

Article

# Hyperspectral Target Detection via Adaptive Joint Sparse Representation and Multi-Task Learning with Locality Information

Yuxiang Zhang <sup>1</sup>, Ke Wu <sup>1</sup>, Bo Du <sup>2,\*</sup>, Liangpei Zhang <sup>3</sup> and Xiangyun Hu <sup>1</sup>

<sup>1</sup> Hubei Subsurface Multi-Scale Imaging Key Laboratory, Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan 430074, China; zyx\_070504@163.com (Y.Z.); tingke2000@126.com (K.W.); xyhu@cug.edu.cn (X.H.)

<sup>2</sup> School of Computer, Wuhan University, Wuhan 430079, China

<sup>3</sup> State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China; zlp62@whu.edu.cn

\* Correspondence: gunspace@163.com; Tel.: +86-138-7146-1059

Academic Editors: Qi Wang, Nicolas H. Younan, Carlos López-Martínez, Xiaofeng Li and Prasad S. Thenkabail  
Received: 16 April 2017; Accepted: 12 May 2017; Published: 14 May 2017

**Abstract:** Target detection from hyperspectral images is an important problem but encounters a critical challenge of simultaneously reducing spectral redundancy and preserving the discriminative information. Recently, the joint sparse representation and multi-task learning (JSR-MTL) approach was proposed to address the challenge. However, it does not fully explore the prior class label information of the training samples and the difference between the target dictionary and background dictionary when constructing the model. Besides, there may exist estimation bias for the unknown coefficient matrix with the use of  $\ell_1/\ell_2$  minimization which is usually inconsistent in variable selection. To address these problems, this paper proposes an adaptive joint sparse representation and multi-task learning detector with locality information (JSRMTL-ALI). The proposed method has the following capabilities: (1) it takes full advantage of the prior class label information to construct an adaptive joint sparse representation and multi-task learning model; (2) it explores the great difference between the target dictionary and background dictionary with different regularization strategies in order to better encode the task relatedness; (3) it applies locality information by imposing an iterative weight on the coefficient matrix in order to reduce the estimation bias. Extensive experiments were carried out on three hyperspectral images, and it was found that JSRMTL-ALI generally shows a better detection performance than the other target detection methods.

**Keywords:** hyperspectral image; target detection; multi-task learning; sparse representation; locality information

## 1. Introduction

Target detection is essentially a binary classification problem, which aims to separate specific target pixels from various backgrounds with prior knowledge of the targets [1,2]. With the characteristic of high spectral resolution [3], 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 [4,5]. 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 [6–9].

The current target detection methods mainly utilize the detailed spectral information from the HSI data and use different techniques to distinguish the targets and the background, such as the statistical

hypothesis testing theory [10–12], filtering or projection technique [13–15], and sparse representation technique [16–19]. These existing target detection methods, using a uniform vector of test pixel's spectrum as input, usually employ all the original bands to both construct the model and perform the detection. In other words, these methods fully and uniformly utilize the discriminative information within all single-band images, without considering the inherent similarity between the adjacent single-band images of HSI. In fact, the spectral resolution of HSIs is so high that the adjacent single-band images present a great spectral similarity or redundancy, and this spectral redundancy provides an obstacle for effective target detection. Many methods via dimension reduction for hyperspectral target detection have been proposed in order to relieve this problem [20–24]. 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 dimension reduction process. To summarize, there exists a dilemma to simultaneously reduce spectral redundancy and preserve discriminative information for Hyperspectral target detection.

In recent years, the multi-task learning (MTL) technique has attracted much interest [25–28] and has been employed to address the above dilemma for hyperspectral target detection in [29], labeled as the joint sparse representation and multi-task learning (JSR-MTL) approach. The approach explores the spectral similarity between the adjacent single-band images to construct multiple sub-HSIs with a band cross-grouping strategy, which leads to multiple related detection tasks. The approach further explores the similarity between the sub-HSIs to analyze the latent sparse representation of each task. Then multiple sparse representation models via the union target and background dictionary are integrated via a unified multitask learning technique. In this way, the redundancy in each detection task can be effectively avoided; and the spectral information behind the high dimension original HSI dataset fully used, so that the discriminative information is not lost [29].

However, there still exist several problems with the JSR-MTL approach. Firstly, it does not fully incorporate the class label (prior) information of the training samples, which only utilizes the class label information in post-processing when calculating the residuals for each class and ignores the class label information when constructing the sparse representation models. Secondly, it encourages shared sparsity among the columns of the coefficient matrix corresponding to the union dictionary, which lead to the same sparsity constraint among the tasks corresponding to both the target dictionary and background dictionary. However, as the size and the spectral variability of the target dictionary are much different from the background dictionary, it is therefore not appropriate to impose the same sparsity constraint for both the coefficient matrices corresponding to the target dictionary and background dictionary. Finally, it does not take the locality information between the test pixel and all the neighboring background training samples into consideration, which may make a contribution for better signal reconstruction, due to the fact that the samples similar to the test pixel are more likely to be selected for signal reconstruction.

To address the above problems, this paper proposes an adaptive joint sparse representation and multi-task learning detector with locality information (JSRMTL-ALI). The proposed method explores the prior class label information of the training samples to construct two joint sparse representation and multi-task learning models, where the test pixel is separately modeled via the target dictionary or background dictionary. Considering also the great difference between the target dictionary and background dictionary, different regularization strategies encoding the task relatedness are employed for the two joint sparse representation and multi-task learning models based on the target dictionary or background dictionary. Besides, a locality information descriptor is introduced to indicate the difference between the central test pixel and the neighboring background training samples. Additionally, inspired by the idea that the coefficient matrix may have estimation bias in [30,31], since the  $\ell_1/\ell_2$  minimization used in the regularization strategy is usually inconsistent in variable selection, a locality information descriptor-based weight is employed to iteratively constrain the regularization term to reduce the estimation bias.

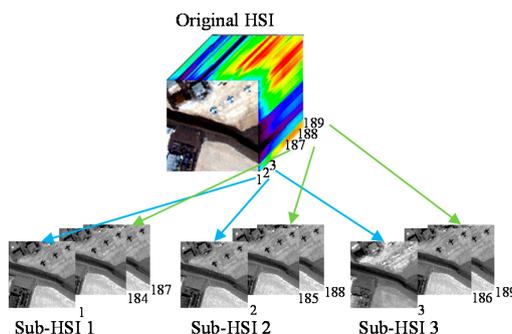
The rest of this paper is organized as follows. Section 2 briefly introduces the original JSR-MTL method. The proposed JSRMTL-ALI method is then presented in Section 3. The experimental results of the proposed method with several HSIs are presented in Section 4. Finally, the discussion and conclusions are drawn in Sections 5 and 6.

## 2. Brief Introduction to the JSR-MTL Method

For the hyperspectral imagery (HSI), as discussed in [29], the adjacent single band images are similar to each other and MTL technology is introduced to utilize the spectral similarity for hyperspectral target detection.

The MTL methodology was proposed by Caruana [28]. It is an inductive transfer method that uses the domain-specific information contained in the training signals of related tasks, which can guarantee that the related tasks can learn from each other and make the inductive transfer method work. There are two key techniques of MTL. One is the construction of multiple tasks with commonality. The other key technique is the relevance analysis of multiple tasks. The multiple tasks can be constructed in various ways, which may depend on the specific application [25,26]. Tasks can be related in various ways. There are two commonly used approaches: (1) tasks may be related by assuming that all the learned functions are close to each other in some norm, such as the linear regression function [25]; and (2) tasks may also be related in that they all share a common underlying representation [32], such as sparsity, a manifold constraint, or a graphical model structure.

In JSR-MTL [29], multiple related detection tasks are constructed through band cross-grouping strategy. In accordance with the band order of the original HSI, the multiple adjacent single-band images are cross-grouped into different groups. Each group then forms a sub-HSI, as shown in Figure 1. Based on the spectral similarity between the adjacent single-band images, multiple sub-HSIs from the original HSI are related with each other. Therefore, these multiple related sub-HSIs naturally correspond to multiple related detection tasks [29].



**Figure 1.** Illustration of the band cross-grouping strategy for the multiple detection tasks. HSI = hyperspectral image.

For the relevance analysis of the multiple tasks, the spectral similarity of the multiple sub-HSIs naturally guarantees the relevance of the multiple detection tasks. Therefore the multiple detection tasks are likely to share a common sparse representation [29], which has shown effectiveness in hyperspectral target detection [16–19].

Considering hyperspectral data  $\mathbf{X} \in R^{h \times w \times B}$  with training samples  $\mathbf{D} = [\mathbf{D}^t, \mathbf{D}^b]$ , where  $\mathbf{D}^t$  is the target dictionary generated via the target training samples  $\{\mathbf{d}_i^t\}_{i=1}^{N_t} \in R^B$ , and  $\mathbf{D}^b$  is the background dictionary generated via the background training samples  $\{\mathbf{d}_i^b\}_{i=1}^{N_b} \in R^B$ .  $N_b$  and  $N_t$  are the number of background and target training samples, respectively. Let  $\mathbf{x}$  be a test pixel in the original HSI, and  $\{\mathbf{x}^k \in R^{B^k}\}_{k=1}^K$  represents the partial test pixel in each sub-HSI.



$$\begin{aligned} \mathbf{x}^1 &= \mathbf{D}^{1b} \mathbf{w}^{1b} + \zeta^{1b} \\ &\vdots \\ \mathbf{x}^K &= \mathbf{D}^{Kb} \mathbf{w}^{Kb} + \zeta^{Kb} \end{aligned} \tag{7}$$

These models can be incorporated into the following joint sparse representation and multi-task learning model

$$\widehat{\mathbf{W}}^b = \arg \min_{\mathbf{w}^{kb}} \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^{kb} \mathbf{w}^{kb}\|_2^2 + \rho^b \Omega(\mathbf{W}^b) \tag{8}$$

where  $\mathbf{W}^b \in R^{N_b \times K}$  is the matrix formed by stacking the vectors  $\mathbf{w}^{kb} \in R^{N_b}$ .  $\Omega(\mathbf{W}^b)$  is the regularization term to further encode the task relatedness.  $\rho^b$  is the regularization parameter to trade off the data fidelity term and the regularization term.

Similarly, a target pixel  $\mathbf{x} \in R^B$  decomposed into  $K$  sub-vectors can be represented as

$$\begin{aligned} \mathbf{x}^1 &= \mathbf{D}^{1t} \mathbf{w}^{1t} + \zeta^{1t} \\ &\vdots \\ \mathbf{x}^K &= \mathbf{D}^{Kt} \mathbf{w}^{Kt} + \zeta^{Kt} \end{aligned} \tag{9}$$

where  $\zeta^{kt}$  is the random noise.  $\mathbf{D}^{kt}$  is the  $B^k \times N_t$  target sub-dictionary.  $\mathbf{w}^{kt} \in R^{N_t}$  is the coefficient sub-vector over the  $k$ -th sub-dictionary  $\mathbf{D}^{kt}$ .

These models can also be incorporated into the following joint sparse representation and multi-task learning model

$$\widehat{\mathbf{W}}^t = \arg \min_{\mathbf{w}^{kt}} \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^{kt} \mathbf{w}^{kt}\|_2^2 + \rho^t \Omega(\mathbf{W}^t) \tag{10}$$

where  $\mathbf{W}^t \in R^{N_t \times K}$  is the matrix formed by stacking the vectors  $\mathbf{w}^{kt} \in R^{N_t}$ .

In the detection problems, we are given a set of training samples with corresponding labels. The above two JSR-MTL models in Equations (8) and (10) make the assumption that a test sample should be represented by atoms from the same classes that the test sample belongs to, which means that the test sample is modeled separately for target and background pixel. Therefore, the above two JSR-MTL models in Equations (8) and (10) are more complete and realistic than the basic JSR-MTL model in [29]. In the above two JSR-MTL models in Equations (8) and (10), the test samples are modeled separately with more reasonable dictionaries, with only the background training samples for the null hypothesis, and the target training samples for the alternative hypothesis. In the case of the basic JSR-MTL model in [29], either the target test samples or the background test samples are represented by both the background and target training samples. In other words, the basic JSR-MTL model in [29] does not fully incorporate the class label (prior) information of the data set; it only utilizes the class label (background and target) information in post-processing when calculating the residuals for each class and ignores it when constructing models and calculating sub-vectors.

What is more, as noted, the regularization terms  $\Omega(\mathbf{W})$  in Equations (8) and (10) are employed to further encode the task relatedness for the background pixel and target pixel, respectively. It can be seen that different assumptions on the task relatedness lead to different regularization terms. Whether the same regularization terms should be used for both the target and background pixel is an interesting problem, which needs further discussion. As we know, in the basic JSR-MTL model [29], multiple partial test pixels  $\mathbf{x}^k$  in each sub-HSI are sparsely represented via the union target and background dictionary  $\mathbf{D}^k = [\mathbf{D}^{kt}, \mathbf{D}^{kb}]$ , and the  $\ell_{2,1}$ -norm of  $\mathbf{W}$  encourages shared sparsity among the columns of the matrix  $\mathbf{W}$  which is formed by stacking the vectors  $\mathbf{w}^k \in R^{N_b + N_t}$ . This will lead to the same sparsity constraint among the columns of the matrix  $\mathbf{W}^b$  and  $\mathbf{W}^t$  corresponding to the target dictionary and background dictionary. This is inappropriate when considering the construction of target and background dictionary. In target detection applications, the number of target pixels is usually small. The target dictionary is

therefore constructed from some of the target pixels in the global image scene [17,18]. 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), which can better represent and capture the spectral signature of the test sample [17,18]. The background dictionary consists of many locally neighboring background training samples whose spectra are likely to be similar to each other. Thus, for the background pixel, multiple columns of the coefficient matrix  $\mathbf{W}^b$  corresponding to multiple background sub-dictionaries are likely to share consistent sparsity among different tasks. However, the case for the target pixel is much different from the background pixel. The size of the target dictionary is smaller than the background dictionary, and the target training samples selected from the whole image are likely to show spectral variability [11]. Therefore it is inappropriate to assume consistent sparsity among multiple columns of the coefficient matrix  $\mathbf{W}^t$  corresponding to multiple target sub-dictionaries. In brief, different regularization terms should be used for the target pixel and background pixel.

For the background pixel in (8), the  $\ell_{2,1}$ -norm can be enforced to the matrix  $\mathbf{W}^b$  as is done in [29]. For the target pixel in Equation (10), the  $\ell_1$ -norm is applied for the matrix  $\mathbf{W}^t$ , which is obtained by the sum of absolute values in the matrix. The difference between  $\ell_1$ -norm and  $\ell_{2,1}$ -norm of the matrix  $\mathbf{W}$  is that,  $\ell_1$ -norm imposes element wise sparsity and does not require consistent feature selection among columns (tasks), while  $\ell_{2,1}$ -norm by grouping rows together can achieve consistent sparsity among different columns (tasks).

Therefore, Equations (8) and (10) can be rewritten as the following adaptive JSR-MTL model, which can be labeled as the JSRMTL-A model:

$$\widehat{\mathbf{W}}^b = \arg \min_{\mathbf{w}^{kb}} \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^{kb} \mathbf{w}^{kb}\|_2^2 + \rho^b \|\mathbf{W}^b\|_{2,1} \tag{11}$$

$$\widehat{\mathbf{W}}^t = \arg \min_{\mathbf{w}^{kt}} \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^{kt} \mathbf{w}^{kt}\|_2^2 + \rho^t \|\mathbf{W}^t\|_1 \tag{12}$$

### 3.2. Locality Information Descriptor-Based Weight

The background dictionary is further discussed in this section. It can be seen from Equation (11) that, all the training samples (atoms) in the background dictionary are treated equally for signal representation, which ignores locality information, such as differences between the neighboring pixels and the central test pixel. However, some surrounding pixels may be quite similar to the center pixel and are likely to be selected for signal representation; some are quite different from the center pixel, such as the pixel which has a different kind of material from the central pixel, which should be limited or even prohibited for signal representation. The differences between the test pixel and the target atoms are not discussed here due to the small size of the target atoms and the global target atoms selection method.

To preserve the locality difference between the central test pixel and the neighboring background atoms, a distance based locality information descriptor is introduced, which can be expressed as

$$\alpha_i^k = \exp\left(\frac{\|\mathbf{x}^k - \mathbf{d}_i^{kb}\|_2^2}{2}\right) \tag{13}$$

where  $\alpha_i^k$  is the sample-specific descriptor for training sample  $i (i = 1, 2, \dots, N_b)$  in the  $k$ -th background sub-dictionary  $\mathbf{D}^{kb}$ . It is clear that a smaller  $\alpha_i^k$  indicates  $\mathbf{x}^k$  is more similar to the atom  $\mathbf{d}_i^{kb}$ , and vice versa.

Once the above descriptor is included, all the atoms in the background dictionary will be adaptively treated for signal representation via the  $\ell_{2,1}$ -norm. However, there may still exist estimation bias for the signal representation. As stated in [30,31] the estimation bias can be large due to the fact that the  $\ell_1/\ell_2$  minimization is generally inconsistent in variable selection. Many efforts have been made to reduce the estimation bias, such as adaptive Lasso method [30] and the reweighted  $\ell_1$

minimization [31]. Inspired by the reweighted  $\ell_1$  minimization, a weight strategy on the  $\ell_{2,1}$ -based regularization term is introduced to reduce the estimation bias as follows.

$$\widehat{\mathbf{W}}^b = \arg \min_{\mathbf{w}^{kb}} \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^{kb} \mathbf{w}^{kb}\|_2^2 + \rho^b \|\Psi \odot \mathbf{W}^b\|_{2,1} \tag{14}$$

where  $\Psi = \{\Psi_i^k\}$  is the weighting matrix, and  $\Psi_i^k$  is the weight for atom  $i (i = 1, 2, \dots, N_b)$  and column (task)  $k (k = 1, 2, \dots, K)$ ,  $\odot$  denotes modifying the element in the coefficients matrix  $\mathbf{W}^b$  by iteratively multiplying a weight during the coefficient optimization.

In order to impose a relatively higher penalty for smaller coefficients and a lower penalty for larger coefficients, the weight can be computed as inversely proportional to the sparse coefficient

$$\varphi_i^k = \frac{1}{|w_i^{kb}|} \tag{15}$$

Combining the above locality information descriptor and weight strategy, we obtain the locality information descriptor-based weight defined as

$$\Psi_i^k = \frac{\varphi_i^k \alpha_i^k}{\max_{i,k} \varphi_i^k \alpha_i^k} \tag{16}$$

### 3.3. Model Optimization

For the model optimization, we use the popular accelerated proximal gradient (APG) algorithm [33,34] to efficiently solve the problem in Equations (12) and (14). The APG algorithm alternately updates a matrix sequence  $\widehat{\mathbf{W}}^t = [w_i^{k,t}]$  and an aggregation matrix sequence  $\widehat{\mathbf{V}}^t = [v_i^{k,t}]$ .

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

$$\begin{aligned} \widehat{\mathbf{w}}^{k,t+1} &= \widehat{\mathbf{v}}^{k,t} - \eta^t \nabla^{k,t}, \quad t \geq 1, \\ \widehat{\mathbf{W}}^{t+1} &= f(\widehat{\mathbf{W}}^{t+1}), k = 1, 2, \dots, K \end{aligned} \tag{17}$$

where  $\nabla^{k,t} = -(\mathbf{D}^k)^T \mathbf{x}^k + (\mathbf{D}^k)^T \mathbf{D}^k \widehat{\mathbf{v}}^{k,t}$ ,  $\eta^t = 1/2^t$  is the step size.  $f(\cdot)$  is a function of  $\widehat{\mathbf{W}}^{t+1} = [w_i^{k,t+1}]$ , which has a different format for (12) and (14).

For (12), the matrix  $\widehat{\mathbf{W}}^{t+1}$  can be updated as follows

$$\widehat{\mathbf{W}}^{t+1} = \begin{cases} P_{\Omega_1}(\widehat{\mathbf{W}}^{t+1} - \frac{\rho}{2^t}), & \Omega_1 : (\widehat{\mathbf{W}}^{t+1})_{i,k \in \Omega_1} > \frac{\rho}{2^t} \\ P_{\Omega_2}(\widehat{\mathbf{W}}^{t+1} + \frac{\rho}{2^t}), & \Omega_2 : (\widehat{\mathbf{W}}^{t+1})_{i,k \in \Omega_2} < -\frac{\rho}{2^t} \\ P_{\Omega_3}(\mathbf{0} \in R^{N_i \times K}), & \Omega_3 : (\Omega_1 \cup \Omega_2)^\perp, \quad i = 1, 2, \dots, N_b \end{cases} \tag{18}$$

where  $P_\Omega$  is the projection of a matrix onto an entry set, and  $\Omega$  is the index of the entry set.

For (14), the matrix  $\widehat{\mathbf{W}}^{t+1}$  can be updated as follows

$$\begin{aligned} \widehat{\mathbf{w}}_i^{t+1} &= \left[ 1 - \frac{\rho}{2^t \|\widehat{\mathbf{w}}_i^{t+1}\|_2} \right]_+ \widehat{\mathbf{w}}_i^{t+1}, \quad i = 1, 2, \dots, N_b \\ \widehat{w}_i^{k,t+1} &= \widehat{w}_i^{k,t+1} \times \frac{\alpha_i^k / |\widehat{w}_i^{k,t+1}|}{\max_{i=1,2,\dots,N_b, k=1,2,\dots,K} \alpha_i^k / |\widehat{w}_i^{k,t+1}|} \end{aligned} \tag{19}$$

where  $[\cdot]_+ = \max(\cdot, 0)$ .

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

$$\widehat{\mathbf{V}}^{t+1} = (1 + \tau^t)\widehat{\mathbf{W}}^{t+1} - \tau^t\widehat{\mathbf{W}}^t \tag{20}$$

where the sequence  $\tau^t$  is conventionally set to  $\tau^t = 2(t - 1)/(1 + \sqrt{1 + 4t^2})$ , as applied in our implementation.

The optimization methods for the problem in Equations (12) and (14) can be summarized as Algorithms 1 and 2, respectively.

**Algorithm 1.** The Coefficients over Target Dictionary Optimization Algorithm.

Input: Data  $\{\mathbf{D}^k, \mathbf{x}^k\}_{k=1}^K$ , regularization parameter  $\rho$

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

Step (1): Initialization:  $\widehat{\mathbf{w}}^{k,0} = (\mathbf{D}^k)^T \mathbf{x}^k$ ,  $\widehat{\mathbf{v}}^{k,0} = \widehat{\mathbf{w}}^{k,0}$ ,  $\tau^0 = -1$ ,  $t = 0$

Step (2): Repeat {Main loop}

- a)  $\widehat{\mathbf{w}}^{k,t+1} = \widehat{\mathbf{v}}^{k,t} - \frac{1}{2^t} \left[ -(\mathbf{D}^k)^T \mathbf{x}^k + (\mathbf{D}^k)^T \mathbf{D}^k \widehat{\mathbf{v}}^{k,t} \right]$ ,  $k = 1, \dots, K$
- b)  $\widehat{w}_i^{k,t+1} = \begin{cases} \widehat{w}_i^{k,t+1} - \rho/2^t, & \Omega_1 \\ \widehat{w}_i^{k,t+1} + \rho/2^t, & \Omega_2 \\ 0, & \Omega_3 \end{cases}$
- c)  $\tau^t = \frac{2(t-1)}{1+\sqrt{1+4t^2}}$ ,  $\widehat{\mathbf{v}}^{k,t+1} = (1 + \tau^t)\widehat{\mathbf{w}}^{k,t+1} - \tau^t\widehat{\mathbf{w}}^{k,t}$
- d)  $t = t + 1$

Until: convergence is attained

**Algorithm 2.** The Coefficients over Background Dictionary Optimization Algorithm.

Input: Data  $\{\mathbf{D}^k, \mathbf{x}^k\}_{k=1}^K$ , regularization parameter  $\rho$ , locality information descriptor  $\{\alpha_i^k\}$

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

Step (1): Initialization:  $\widehat{\mathbf{w}}^{k,0} = (\mathbf{D}^k)^T \mathbf{x}^k$ ,  $\widehat{\mathbf{v}}^{k,0} = \widehat{\mathbf{w}}^{k,0}$ ,  $\tau^0 = -1$ ,  $t = 0$

Step (2): Repeat {Main loop}

- a)  $\widehat{\mathbf{w}}^{k,t+1} = \widehat{\mathbf{v}}^{k,t} - \frac{1}{2^t} \left[ -(\mathbf{D}^k)^T \mathbf{x}^k + (\mathbf{D}^k)^T \mathbf{D}^k \widehat{\mathbf{v}}^{k,t} \right]$ ,  $k = 1, \dots, K$
- b)  $\widehat{\mathbf{w}}_i^{t+1} = \left[ 1 - \frac{\rho}{2^t \|\widehat{\mathbf{w}}_i^{t+1}\|_2} \right]_+ \widehat{\mathbf{w}}_i^{t+1}$ ,  $i = 1, 2, \dots, N_b$
- $\widehat{w}_i^{k,t+1} = \widehat{w}_i^{k,t+1} \times \frac{\alpha_i^k / |\widehat{w}_i^{k,t+1}|}{\max_{i^k} \alpha_i^k / |\widehat{w}_i^{k,t+1}|}$
- c)  $\tau^t = \frac{2(t-1)}{1+\sqrt{1+4t^2}}$ ,  $\widehat{\mathbf{v}}^{k,t+1} = (1 + \tau^t)\widehat{\mathbf{w}}^{k,t+1} - \tau^t\widehat{\mathbf{w}}^{k,t}$
- d)  $t = t + 1$

Until: convergence is attained

3.4. Final Sketch of the JSRMTL-ALI Detector

Once given the recovery of the coefficient vectors  $\widehat{\mathbf{w}}^{kb}$  and  $\widehat{\mathbf{w}}^{kt}$  corresponding to the background dictionary  $\mathbf{D}^{kb}$  and target dictionary  $\mathbf{D}^{kt}$  for each task, 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 sub-dictionaries  $\{\mathbf{D}^{kb}\}_{k=1}^K$  and  $\{\mathbf{D}^{kt}\}_{k=1}^K$  as follows.

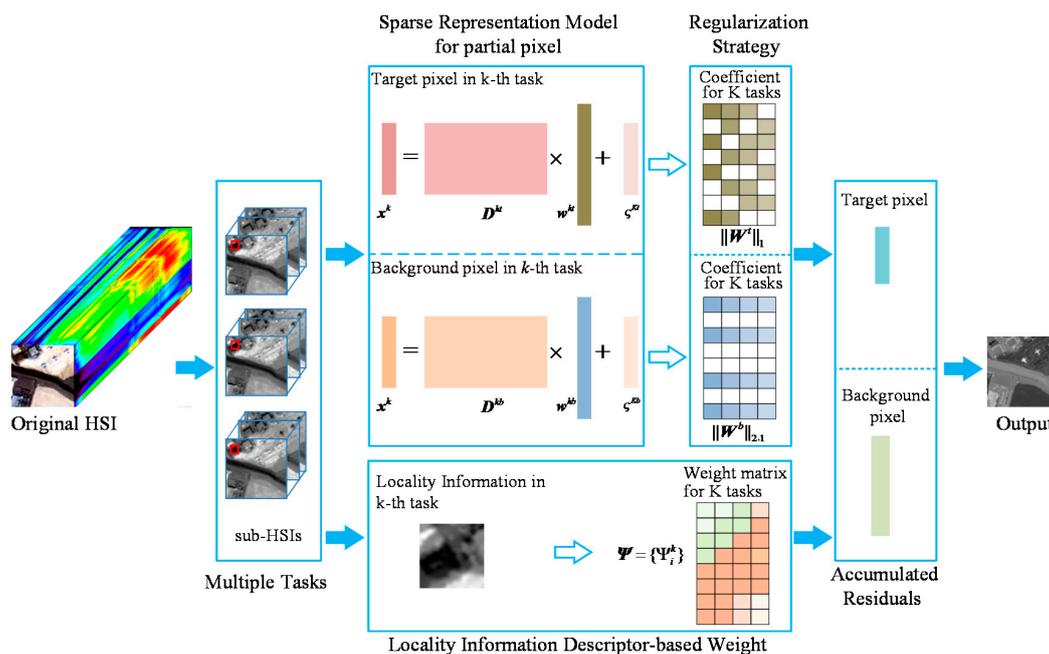
$$r^b = \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^{kb} \widehat{\mathbf{w}}^{kb}\|_2$$

$$r^t = \sum_{k=1}^K \|\mathbf{x}^k - \mathbf{D}^{kt} \widehat{\mathbf{w}}^{kt}\|_2$$
(21)

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

$$D(\mathbf{x}) = r^b - r^t$$
(22)

Finally, a visual illustration of the proposed LWJSR-MTL algorithm for HSIs is shown in Figure 2. Given a hyperspectral image, multiple sub-HSIs are extracted via the band cross-grouping strategy. We construct the multiple-signals for each pixel  $\{\mathbf{x}^k\}_{k=1}^K$ , multiple background dictionary  $\{\mathbf{D}^{kb}\}_{k=1}^K$  with the local dual window, and multiple target dictionary  $\{\mathbf{D}^{kt}\}_{k=1}^K$  via the target training samples. Each pixel is represented by the multi-task sparse representation model via the target dictionary and background dictionary, respectively. The coefficient matrices corresponding to the target dictionary and dictionary are recovered via the Algorithms 1 and 2, respectively. Finally, the detection decision rules in favor of the target class or the background class with the lowest total reconstruction error difference accumulated over all the tasks.



**Figure 2.** Schematic illustration of the adaptive joint sparse representation and multi-task learning detector with locality information (JSRMTL-ALI) algorithm.

## 4. Experiments and Analysis

### 4.1. Dataset Description

Three hyperspectral datasets were used in this study to evaluate the effectiveness of the proposed detector introduced in Section 3.

The first dataset 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. 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  $100 \times 100$  pixels was used for the experiments. The image scene is shown in Figure 3a. There are three planes in the image, which consist of 58 pixels, as shown in Figure 3b. We selected one pixel from each plane as the target atoms,  $N_t = 3$ .

The second dataset was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines test site in Northwest Indiana and consists of  $145 \times 145$  pixels and 224 spectral reflectance bands in the wavelength range  $0.4\text{--}2.5 \mu\text{m}$ . The false color image of the Indian Pines image is shown in Figure 4a. We also reduced the number of bands to 200 by removing bands covering the regions of water absorption: 104–108, 150–163, and 220, as referred to in [35]. This image contains 16 ground-truth classes via a ground truth labels, and the stone-steel-towers was selected as the target of interest to be detected, which has 93 pixels, as shown in Figure 4b. We selected three pixels from the target as the target atoms,  $N_t = 3$ .

The third data set was acquired by the Nuance Cri hyperspectral sensor. This sensor can acquire imagery with a spectral resolution of 10 nm. The image scene covers an area of  $400 \times 400$  pixels, as shown in Figure 5a, with 46 spectral bands in wavelengths ranging from 650 to 1100 nm. There are ten rocks located in the grassy scene, which consist of 1254 pixels, as shown in Figure 5b. We selected one pixel from each rock as the target atoms,  $N_t = 10$ .

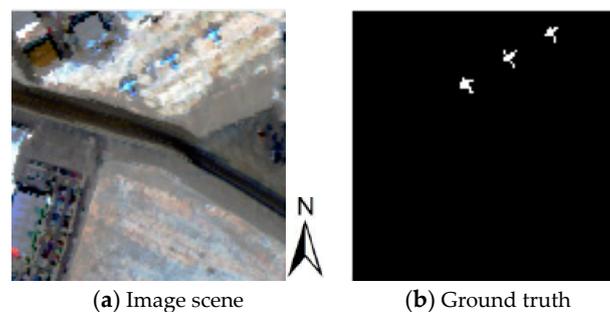


Figure 3. The AVIRIS dataset.

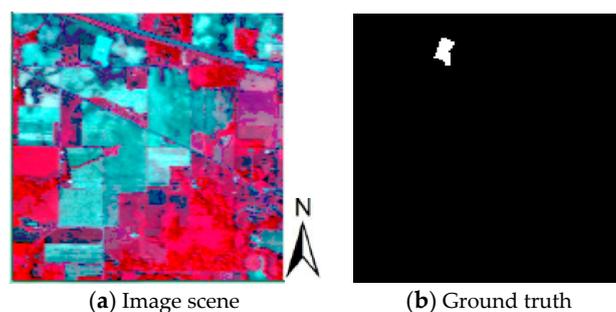


Figure 4. The Indian dataset.

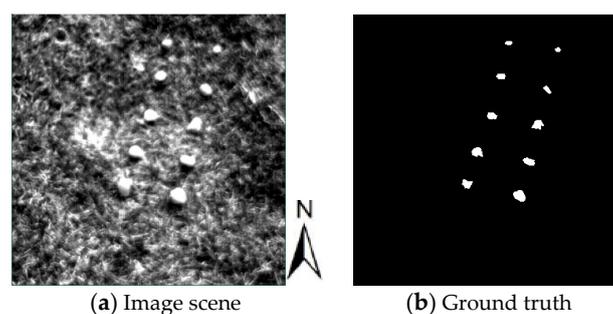


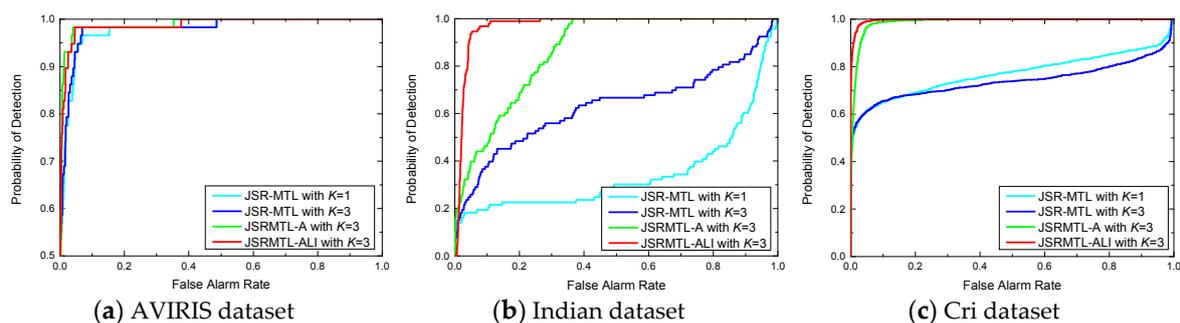
Figure 5. The Cri dataset.

#### 4.2. Evaluation of JSRMTL-ALI Model

Firstly, the effectiveness of the JSRMTL-ALI model was investigated and compared with the original JSR-MTL model in Equation (3) and the adaptive JSR-MTL (labeled as JSRMTL-A) model in Equations (11) and (12). We took three detection tasks ( $K = 3$ ) as an example, and the JSR-MTL with  $K = 1$  which indicates the detection performance without the multi-task learning technique. For simplicity, all regularization parameters used in the four models are set with the same value ( $\rho = 0.1$ ). The sizes of the OWR for the three datasets were respectively set as  $17 \times 17$ ,  $23 \times 23$ , and  $23 \times 23$ . The sizes of the IWR are related to the size of the target, and were set as  $7 \times 7$ ,  $15 \times 15$ , and  $15 \times 15$  for the AVIRIS, Indian, and Cri datasets, respectively. The numbers of the background training samples for the three datasets were therefore  $N_b = 240$ ,  $N_b = 304$ , and  $N_b = 304$ , respectively. The detection performance for the four models with three datasets are provided by the area under the receiver operation characteristics (ROC) curves, as shown in Figure 6.

For the AVIRIS dataset, as shown Figure 6a, the ROC curve of JSRMTL-ALI is not above that of JSRMTL-A, however, it is above that of JSR-MTL with  $K = 3$  and  $K = 1$ . For the Indian dataset, as shown in Figure 6b, the ROC curve of JSRMTL-ALI is always above those of the other models, and the ROC curve of JSRMTL-A is always above that of JSR-MTL with  $K = 3$  and  $K = 1$ . For the Cri dataset, as shown Figure 6c, the ROC curve of JSRMTL-ALI is successively above that of JSRMTL-A, JSR-MTL with  $K = 1$ , and JSR-MTL with  $K = 3$ .

Overall, the results show that the performance of the JSR-MTL model is generally better than that without the multi-task learning technique, especially for the AVIRIS and Indian datasets. The JSRMTL-A model can also obtain a better detection performance compared to the JSR-MTL model for all three datasets, which shows the effectiveness of the adaptive JSR-MTL (JSRMTL-A) model. This result demonstrates that it is useful to explore the prior class label information of the training samples and the difference between the target dictionary and background dictionary for hyperspectral target detection. What is more, the JSRMTL-ALI model can further improve the detection performance of the JSRMTL-A model, especially for the Indian and Cri datasets. This result confirms that the locality information descriptor-based weight can improve the detection performance, which can remain as the locality information between the central test pixel and neighboring background training samples, and also reduce the estimation bias caused by the  $\ell_1/\ell_2$  minimization. In addition, we can further adjust the number of detection tasks, the regularization parameter, and the window size to obtain an even better performance.



**Figure 6.** Receiver operation characteristic (ROC) curves for the effectiveness investigation of JSRMTL-ALI model.

#### 4.3. Parameter Analysis for the JSRMTL-ALI Algorithm

In this section, we examine the effect of the parameters on the detection performance of the JSRMTL-ALI algorithm with the three datasets. We fixed the other parameters and focused on one specific parameter at a time. There are three key parameters in the JSRMTL-ALI algorithm: the detection task number parameter  $K$ , the regularization parameter  $\rho$ , and the size of the dual window. As is done in [29], the range of  $K$  was set as  $[1, 2, 3, 4, 5, 6, 7, 8, 9]$  and the range of  $\rho$  was set as

[1, 0.5, 10<sup>-1</sup>, 10<sup>-2</sup>, 10<sup>-3</sup>, 10<sup>-4</sup>, 10<sup>-5</sup>]. 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, 15 × 15, and 15 × 15 for the AVIRIS, Indian, and Cri datasets, respectively. The range of the size of the OWR for AVIRIS dataset was set as [17, 19, 21, 23, 25], and it was set as [23, 25, 27, 29, 31] for the Indian and Cri datasets. The experimental results are provided through the AUC values, as shown in Figures 7–9. The X-axes and the Y-axes respectively represent the value range of the corresponding parameter and the AUC values.

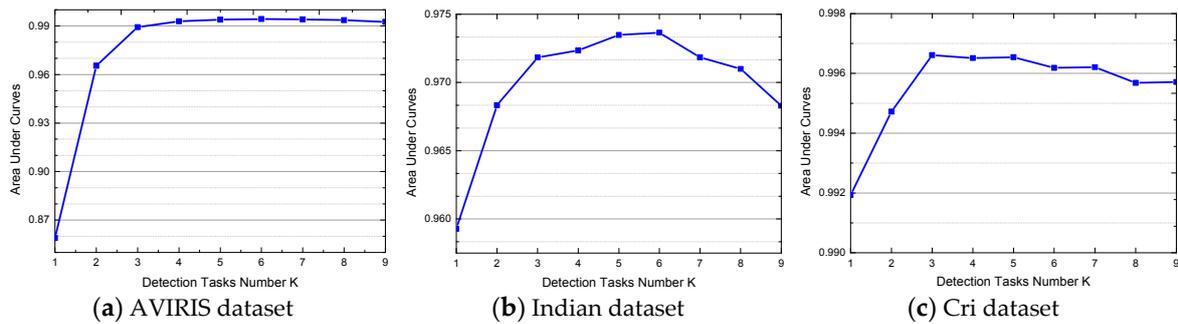


Figure 7. Detection performance of JSRMTL-ALI versus the detection task number  $K$ .

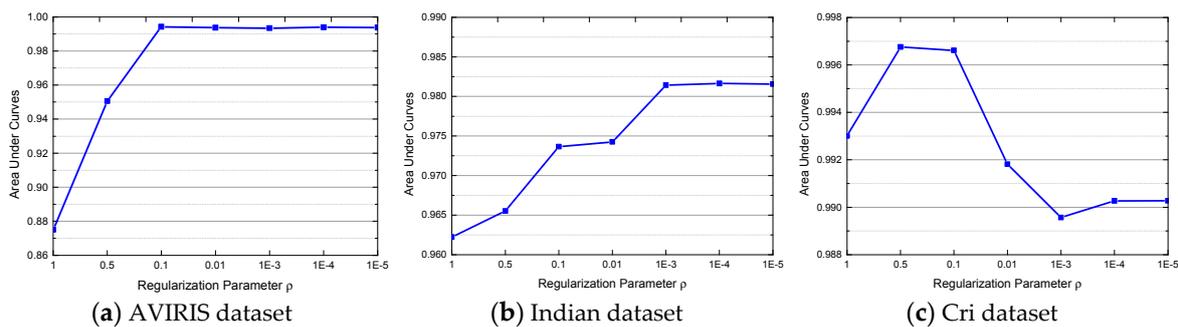


Figure 8. Detection performance of JSRMTL-ALI versus the detection task number  $\rho$ .

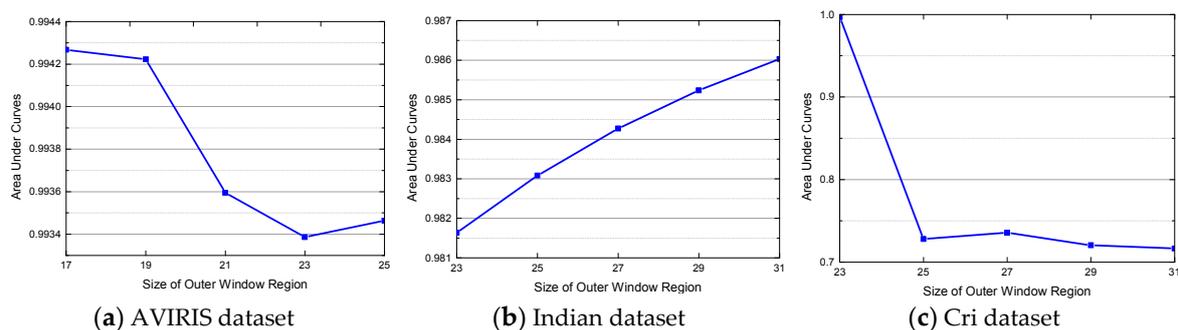


Figure 9. Detection performance of JSRMTL-ALI versus the size of the outer window region (OWR).

For the AVIRIS dataset in Figure 7a, the AUC value of the JSRMTL-ALI algorithm improves as the detection task number parameter  $K$  increases to 6. After that, the detection performance slowly decreases as  $K$  increases to 9. For the Indian dataset in Figure 7b, the AUC value generally improves as  $K$  increases to 6 and then decreases as  $K$  increases to 9. For the Cri dataset in Figure 7c, the AUC value improves as  $K$  increases to 3, gently decreases as  $K$  increases to 9. Based on these results, it can be generally concluded that the performance of the JSRMTL-ALI algorithm improves as the detection

task number parameter  $K$  increases and then begins to decrease after the maximum value. The reason for this may be as follows. As discussed in [29], a large detection task number  $K$  results in too many detection tasks which leads to too many unknown coefficients; however, the rows of the dictionary for the multiple representation models will be significantly decreased. This can lead to a weakened estimation for the unknown coefficient matrix, which will affect the detection performance. Besides, the advantage of the multi-task learning technique for hyperspectral image lies in the fact that it can explore the relatedness within the corresponding single-band images in the same position in each sub-HIS. However, a large detection task number is highly likely to reduce the relatedness within multiple sub-HSIs, and the effectiveness of MTL will decrease in return.

For the AVIRIS dataset in Figure 8a, the AUC value of the JSRMTL-ALI algorithm improves when the regularization parameter  $\rho$  decreases from 1 to  $10^{-1}$ , and the AUC values gradually decrease as  $\rho$  decreases from  $10^{-2}$  to  $10^{-5}$ . For the Indian dataset in Figure 8b, the AUC value improves as  $\rho$  decreases from 1 to  $10^{-4}$ , and decreases as  $\rho$  decreases to  $10^{-5}$ . For the Cri dataset in Figure 8c, the AUC value improves as  $\rho$  decreases from 1 to 0.5, and generally decreases as  $\rho$  decreases to  $10^{-5}$ . Based on these results, it can be generally concluded that a too small or too large regularization parameter  $\rho$  can decrease the detection performance of JSRMTL-ALI. The reasons may be listed as follows. A too small regularization parameter makes the dominant part of Equations (11) and (14) 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 JSRMTL-ALI. A too large regularization parameter makes the dominant part of Equations (11) and (14) become the second term, which will weaken the effect of the data representation, and again affect the final detection performance of JSRMTL-ALI.

For the AVIRIS dataset in Figure 9a, the AUC value of the JSRMTL-ALI algorithm decreases as the size of the OWR increases to 23, and then slightly increases as the size of the OWR increases to 25. For the Indian dataset in Figure 9b, the AUC value improves as the size of the OWR increases to 31. For the Cri dataset in Figure 9c, the AUC value of the JSRMTL-ALI algorithm generally decreases as the size of the OWR increases to 31. Based on these results, it can be seen that the detection performance decreases as the size of Outer Window Region (OWR) increases for the AVIRIS dataset and Cri dataset, while the case is totally different for the Indian dataset. Although the regular patten of the size of the OWR for all datasets is not obvious; it can still generally be concluded that a too large or too small size of OWR can affect the detection performance of JSRMTL-ALI. 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. Both cases will lead to a weakened detection performance. Therefore, it is not easy to select a proper value for the size of OWR in a practical application.

#### 4.4. Detection Performance

In this section, the detection performance of the proposed JSRMTL-ALI algorithm was further analyzed and compared with traditional detectors of local adaptive coherence/cosine estimator (LACE), local constrained energy minimization (LCEM), reweighted adaptive coherence/cosine estimator (rACE) [10], hierarchical constrained energy minimization (hCEM) [14], STD [17], RBBHD [18], and JSR-MTL [29]. The parameters of the JSRMTL-ALI algorithm were set as the optimal parameter values for the three datasets. The detection task number parameter  $K$  was respectively set as 6, 6, and 3 for the three datasets. The regularization parameter  $\rho$  was respectively set as  $10^{-1}$ ,  $10^{-4}$ , 0.5 for the three datasets. The size of the OWR was set as 17, 31, and 23 for the three datasets. For the comparison methods, the parameters were also tested, such as the sparsity level for the sparsity-based detectors (STD, SRBBHD), and so on. The optimal parameter values were experimentally set for the comparison methods. For all the detectors, we used the same given target spectra as a priori target spectra. 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, JSR-MTL, and JSRMTL-ALI. The detection performance of the eight detectors are provided through the receiver operation characteristics (ROC) curves, as shown in Figure 10.

For the AVIRIS dataset, as shown in Figure 10a, the ROC curve of JSRMTL-ALI is above that of the other detectors, except for rACE. For the Indian dataset, as shown in Figure 10b, the ROC curve of JSRMTL-ALI is always above those of the other detectors. For the Cri dataset, as shown in Figure 10c, rACE and hCEM obtain the best result, and the ROC curve of JSRMTL-ALI is above those of the rest of the detectors.

Overall, the results generally show that the JSRMTL-ALI algorithm obtains a better detection performance than the other detectors, especially for the Indian dataset. For the AVIRIS and Cri dataset, JSRMTL-ALI does not perform as well as rACE or hCEM. However, the detection performances of rACE and hCEM are much different for the three datasets and a robust detection performance is not shown. For example, rACE obtains a good performance for AVIRIS and Cri datasets, while obtains a weak performance for the Indian dataset. hCEM obtains a good performance for Indian and Cri datasets, while obtaining a weak performance for the AVIRIS dataset.

The separability between target and background was evaluated via separability maps, as shown in Figure 11. 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 [1]. The orange boxes illustrate the distribution of the target pixel values, and the line in the middle of the box is the mean of the pixels. In a similar way, the green 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 AVIRIS dataset, as shown in Figure 11a, STD and rACE can effectively suppress the background information; and LACE, LCEM, STD, SRBBHD, rACE and hCEM can effectively suppress the middle 80% of the background pixels. Compared to these detectors, the gaps between the target box and the background box for rACE, JSR-MTL, and JSRMTL-ALI are very obvious, and the gap for JSRMTL-ALI is larger than JSR-MTL. The target box and the background box for rACE, JSR-MTL, and JSRMTL-ALI are overlapping, but the overlapped region for JSR-MTL is slightly less. For the Indian dataset, as shown in Figure 11b, rACE can specially, effectively suppress the middle 80% of the background pixels. Compared to these detectors, the gap between the two boxes for JSRMTL-ALI is very obvious, and the two boxes for the other detectors are overlapping. For the Cri dataset, as shown in Figure 11c, JSRMTL-ALI, hCEM, 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 JSRMTL-ALI algorithm can perform well at the distinguishing target from the background.

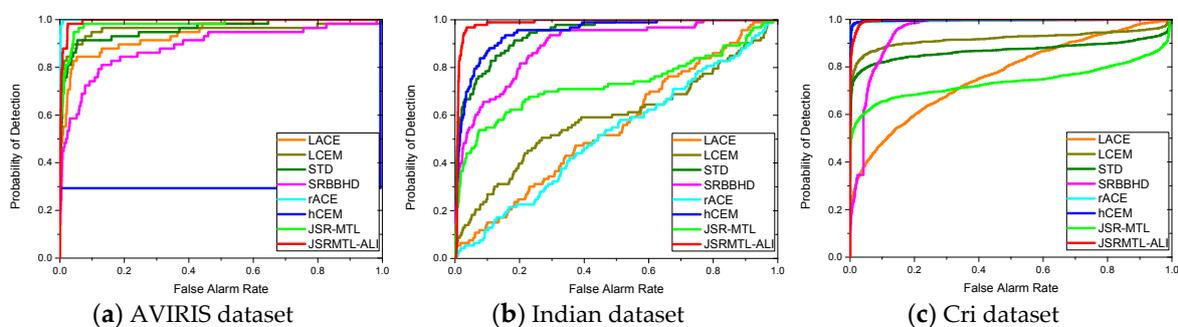


Figure 10. Detection performance of eight detectors for three datasets.

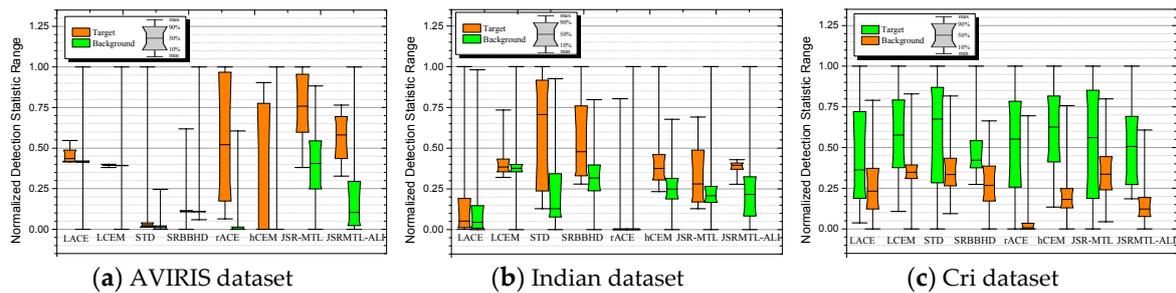


Figure 11. The separability maps of eight detectors for three datasets.

Finally, 2-D plots of the detection map of all the comparison algorithms with the three data sets are shown in Figures 12–14. For the AVIRIS dataset, as shown in Figure 12, we can see that the proposed JSRMLT-ALI shows high statistical values for the target pixels as well as STD, rACE, and JSR-MTL. However, compared with JSRMLT-ALI, STD and JSR-MTL also show high values for some tree or grass pixels, particularly in the bottom/right left corner in the image. Also rACE shows a good performance for suppressing background. For the Indian dataset, as shown in Figure 13, none of these detectors show a clearly distinguishable statistic map, but JSRMLT-ALI generally shows relatively higher statistical values for the target pixels compared with all the other detectors. For the Cri dataset, as shown in Figure 14, the proposed JSRMLT-ALI shows low statistical values for the background pixels as well as rACE and hCEM. However, compared with rACE, JSRMLT-ALI does not show a clearly distinguishable statistic map between target and background.

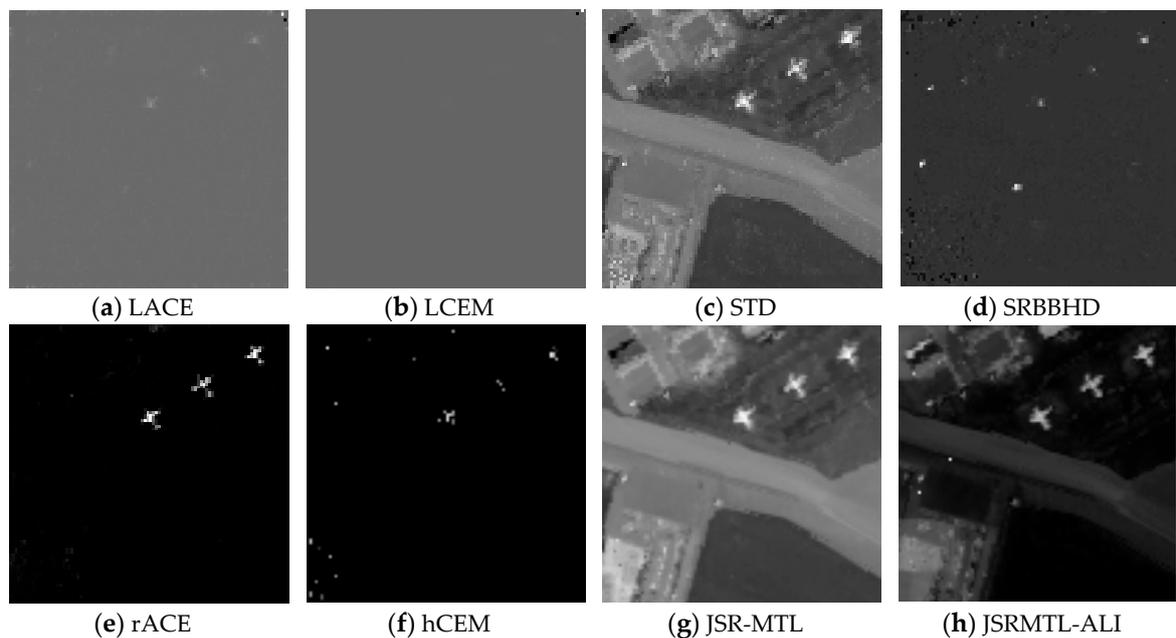


Figure 12. Two-dimensional plots of the detection map for the AVIRIS dataset.

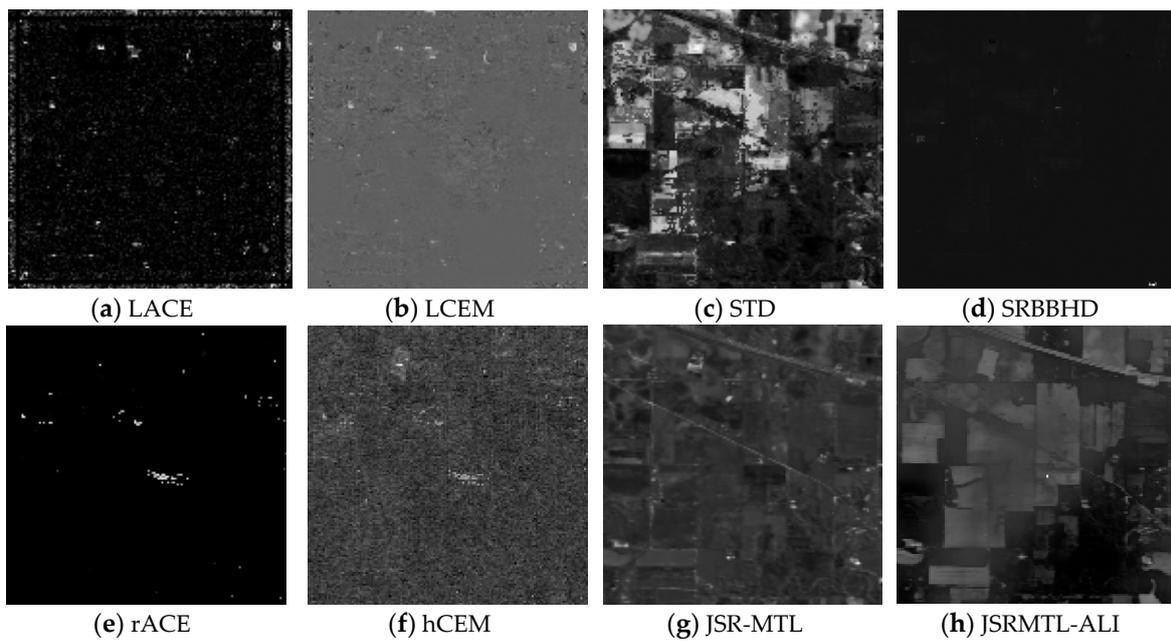


Figure 13. Two-dimensional plots of the detection map for the Indian dataset.

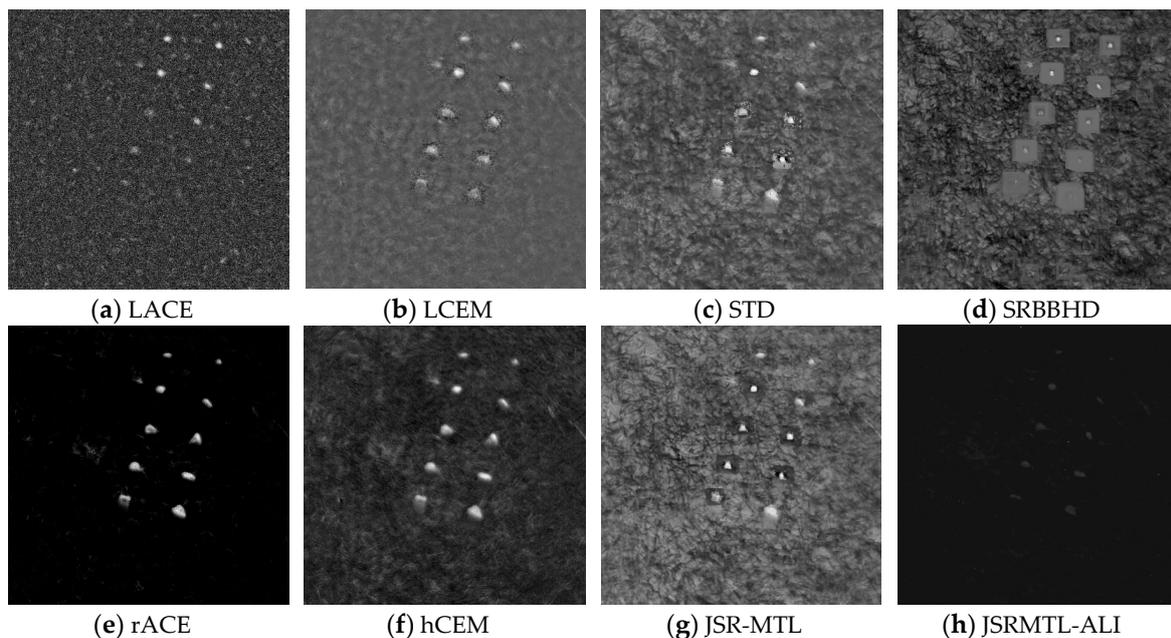


Figure 14. Two-dimensional plots of the detection map for the Cri dataset.

## 5. Discussion

An adaptive joint sparse representation and multi-task learning detector with locality information (JSRMTL-ALI) is proposed in this paper. In order to fully explore the prior class label information of the training samples, JSRMTL-ALI constructs two joint sparse representation and multi-task learning models corresponding to the target and background classes. In order to consider the difference between the target dictionary and background dictionary, JSRMTL-ALI then employs different regularization strategies encoding the task relatedness for the two models, where the  $\ell_{2,1}$ -norm is enforced to the coefficient matrix  $\mathbf{W}^b$  corresponding to the background dictionary, while the  $\ell_1$ -norm is applied for the coefficient matrix  $\mathbf{W}^t$  corresponding to the target dictionary. These two contributions lead to the

so-called JSRMTL-A model. What is more, in order to keep the locality information between the central test pixel and neighboring background training samples, JSRMTL-ALI employs the locality information descriptor-based weight to the joint sparse representation and multi-task learning model corresponding to the background class, which can also reduce the estimation bias caused by the  $\ell_1/\ell_2$  minimization.

From the above experimental results, it can be seen that Figure 6 shows the superiority of the JSRMTL-A model to the traditional JSR-MTL. The JSRMTL-ALI model can generally further improve the detection performance of the JSRMTL-A model. In the detection performance analysis section, as shown in Figure 10, the detection performance of the JSRMTL-ALI generally outperforms the other detectors for all the datasets, especially the robustness of the JSRMTL-ALI compared to the rACE and hCEM algorithms. From the separability maps as shown in Figure 11 and detection map as shown in Figure 12, it can be seen that, the proposed JSRMTL-ALI algorithm generally performs well at a distinguishing target from the background. However, the performance of suppressing background for the JSRMTL-ALI algorithm is not as good as rACE, which needs further consideration in the future.

There are three key parameters of the JSRMTL-ALI algorithm, which have been analyzed as depicted in Figures 7–9. As shown in Figure 7, it can be seen that a large detection task number  $K$  can affect the detection performance of JSRMTL-ALI. A larger value for detection task number  $K$  is recommended for the dataset with more bands, such as 6 for the AVIRIS and Indian datasets and a lower value is recommended for the dataset with fewer bands, such as 3 for the Cri dataset. Then as shown in Figure 8, it can be seen that a too small or too large regularization parameter  $\rho$  can decrease the detection performance of JSRMTL-ALI, and a proper value should be set for  $\rho$ , such as 0.1. Based on the results as shown in Figure 9, it can be seen that, it is not easy to recommend a regular value for the size of OWR in practical application. Our future research will investigate the construction of a global background dictionary in order to avoid tuning the size of the OWR.

## 6. Conclusions

In this paper, the adaptive joint sparse representation and multi-task learning detector with locality information (JSRMTL-ALI) algorithm was proposed. Based on the prior class label information of the training samples, this algorithm constructs an adaptive joint sparse representation and multi-task learning (JSRMTL-A) model, where the test pixel (target pixel or background pixel) is separately modeled via the target dictionary or background dictionary. Considering the great difference between the target dictionary and background dictionary, different regularization strategies encoding the task relatedness are employed for the two joint sparse representation and multi-task learning models based on the target dictionary or background dictionary. A locality information descriptor is then introduced to indicate the difference between the central test pixel and neighboring background training samples. A descriptor based weight strategy is applied to reduce the estimation bias caused by  $\ell_1/\ell_2$  minimization used in the JSRMTL-A model. The detection decision rules in favor of the target class or the background class with the lowest total reconstruction error difference accumulated over all the tasks.

Experiments in hyperspectral target detection with three datasets confirmed the superior performance of the multiple detection task combination in the proposed JSRMTL-ALI algorithm. With the integration of the JSRMTL-A model and local information descriptor based weight strategy, the JSRMTL-ALI shows its superiority to the traditional JSR-MTL for hyperspectral target detection. In general, the JSR-MTL presents a better detection performance and better separability than the other common detectors.

**Acknowledgments:** The authors would like to thank the handling editor and anonymous reviewers for their careful reading and helpful remarks. This work was supported in part by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant CUG170617, Grant CUGL140410 and Grant 26420160125; in part by the China Postdoctoral Science Found under Grant 2017M612533; in part by the Natural Science Foundation of Hubei Province, China under Grant 2014CFA052; in part by the National Natural Science Foundation of China under Grant 61471274, Grant 61372153 and Grant 41630317.

**Author Contributions:** All the authors made significant contributions to the work. Yuxiang Zhang and Bo Du conceived, designed and performed the experiments; Ke Wu and Xiangyun Hu analyzed the data; Liangpei Zhang provided advice for the preparation and revision of the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Manolakis, D.; Truslow, E.; Pieper, M.; Cooley, T.; Brueggeman, M. Detection Algorithms in Hyperspectral Imaging Systems: An Overview of Practical Algorithms. *IEEE Signal Process. Mag.* **2014**, *31*, 24–33. [[CrossRef](#)]
2. Nasrabadi, N.M. Hyperspectral Target Detection: An Overview of Current and Future Challenges. *IEEE Signal Process. Mag.* **2014**, *31*, 34–44. [[CrossRef](#)]
3. Kang, X.; Li, S.; Benediktsson, J.A. Feature Extraction of Hyperspectral Images with Image Fusion and Recursive Filtering. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3742–3752. [[CrossRef](#)]
4. Landgrebe, D. Hyperspectral Image Data Analysis. *IEEE Signal Process. Mag.* **2002**, *19*, 17–28. [[CrossRef](#)]
5. Yuan, Y.; Ma, D.; Wang, Q. Hyperspectral Anomaly Detection by Graph Pixel Selection. *IEEE Trans. Cybern.* **2016**, *46*, 3123–3134. [[CrossRef](#)] [[PubMed](#)]
6. Stefanou, M.S.; Kerekes, J.P. Image-derived prediction of spectral image utility for target detection applications. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1827–1833. [[CrossRef](#)]
7. Datt, B.; McVicar, T.R.; Niel, T.G.V.; Jupp, D.L.B.; Pearlman, J.S. Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1246–1259. [[CrossRef](#)]
8. Eismann, M.T.; Stocker, A.D.; Nasrabadi, N.M. Automated hyperspectral cueing for civilian search and rescue. *Proc. IEEE* **2009**, *97*, 1031–1055. [[CrossRef](#)]
9. Manolakis, D.; Marden, D.; Shaw, G.A. Hyperspectral Image Processing for Automatic Target Detection Applications. *Linc. Lab. J.* **2003**, *14*, 79–116.
10. Wang, T.; Du, B.; Zhang, L. An automatic robust iteratively reweighted unstructured detector for hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2367–2382. [[CrossRef](#)]
11. Gao, L.; Yang, B.; Du, Q.; Zhang, B. Adjusted Spectral Matched Filter for Target Detection in Hyperspectral Imagery. *Remote Sens.* **2015**, *7*, 6611–6634. [[CrossRef](#)]
12. Liu, Y.; Gao, G.; Gu, Y. Tensor Matched Subspace Detector for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2017**, *54*, 1967–1974. [[CrossRef](#)]
13. Geng, X.; Ji, L.; Sun, K.; Zhao, Y. CEM: More Bands, Better Performance. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1876–1880. [[CrossRef](#)]
14. Zou, Z.; Shi, Z. Hierarchical Suppression Method for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 330–342. [[CrossRef](#)]
15. Harsanyi, J.C.; Chang, C.I. Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 779–785. [[CrossRef](#)]
16. Huang, Z.; Shi, Z.; Yang, S. Nonlocal Similarity Regularized Sparsity Model for Hyperspectral Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1532–1536. [[CrossRef](#)]
17. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Sparse representation for target detection in hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2011**, *5*, 629–640. [[CrossRef](#)]
18. Zhang, Y.; Du, B.; Zhang, L. A Sparse Representation Based Binary Hypothesis Model for Target Detection in Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1346–1354. [[CrossRef](#)]
19. Niu, Y.; Wang, B. Extracting Target Spectrum for Hyperspectral Target Detection: An Adaptive Weighted Learning Method Using a Self-Completed Background Dictionary. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1604–1617. [[CrossRef](#)]
20. Farrell, M.D.; Mersereau, R.M. On the impact of PCA dimension reduction for hyperspectral detection of difficult targets. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 192–195. [[CrossRef](#)]
21. Fowler, J.E.; Du, Q. Anomaly Detection and Reconstruction from Random Projections. *IEEE Trans. Image Process.* **2012**, *21*, 184–195. [[CrossRef](#)] [[PubMed](#)]
22. Wang, Q.; Lin, J.; Yuan, Y. Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 1279–1289. [[CrossRef](#)] [[PubMed](#)]

23. Binol, H.; Ochilov, S.; Alam, M.S.; Bal, A. Target oriented dimensionality reduction of hyperspectral data by Kernel Fukunaga—Koontz Transform. *Opt. Laser Eng.* **2016**, *89*, 123–130. [[CrossRef](#)]
24. Sun, K.; Geng, X.; Ji, L. A New Sparsity-Based Band Selection Method for Target Detection of Hyperspectral Image. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 329–333. [[CrossRef](#)]
25. Jalali, A.; Ravikumar, P.; Sanghavi, S.; Ruan, C. A Dirty Model for Multi-task Learning. In Proceedings of the Neural Information Processing Systems Conference, Hyatt Regency, Vancouver, BC, Canada, 6–11 December 2010.
26. Yuan, X.; Liu, X.; Yan, S. Visual classification with multi-task joint sparse representation. *IEEE Trans. Image Process.* **2012**, *21*, 4349–4360. [[CrossRef](#)] [[PubMed](#)]
27. Yuan, Y.; Lin, J.; Wang, Q. Hyperspectral Image Classification via Multi-Task Joint Sparse Representation and Stepwise MRF Optimization. *IEEE Trans. Cybern.* **2016**, *46*, 2966–2977. [[CrossRef](#)] [[PubMed](#)]
28. Caruana, R. Multitask learning. *Mach. Learn.* **1997**, *28*, 41–75. [[CrossRef](#)]
29. Zhang, Y.; Du, B.; Zhang, L. Joint Sparse Representation with Multitask Learning for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 894–906. [[CrossRef](#)]
30. Zou, H. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [[CrossRef](#)]
31. Candes, E.J.; Wakin, M.B.; Boyd, S.P. Enhancing sparsity by reweighted  $\ell$  minimization. *J. Fourier Anal. Appl.* **2008**, *14*, 877–905. [[CrossRef](#)]
32. Ben-David, S.; Schuller, R. Exploiting task relatedness for multiple task learning. In Proceedings of the Conference on Computational Learning Theory, Washington, DC, USA, 24–27 August 2003.
33. Chen, X.; Pan, W.; Kwok, J.; Garbonell, J. Accelerated gradient method for multi-task sparse learning problem. In Proceedings of the IEEE International Conference on Data Mining, Miami, FL, USA, 6–9 December 2009; pp. 746–751.
34. Tseng, P. On accelerated proximal gradient methods for convex-concave optimization. *SIAM J. Optim.* **2008**, submitted.
35. Kang, X.; Li, S.; Fang, L.; Benediktsson, J.A. Intrinsic Image Decomposition for Feature Extraction of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2241–2253. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).