameter values were experimentally set for the comparison methods.

The AUC values of these detectors were first compared, as shown in Fig. 14. The results generally show that the JSR-MTL algorithm obtains a better detection performance than the other detectors, especially for the Synthetic and HYDICE data sets. For the AVIRIS data set, JSR-MTL does not perform as well as rACE; however, the detection performance of rACE is very different for the three data sets.

The ability of the proposed JSR-MTL algorithm to distinguish target from background was also investigated, and compared with the results of the other detectors. The separability between target and background was evaluated via separability maps, as shown in Fig. 15. After statistical calculation of the detection values of each pixel, boxes were drawn to enclose the main parts of the pixels, excluding the biggest 10% and the smallest 10%. There are target and background columns for each detector. The lines at the top and bottom of each column are the extreme values, which are normalized to [0, 1]. The green boxes illustrate the distribution of the target pixels values, and the line in the middle of the box is the mean of the pixels. In a similar way, the orange boxes enclose the middle 80% of the pixels of the background pixels. The position of the boxes reflects the tendency and compactness of the distribution of the pixels. In other words, the position reflects the separability between target and background.

For the Synthetic data set, as shown in Fig. 15(a), LACE and hCEM can effectively suppress the background information, and rACE and SRBBHD can effectively suppress the middle 80% of the background pixels. The target box and the background box for LCEM, STD, and JSR-MTL are overlapping, but the overlapped region for JSR-MTL is slightly less. For the HYDICE data set, as shown in Fig. 15(b), LACE, LCEM, hCEM, rACE, and SRBBHD can specially effectively suppress the middle 80% of the background pixels. Compared with these detectors, the gaps between the two boxes for STD and JSR-MTL are very obvious, and the gap for JSR-MTL is even larger. For the AVIRIS data set, as shown in Fig. 15(c), LACE, LCEM, hCEM, rACE, and SRBBHD can effectively suppress the middle 80% of the background pixels. STD, JSR-MTL, and rACE can gradually and successively increase the gap between the target box and the background box. Based on these results, it can be seen that the proposed JSR-MTL algorithm can perform well at distinguishing target from background.

V. CONCLUSION

In this paper, the JSR-MTL algorithm has been proposed to utilize the inherent spectral similarity for hyperspectral target detection. Based on the consistency of the MTL technique and the hyperspectral target detection problem, this algorithm introduces the MTL technique to utilize the inherent spectral similarity between adjacent single-band images. A band cross-grouping strategy is introduced to construct multiple detection tasks with commonality. The sparse representation model is then employed for the test pixel in each task, which further guarantees the relevance of the multiple detection tasks. The MTL technique is applied to integrate the multiple detection tasks for target detection. The detection decision rules in favor

of the target class or the background class with the lowest total reconstruction error accumulated over all the tasks.

Experiments in hyperspectral target detection with three data sets confirmed the superior performance of the multiple detection task combination in the proposed JSR-MTL algorithm. With the band cross-grouping strategy, commonality exists between the sub-HSIs, which can be useful when combining the sub-HSIs via MTL for hyperspectral target detection. In general, JSR-MTL presents a better detection performance and better separability than the other common detectors. Still, it is inconvenient to determine the size of the OWR with the local background dictionary. The global dictionary can avoid the size of the OWR and we will work on the global background dictionary construction in the future to further investigate the performance of the JSR-MTL algorithm.

ACKNOWLEDGMENT

The authors thank the handling editor and anonymous reviewers for their careful reading and helpful remarks.

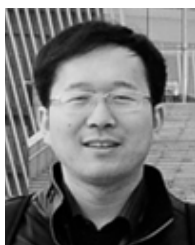
REFERENCES

- [1] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggemanand, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 24–33, Jan. 2014.
- [2] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.
- [3] A. F. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.
- [4] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
- [5] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing-Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2008.
- [6] M. S. Stefanou and J. P. Kerekes, "Image-derived prediction of spectral image utility for target detection applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1827–1833, Apr. 2010.
- [7] N. K. Patel, C. Patnaik, S. Dutta, A. M. Shekh, and A. J. Dave, "Study of crop growth parameters using airborne imaging spectrometer data," *Int. J. Remote Sens.*, vol. 22, no. 12, pp. 2401–2411, 2001.
- [8] 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.
- [9] B. Horig, 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.
- [10] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, Jun. 2009.
- [11] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.
- [12] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.
- [13] N. M. Nasrabadi, "Regularized spectral matched filter for target recognition in hyperspectral imagery," *IEEE Signal Process. Lett.*, vol. 15, no. 1, pp. 317–320, Mar. 2008.
- [14] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.
- [15] X. Geng, L. Ji, K. Sun, and Y. Zhao, "CEM: More bands, better performance," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1876–1880, Nov. 2014.
- [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] 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.
- [18] Z. Huang, Z. Shi, and Z. Qin, "A dimension reduction model for sparse hyperspectral target detection with weighted ℓ_1 -minimization," in *Proc. IEEE 5th Int. Congr. Image Signal Process.*, Oct. 2012, pp. 972–976.
- [19] Z. Huang, Z. Shi, and S. Yang, "Nonlocal similarity regularized sparsity model for hyperspectral target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1532–1536, Nov. 2013.
- [20] 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.
- [21] 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.
- [22] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [23] H. Kwon and N. M. Nasrabadi, "Kernel spectral matched filter for hyperspectral imagery," *Int. J. Comput. Vis.*, vol. 71, no. 2, pp. 127–141, 2007.
- [24] 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.
- [25] L. Capobianco, A. Garzelli, and G. Camps-Valls, "Target detection with semisupervised kernel orthogonal subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3822–3833, Nov. 2009.
- [26] A. O. Kozal, M. Teke, and H. A. Ilgin, "Comparative analysis of hyperspectral dimension reduction methods," in *Proc. IEEE Signal Process. Commun. Appl. Conf.*, Apr. 2013, pp. 1–4.
- [27] H. Gholizadeh, M. J. V. Zoj, and B. Mojaradi, "Impact of informative band selection on target detection performance," *Proc. SPIE*, vol. 8180, p. 81801C, Oct. 2011.
- [28] M. D. Farrell and R. M. Mersereau, "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 192–195, Apr. 2005.
- [29] J. E. Fowler and Q. Du, "Anomaly detection and reconstruction from random projections," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 184–195, Jan. 2012.
- [30] Z. Shi, J. Wu, S. Yang, and J. Zhiguo, "RX and its variants for anomaly detection in hyperspectral images," *Infr. Laser Eng.*, vol. 41, no. 3, pp. 796–802, 2012.
- [31] Y. Dong, B. Du, and L. Zhang, "Target detection based on random forest metric learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 4, pp. 1830–1838, Apr. 2015.
- [32] A. Jalali, P. Ravikumar, S. Sanghavi, and C. Ruan, "A dirty model for multitask learning," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2010, vol. 28, no. 1, pp. 964–972.
- [33] X.-T. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4349–4360, Oct. 2012.
- [34] J. Attenberg, K. Weinberger, and A. Dasgupta, "Collaborative email-spam filtering with the hashing trick," in *Proc. 6th Conf. Email Anti-Spam*, 2009, pp. 1–4.
- [35] O. Chapelle, P. Shivaswamy, K. Weinberger, Y. Zhang, T. Tseng, and S. Vadrevu, "Multi-task learning for boosting with application to Web search ranking," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 1189–1198.
- [36] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, 1997.
- [37] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.
- [38] 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.
- [39] T. Liu, D. Tao, M. Song, and S. Maybank, "Algorithm-dependent generalization bounds for multi-task learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PP, no. 99, p. 1, Mar. 2016.
- [40] Y. Li, X. Tian, T. Liu, and D. Tao, "Multi-task model and feature joint learning," in *Proc. 24th Int. Joint Conf. Artif. Intell. (IJCAI)*, Jun. 2015, pp. 3643–3649.
- [41] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing features: Efficient boosting procedures for multiclass object detection," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2004, pp. 762–769.

- [42] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Mach. Learn.*, vol. 73, no. 3, pp. 243–272, 2008.
- [43] S. Ben-David and R. Schuller, "Exploiting task relatedness for multiple task learning," in *Proc. Conf. Comput. Learn. Theory (COLT)*, 2003, pp. 567–580.
- [44] K. Lounici, M. Pontil, A. B. Tsybakov, and S. V. D. Geer, "Taking advantage of sparsity in multi-task learning," in *Proc. Conf. Comput. Learn. Theory (COLT)*, 2009.
- [45] X. Chen, W. Pan, J. T. Kwok, and J. G. Carbonell, "Accelerated gradient method for multi-task sparse learning problem," in *Proc. 9th IEEE Int. Conf. Data Mining*, Dec. 2009, pp. 746–751.
- [46] J. Liu and J. Ye, "Efficient ℓ_1/ℓ_q norm regularization," *Comp. Sci.*, Sep. 2010.
- [47] P. Tseng, "On accelerated proximal gradient methods for convex-concave optimization," *SIAM J. Optim.*, 2008.
- [48] Z. Zou and Z. Shi, "Hierarchical suppression method for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 330–342, Jan. 2016.
- [49] R. N. Clark, G. A. Swayze, and A. J. Gallagher, "The U.S. geological survey digital spectral library: Version 1: 0.2 to 3.0 microns," U.S. Geol. Survey, Denver, CO, USA, Tech. Rep. 93-592, 1993.
- [50] Y.-C. C. Chang, H. Ren, C.-I. Chang, and R. S. Rand, "How to design synthetic images to validate and evaluate hyperspectral imaging algorithms," *Proc. SPIE*, vol. 6966, p. 69661P, Apr. 2008.
- [51] Y. Zhang, B. Du, and L. Zhang, "Regularization framework for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 313–317, Jan. 2014.
- [52] Y. Gu, C. Wang, S. Wang, and Y. Zhang, "Kernel-based regularized-angle spectral matching for target detection in hyperspectral imagery," *Pattern Recognit. Lett.*, vol. 32, no. 2, pp. 114–119, 2011.
- [53] L. Zhang, L. Zhang, 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.
- [54] T. Wang, B. Du, and L. Zhang, "An automatic robust iteratively reweighted unstructured detector for hyperspectral imagery," *IEEE Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2367–2382, Jun. 2014.



Yuxiang Zhang (S'13) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2011. And she received the Ph.D. degree in the sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2016. She is currently an associate professor with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China. Her research interests include hyperspectral image processing, target detection, sparse representation, and signal processing.



Bo Du (M'10–SM'15) received the B.S. and Ph.D. degrees in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

He is currently a Professor with the School of Computer, Wuhan University. He has authored over 40 research papers published in the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), the IEEE TRANSACTIONS ON

IMAGE PROCESSING (TIP), the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS), and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL). His current research interests include pattern recognition, hyperspectral image processing, and signal processing.

Dr. Du received the best reviewer awards from the IEEE Geoscience and Remote Sensing Society (GRSS) for his service to the IEEE JSTARS in 2011 and the ACM RISING STAR awards for his academic progress in 2015. He was the Session Chair of the International Geoscience And Remote Sensing Symposium 2016 and the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. He also serves as a Reviewer of 20 Science Citation Index magazines including the IEEE TGRS, TIP, JSTARS, and GRSL.



Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He was a Principal Scientist of the China State Key Basic Research Project appointed by the Ministry of National Science and Technology of China to

Lead the Remote Sensing Program in China from 2011 to 2016. He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, and also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He has authored over 450 research papers and five books. He holds 15 patents. His current research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Engineering and Technology, an Executive Member (Board of Governor) of the China National Committee of International Geosphere-Biosphere Programme, an Executive Member of the China Society of Image and Graphics. He was a recipient of the 2010 Best Paper Boeing Award and the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing. He received the best reviewer awards from the IEEE Geoscience and Remote Sensing Society (GRSS) for his service to the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS) in 2012 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2014. He was the General Chair of the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing and the Guest Editor of the IEEE JSTARS. His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium student paper contest in recent years. He regularly serves as the Co-Chair of the series SPIE conferences on multispectral image processing and pattern recognition, conference on Asia remote sensing, and many other conferences. He edits several conference proceedings, issues, and geoinformatics symposiums. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geospatial Information Science*, and the *Journal of Remote Sensing*, and the Guest Editor of the *Journal of Applied Remote Sensing* and the *Journal of Sensors*. He is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and also the Founding Chair of the IEEE GRSS Wuhan Chapter.



Tongliang Liu (SM'13) received the B.E. degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2012, and the Ph.D. degree in information systems from the University of Technology Sydney, Sydney, Australia, in 2016.

He was a Visiting Ph.D. Student with the Barcelona Graduate School of Economics, Barcelona, Spain, and with the Department of Economics, Pompeu Fabra University, Barcelona,

for six months. He is currently a Lecturer with the School of Software and the Faculty of Engineering and Information Technology, University of Technology Sydney. He has authored or coauthored over ten research papers, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the Neural Computation, the International Conference on Machine Learning, the Knowledge Discovery and Data Mining, the International Joint Conference on Artificial Intelligence, and the Association for the Advance of Artificial Intelligence. His current research interests include statistical learning theory, computer vision, and optimization.

Dr. Liu received the Best Paper Award in the IEEE International Conference on Information Science and Technology 2014.