A Randomized Subspace Learning Based Anomaly Detector for Hyperspectral Imagery

Weiwei Sun [1,2,*]**, Long Tian** [3]**, Yan Xu** [3]**, Bo Du** [4] **and Qian Du** [3]

[1] Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China
[2] State Key Laboratory of Information Engineering on Survey, Mapping and Remote Sensing,
    Wuhan University, Wuhan 430079, China
[3] Department of Electrical and Computer Engineering, Mississippi State University, Starkville,
    MS 39762, USA; lt766@msstate.edu (L.T.); yx131@msstate.edu (Y.X.); du@ece.msstate.edu (Q.D.)
[4] School of Computer Science, Wuhan University, Wuhan 430079, China; remoteking@whu.edu.cn
* Correspondence: sunweiwei@nbu.edu.cn

Abstract: This paper proposes a randomized subspace learning based anomaly detector (RSLAD) for hyperspectral imagery (HSI). Improved from robust principal component analysis, the RSLAD assumes that the background matrix is low-rank, and the anomaly matrix is sparse with a small portion of nonzero columns (i.e., column-wise). It also assumes the anomalies do not lie in the column subspace of the background and aims to find a randomized subspace of the background to detect the anomalies. First, random techniques including random sampling and random Hadamard projections are implemented to construct a coarse randomized columns subspace of the background with reduced computational cost. Second, anomaly columns are searched and removed from the coarse randomized column subspace by solving a series of least squares problems, resulting in a purified randomized column subspace. Third, the nonzero columns in the anomaly matrix are located by projecting all the pixels on the orthogonal subspace of the purified subspace, and the anomalies are finally detected based on the L_2 norm of the columns in the anomaly matrix. The detection performance of RSLAD is compared with four state-of-the-art methods, including global Reed-Xiaoli (GRX), local RX (LRX), collaborative-representation based detector (CRD), and low-rank and sparse matrix decomposition base anomaly detector (LRaSMD). Experimental results show good detection performance of RSLAD with lower computational cost. Therefore, the proposed RSLAD offers an alternative option for hyperspectral anomaly detection.

Keywords: low-rank recovery; randomized subspace learning; anomaly detection; hyperspectral imagery

1. Introduction

Hyperspectral imaging collects detailed spectral information of ground objects on the earth surface using hundreds of narrow and continuous bands [1,2]. It has distinctive advantages in detecting small and low-probability ground objects using the techniques of target detection [3–6]. In particular, with the forthcoming generation of hyperspectral sensors (e.g., EnMAP, HISUI, and Hispery), there is a tremendous need to develop intelligent methods and protocols for target detection to fully benefit from a wider range of spectral bands. Hyperspectral target detection can be applied in many realistic applications, including biophysical parameter retrieval [7], classification of complicated environments [8] and military target detection [9]. Compared with supervised target detection, unsupervised target detection, i.e., anomaly detection, does not require any prior knowledge of target spectral characteristic [10–12]. Anomaly detection methods have witnessed increasing interest due to their interesting applications [13–16]. In anomaly detection, common ground objects that dominate the image scene are defined as the background, whereas small and low-probability ground objects that

spectrally differ from the background are considered as anomalies [3–6]. We limit our discussion on anomaly detection in this paper.

The classical statistical method of anomaly detection is the Reed-Xiaoli (RX) detector [15,16]. It assumes that the background follows a multivariate normal distribution and uses the probability density of a multivariate normal distribution to estimate the chance of a pixel under test to be part of the background. The RX has two versions: global RX (GRX) and local RX (LRX). In the GRX mode, the background is estimated from the full image scene, whereas the background in the LRX model is estimated from the local neighborhood around the testing pixel [12]. The Gaussian distribution assumption in the RX may contradict the reality of the HSI data and fails to capture the complexity of the background [14,17]. Other advanced RX based methods were then proposed to model the nonstationarity of the background clutter, e.g., using a mixture of multivariate Gaussian distributions in global Gaussian mixture model methods (GGMM) [18], using the clustering techniques in the cluster-based detector (CBD) [19] and using the nonlinear kernel schemes in the kernel RX (KRX) [20]. On the other hand, since the RX may include anomalous pixels when estimating background statistics, more robust background estimation approaches were proposed for more accurate background modeling. Typical examples are the subspace-based approaches [21,22], the multivariate outlier detection methods [23,24] and the random selection based method [25]. The subspace-based anomaly detectors model the background as a subspace and find the anomalies using the projection separation statistics [3,21,22]. The projection basis of the subspace is usually generated by eigenvalue decomposition of the background covariance matrix or by background endmember extraction using spectral unmixing [3]. The multivariate outlier detection methods consider the anomaly detection problem as outlier detection in a projected subspace of hyperspectral data and suppress anomalous signatures during background modeling. Representative detectors are blocked adaptive computationally efficient outlier nominator detector (BACON) [23] and the minimum covariance determinant detector [24]. The random selection based anomaly detector (RSAD) selects adequate and representative background pixels to compute a robust background statistic [25].

The aforementioned statistical methods belong to parametric anomaly detectors and require the selection, validation and estimation of probability density function of background pixels [26]. An inadequate estimation of the underlying distribution leads to poor detection performance [3,6]. Therefore, nonparametric anomaly detectors may be preferred to avoid prior assumptions about background distribution, e.g., support vector approaches [26,27] and manifold learning approaches [28,29]. Support vector approaches are based on one-class classifier and model the background by a minimum enclosing hypersphere to identify anomalies lying outside of the background support regression. Typical detectors include local support vector data description (LSVDD) detector [26] and sparse kernel-based anomaly detector [27]. A common issue of support vector approaches is how to select a proper kernel function. Manifold learning approaches assume the nearest neighbor structures of all pixels as the local manifolds. The anomalies are selected from points that lie far away from main manifolds in the embeddings, using the detectors of robust locally linear embedding [28] or the local tangent space alignment [29]. However, the performance of these algorithms highly depends on the neighborhood size and the embedding dimensions. More recent nonparametric anomaly detectors include discriminative metric learning based anomaly detector [14], transferred deep convolutional neural network based anomaly detector [30], and tensor composition based anomaly detector [31].

Recently, with the popularity of sparsity theory in matrix factorization, the sparse anomaly detectors attract more attentions in hyperspectral field. Current sparsity based anomaly detectors can be roughly categorized into two main aspects, i.e., sparse representation (SR) based and robust principal component analysis (RPCA) based. The SR based anomaly detectors assume the background pixel can be sparsely represented by a defined dictionary and use sparse representation to separate an anomaly from the background [17]. The RPCA based anomaly detectors assume that the background is low-rank and the anomalies are sparse in the image scene. The collaborative-representation based

detector (CRD) assumes that each background pixel can be approximately represented by its spatial neighborhoods. The anomalies are estimated by subtracting the predicted background from the original HSI data [32]. The low-rank and sparse representation (LRASR) based detector regards that a background pixel can be approximately represented by a background dictionary and the anomalies are estimated from the residual sparse image [33]. The low-rank representation and learned dictionary (LRRaLD) improves decomposition process of the regular low-rank representation model using the random selected dictionary and could obtain more robust detection results within less computation time [34]. The sparsity score estimation framework (SSEF) detector counts the frequency of each dictionary atom for hyperspectral data construction in sparse representation, and it estimates the anomalies using the sparsity score matrix of all pixels [35]. The low-rank and sparse decomposition (LSD) formulates the detection of anomalies as a RPCA problem in the local image region and finds the anomalies by soring each pixel by the norm of its corresponding sparse coefficient vector [36]. The low-rank and sparse matrix decomposition based anomaly detection (LRaSMD) [12] improved the RPCA model by separating the noise term from the anomaly term in the sparse noise matrix. The pixels that have small L_2 norm values of the sparse coefficient vectors are selected as the anomalies. Later, the LRaSMD is further improved by replacing the L_2 norm with the Mahalanobis distance [37].

In this paper, inspired by RPCA [38–40], we propose a randomized subspace learning based anomaly detector (RSLAD) for hyperspectral anomaly detection. The RSLAD assumes that the background is low-rank, and the anomaly matrix is sparse and has a small portion of nonzero columns (i.e., column-wise). Meanwhile, the RSLAD assumes that the background pixels lie in the column subspace of the background whereas the anomalies do not. It aims to find a randomized subspace of the background where anomalies are more likely to be excluded. Random techniques are utilized to find a coarse randomized subspace of the background. Random sampling and the random Hadamard projections could separately sketch the original data from columns and rows and greatly reduce the computational requirements of subspace learning. The anomaly columns are excluded from the coarse randomized subspace by solving a series of least squares problems, resulting in purified randomized subspace of the background. The anomalies are then located by projecting the data onto the orthogonal subspace projection of the purified column subspace.

Compared with current sparsity based anomaly detectors, our method favors three main contributions:

(1) The RSLAD has more advanced assumptions than RPCA. The RSLAD assumes that the anomaly matrix has a small portion of nonzero columns and these nonzero columns do not lie in the column subspace of the background. This assumption reduces the impact from anomalies when constructing the column subspace of the background. In contrast, the RPCA based anomaly detectors assume that nonzero elements in the sparse anomaly matrix are uniformly scattered without any specific structure. Accordingly, nonzero entries in the anomaly matrix would negatively affect all the columns of the background when optimizing the convex programs.

(2) The idea behind RSLAD is more advanced than current sparsity based anomaly detectors. It is to find a randomized subspace of the background and investigates the low-dimensionality of the background column subspace and the independence between anomalies. It estimates the randomized column subspace of the background and alleviates the effects from anomalies in estimating the low-rank background.

(3) The RSLAD does not actually solve a complicated convex optimization problem, and it offers good performance with a low computational cost due to the use of random selection and projection. The low computational complexity makes it more appealing in practical applications.

The forthcoming sections of our paper are arranged as follows. Section 2 describes the modeling of background and anomalies in RSLAD. Section 3 presents the methodology of RSLAD. Section 4 compares with state-of-the-art methods and analyzes the detection performance of RSLAD using four real hyperspectral images. Section 5 discusses the experimental results. Section 6 draws conclusions of our paper.

2. Modeling of Background and Anomalies in RSLAD

Consider a hyperspectral data as a collection of band vectors $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N} \in \mathbb{R}^{M \times N}$, where M is the number of bands and N is the number of pixels. Let the background matrix and the anomaly matrix be denoted as $\mathbf{B} \in \mathbf{R}^{M \times N}$ and $\mathbf{S} \in \mathbf{R}^{M \times N}$ respectively. The anomaly detection is to separate the anomalies from the background, and accordingly, the HSI matrix $\mathbf{Y}$ can be expressed as the sum of background matrix $\mathbf{B}$ and anomaly matrix $\mathbf{S}$.

The background matrix $\mathbf{B}$ consists of spectral vectors of main ground objects in the image scene and is assumed to lie on a low-dimensional subspace with low-rank properties. The anomaly matrix $\mathbf{S}$ collects spectral vectors of small and low-probability ground objects (i.e., anomalies), and hence it is column-wise sparse with a small portion of nonzero columns $\mathcal{C}$. Obviously, the corresponding columns $\mathcal{C}$ in $\mathbf{B}$ is zero [38,41]. The matrices $\mathbf{Y}$, $\mathbf{B}$, and $\mathbf{S}$ are related as

$$\mathbf{Y} = \mathbf{B} + \mathbf{S}, \; s.t., \begin{cases} \text{Rank}(\mathbf{B}) = r, \; \mathbf{b}_i = \mathbf{0}, \; \forall i \in \mathcal{C} \\ \left(\mathbf{I} - \mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\right)\mathbf{s}_i \neq \mathbf{0}, \forall i \in \mathcal{C} \end{cases} \tag{1}$$

where $\mathbf{b}_i$ and $\mathbf{s}_i$ are the i-th column of $\mathbf{B}$ and $\mathbf{S}$, respectively, r is the dimensonality of subspace of $\mathbf{B}$, and $\mathbf{U}$ is the basis of column subspace of $\mathbf{B}$. The constraint Rank($\mathbf{B}$) = r is to guarantee the low-rank property of $\mathbf{B}$. The constraint $\left(\mathbf{I} - \mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\right)\mathbf{s}_i \neq \mathbf{0}, \forall i \in \mathcal{C}$ is to restrict that the anomalies do not lie in the column subspace of $\mathbf{B}$.

Figure 1 illustrates the difference between the model in Equation (1) and the original RPCA [38]. The RPCA assumes that the matrix $\mathbf{S}$ is sparse with nonzero entries being scattered uniformly at random. The nonzero entries can have arbitrarily large magnitude. Consequently, all the columns of $\mathbf{B}$ can be affected by the nonzero elements of $\mathbf{S}$ [41,42]. On the contrary, Equation (1) assumes that the anomaly matrix $\mathbf{S}$ is column-wise sparse, where only a portion of its columns are nonzero. A small portion of the columns of the anomaly matrix $\mathbf{S}$ are nonzero, and these nonzero columns do not lie in the column space of $\mathbf{B}$. Therefore, a portion of the columns that formulate the space of $\mathbf{B}$ are unaffected by the nonzero columns of $\mathbf{S}$ [42,43].
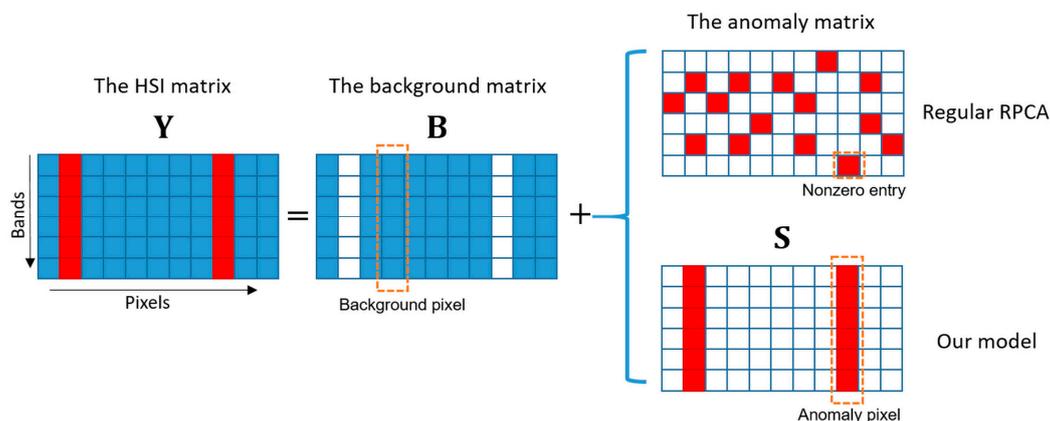

Figure 1. The modeling of background and anomalies. The red columns are spectral vectors of anomalies, where each red square denotes the spectral reflectance or digital number of an anomalous pixel in a band; the blues are spectral vectors of background pixels; and the whites denote zero columns in the matrix.

3. The Proposed RSLAD for HSI Anomaly Detection

The background matrix $\mathbf{B}$ and the anomaly matrix $\mathbf{S}$ are unknown, and it is difficult to find a closed form solution of Equation (1). It is often transformed into a convex optimization problem [41] with an objection function combining the nuclear norm minimization of $\mathbf{B}$ and the $L_{1,2}$ norm minimization

of **S**. The nuclear norm and $L_{1,2}$ norm could ensure the low-rank of **B** and the column space of **S**, respectively. However, the approach could yield a robust estimate of **B** only when the fractions of anomalies in the image scene are less than a constant threshold [43].

In Equation (1), once the column subspace **U** of the background is found, the anomalies can be also correctly located with the orthogonal subspace projection constraint $(\mathbf{I} - \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T)\mathbf{y}_i \neq 0$, where $\mathbf{y}_i$ is a column (corresponds to one pixel) of the data matrix **Y**. The explanation is that the background pixels lie in the column subspace of **B**, whereas the anomaly pixels lie out of the column subspace of **B**. Accordingly, the idea of our RSLAD is to find a column subspace of the background where anomalies being excluded.

3.1. Constructing Coarse Randomized Subspace by Data Sketching

In Equation (1), the column subspace of the background matrix **B** is a low-dimensional subspace, and it can be spanned by a small subset of its columns. Assuming $\mathbf{Y}_\Psi$ is the matrix of p randomly sampled columns of the background matrix **B**, and $\mathbf{V}_\Psi$ is an orthonormal basis for the row space of **B**, random sampling states that, if the sampled number satisfies $p \geq 10\mu n \log \frac{2n}{\delta}$, then $\mathbf{Y}_\Psi$ and **Y** have the same column subspace with probability at least $(1 - \delta)$, and the $\mathbf{Y}_\Psi$ is incoherent with the basis of row subspace $\mathbf{V}_\Psi$ by satisfying $\max_i ||\mathbf{e}_i^T \mathbf{V}_\Psi||_2^2 \leq \frac{6\mu n}{p}$ with probability at least $(1 - \delta)$ [43]. n is the rank of background **B**, $\mathbf{e}_i$ is an identity vector with all entries equal to 1, and μ is the incoherence parameter of row subspace of **B**. Therefore, the column subspace $\mathbf{Y}_\Psi$ of the background matrix **B** can be learned from a random subset of its columns when its row space is incoherent with the orthonormal basis.

Random sampling can be used to extract the low-dimensional subspace structure of the background. It also sketches the data matrix **Y** from columns and reduce its computational requirements. Suppose the p pixels (i.e., columns) are randomly sampled, the random sampling can be defined as

$$\mathbf{Y}_\Psi = \mathbf{Y}\Psi \tag{2}$$

where $\mathbf{Y}_\Psi \in \mathbf{R}^{M \times p}$ is the sub-matrix of **Y** with p selected columns, and Ψ is the random sampling matrix whose columns are randomly selected with replacement from the $N \times N$ identity matrix $\mathbf{I}_N$.

On the other hand, numerous bands in **Y** bring high computational complexity and memory. Random projections originate from the famous Johnson-Lindenstrauss lemma and have been proven to be a computationally efficient and sufficiently accurate method for reducing the dimensionality of the HSI data. Compared with the regular Gaussian random projections, the random Hadamard projections have lower computational costs and better performance [44,45]. Therefore, we adopt the random Hadamard projections to sketch the matrix $\mathbf{Y}_\Psi$ from rows and reduce the spectral dimensionality of $\mathbf{Y}_\Psi$. The random Hadamard matrix based dimensionality reduction is defined as

$$\mathbf{Y}_\Phi = \Phi^T \mathbf{Y}_\Psi = \left(\sqrt{\frac{K}{M}} \mathbf{DH}_N \mathbf{P}\right)^T \mathbf{Y}_\Psi \tag{3}$$

where $\mathbf{Y}_\Phi \in \mathbf{R}^{K \times p}$ is the projected matrix and $\Phi = \sqrt{\frac{K}{M}} \mathbf{DH}_M \mathbf{P} \in \mathbf{R}^{M \times K}$ is the random Hadamard projection matrix. $\mathbf{D} \in \mathbf{R}^{M \times M}$ is a diagonal matrix with diagonal entries sampled uniformly from $\{-1, 1\}$. $\mathbf{H}_N \in \mathbf{R}^{M \times M}$ is the Hadamard matrix defined recursively for any M that is an integer power of 2 as $\mathbf{H}_N = \begin{bmatrix} \mathbf{H}_{2/M} & \mathbf{H}_{2/M} \\ \mathbf{H}_{2/M} & -\mathbf{H}_{2/M} \end{bmatrix}$, where $\mathbf{H}_1 = 1$. $\mathbf{P} \in \mathbf{R}^{M \times K}$ is a uniform sampling matrix that randomly samples K columns of $\mathbf{DH}_N$, where each column of **P** is randomly selected with replacement from the $M \times M$ identity matrix $\mathbf{I}_M$.

3.2. Purifying Randomized Column Subspace of Background

Due to the presence of anomalous pixels, $\mathbf{Y}_\Phi$ may include both background and anomalous columns. However, the number of background columns are much larger than that of anomaly columns

with higher probability. Therefore, the anomaly columns should be removed from $\mathbf{Y_\Phi}$ to purify the column subspace of background.

The columns of $\mathbf{Y_\Phi}$ from random sampling and random projections can span the columns of background with high probability. Accordingly, any background column of $\mathbf{Y_\Phi}$ lies in the span of other background columns of $\mathbf{Y_\Phi}$ with high probability. That is, for each background column $\mathbf{y_\Phi^i}$, let the matrix $\mathbf{Y_{\Phi(-i)}}$ be $\mathbf{Y_\Phi}$ but with the *i*-th column being removed, it can be expressed as a linear combination of columns of $\mathbf{Y_{\Phi(-i)}}$. In contrast, the anomaly column does not lie in the span of background columns. Since real hyperspectral data are always contaminated by noises, the anomaly columns from $\mathbf{Y_\Phi}$ can be located by solving the following least squares problem:

$$\mathbf{RSE}_i = \underset{\mathbf{z}_i}{\operatorname{argmin}} ||\mathbf{y_\Phi^i} - \mathbf{Y_{\Phi(-i)}z_i}||_2, s.t., \begin{cases} \mathbf{RSE}_i > \varepsilon, & i \text{ belong to anomaly columns} \\ \mathbf{RSE}_i \leq \varepsilon, & i \text{ belong to background columns} \end{cases} \tag{4}$$

where $\varepsilon > 0$ is the defined residual threshold because of noise. After locating the anomaly columns in $\mathbf{Y_\Phi}$, the purified matrix $\mathbf{U}$ can be obtained.

3.3. Detecting Anomalies Using Orthogonal Subspace Projection

The purified matrix $\mathbf{U}$ of $\mathbf{Y_\Psi}$ contains linearly independent background columns and can be considered as a basis of the background matrix $\mathbf{B}$. Since the anomalies do not lie in the column subspace of the background $\mathbf{B}$, the projection of the HSI pixels on the orthogonal subspace of $\mathbf{U}$ can be used to locate nonzero columns of $\mathbf{S}$ [43]. The anomaly matrix $\mathbf{S}$ is estimated from Equation (5)

$$\mathbf{S} = (\mathbf{I} - \mathbf{U}(\mathbf{U^T U})^{-1}\mathbf{U^T})\mathbf{Y} \tag{5}$$

Due to the impact of noise, the anomaly matrix $\mathbf{S}$ does not necessarily have a small portion of nonzero columns. Therefore, the L_2 norm is adopted to calculate the anomalous value for each pixel, and the pixels with the anomalous value above a manually selected threshold are determined to be anomalies.

3.4. The Summary of RSLAD for HSI Anomaly Detection

The RSLAD assumes that the background is low-rank, the anomaly matrix is sparse and column-wise, and the anomalies do not lie in the column subspace of the background $\mathbf{B}$. It seeks the randomized subspace of the background $\mathbf{B}$ and detects the anomalies by projecting the HSI dataset $\mathbf{Y}$ on the complement subspace of the randomized subspace. Algorithm 1 lists the detailed procedure of RSLAD for anomaly detection. The RSLAD uses random sampling and the random Hadamard projections to sketch the HSI data from columns and rows, respectively. That greatly reduces the computational complexity and memory requirements of the HSI data in subspace learning. More importantly, random sampling constructs a coarse low-dimensional randomized subspace $\mathbf{Y_\Phi}$ of the background. After that, the RSLAD removes the anomalies from the coarse randomized subspace $\mathbf{Y_\Phi}$ by solving a series of least square problems, and the purified randomized subspace is obtained. The least square problem investigates the linear dependence of background pixels and the linear independence between the column subspace of background pixels and anomaly pixels. Finally, the RSLAD detects the anomalies in the HSI data $\mathbf{Y}$ by projecting it on the complement subspace of $\mathbf{U}$. The sketch map of anomaly detection using RSLAD is illustrated in Figure 2.
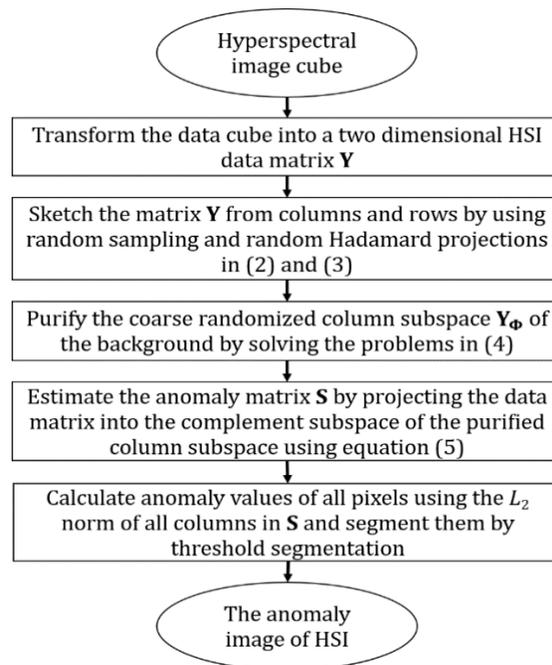
Figure 2. The sketch map of RSLAD for anomaly detection.

Algorithm 1. The Procedure of RSLAD Method.

Input: the HSI band matrix $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N} \in \mathbf{R}^{M \times N}$, the number of sampled pixels p, the projected dimension K, and the residual threshold ε.

(1) Construct a coarse randomized column subspace
    (a)   Obtain the reduced matrix $\mathbf{Y_\Psi}$ by random sampling in (2);
    (b)   Obtain the projected matrix $\mathbf{Y_\Phi}$ by random Hadamard projections in (3);

---

(2) Purify the coarse column subspace $\mathbf{Y_\Phi}$ of the background
For all pixels **do**
    (a)   Define the matrix $\mathbf{Y_{\Phi(-i)}}$ to be $\mathbf{Y_\Phi}$ but with the i-th column removed;
    (b)   Compute the residual error $\mathbf{RSE}_i$ by using the linear combination of $\mathbf{Y_{\Phi(-i)}}$;
    (c)   Locate the anomaly columns in $\mathbf{Y_\Phi}$ using the residual threshold in (4);
End for
    (a)   Obtain the purified matrix $\mathbf{U}$ from $\mathbf{Y_\Psi}$ and set it as the basis of the background $\mathbf{B}$

---

(3) Detecting the anomalies using orthogonal subspace projection
    (a)   Compute the anomaly matrix $\mathbf{S}$ using complement subspace projection of $\mathbf{U}$ in (5);
    (b)   Compute anomalous values of all pixels via the L_2 norm of columns in the anomaly matrix $\mathbf{S}$

Output: Anomaly detection map

---

4. Experimental Results

4.1. The HSI Dataset Descriptions

The first dataset is the Pavia Center (PaviaC) dataset acquired by the reflective optics system imaging spectrometer (ROSIS) sensor [12,37]. It covers the Pavia Center in northern Italy and has accurate ground truth information. The number of bands in the initial dataset is 115 with 1.3 m spatial resolutions covering the spectrum range from 430 to 860 nm. In the experiment, the digital numbers of a smaller image were used as the input data, containing 108×120 pixels and 102 bands after removing low signal-to-noise ratio (SNR) bands. As shown in Figure 3a, three ground objects constitute the background: bridge, water, and shadow. A total of 47 pixels representing vehicles on the

bridge and the bare soil near the bridge were commonly selected as anomalies. The reason is that they cover a very small number of pixels and are spectrally different from main ground objects. Figure 3b shows the ground objects of the anomalies and Figure 3c plots spectral curves of anomalies and main ground objects.
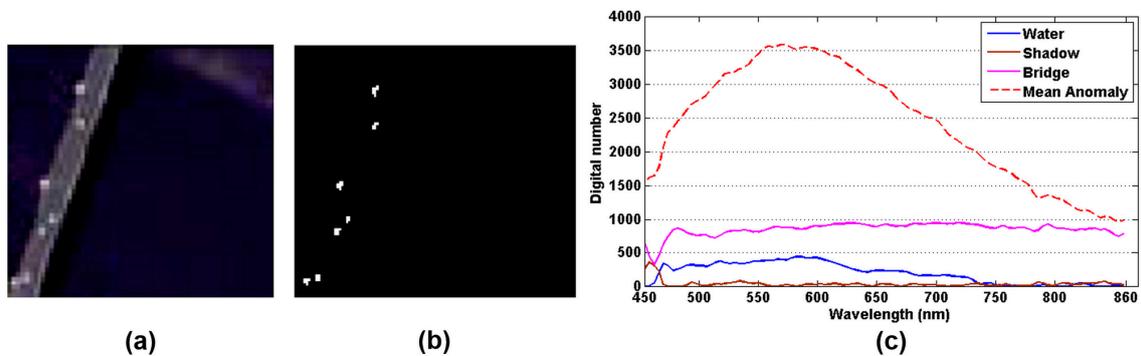

(a)　　　　　　　　　　　(b)　　　　　　　　　　　(c)

Figure 3. (**a**) The image scene of PaviaC dataset by the ROSIS sensor; (**b**) ground truth map of 47 pixels representing anomalies; and (**c**) spectral curves of background and average anomalous signature.

The second dataset includes the San Diego data collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over San Diego, CA, USA [46,47]. The initial images have 3.5 m spatial resolution and 224 spectral channels ranging the spectrum from 370 nm to 2510 nm. In the experiment, a subset image with the image size of 100 × 85 pixels was selected, and the digital numbers were used as the input data. After removing the bad bands [1–6, 33–35, 97, 107–113, 153–166, 221–224] due to water absorption and low signal-to-noise ratio, the 189 bands were used in the experiment. In the image scene shown in Figure 4a, main ground objects of the background are roof, road, shadow and grass. Three planes occupying 58 pixels were commonly regarded as anomalies because they cover a very small number of pixels and are spectrally different from main ground objects. Figure 4b shows the ground objects of the anomalies and Figure 4c plots spectral curves of anomalies and main ground objects.
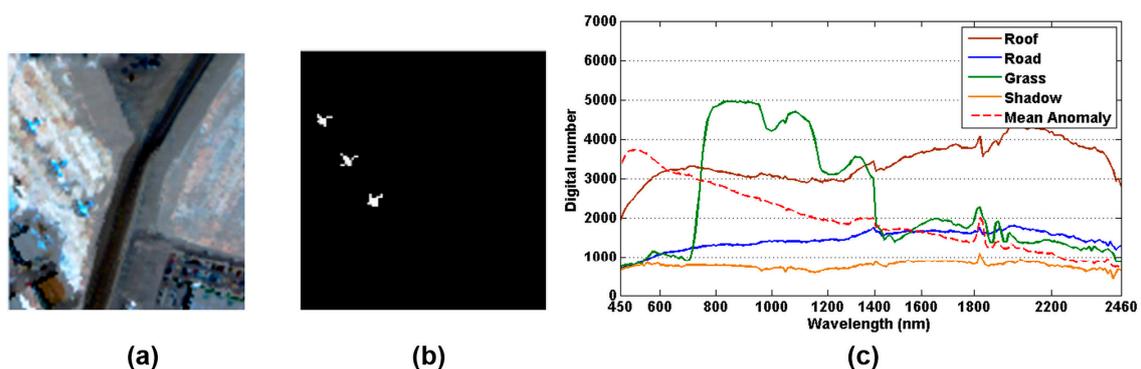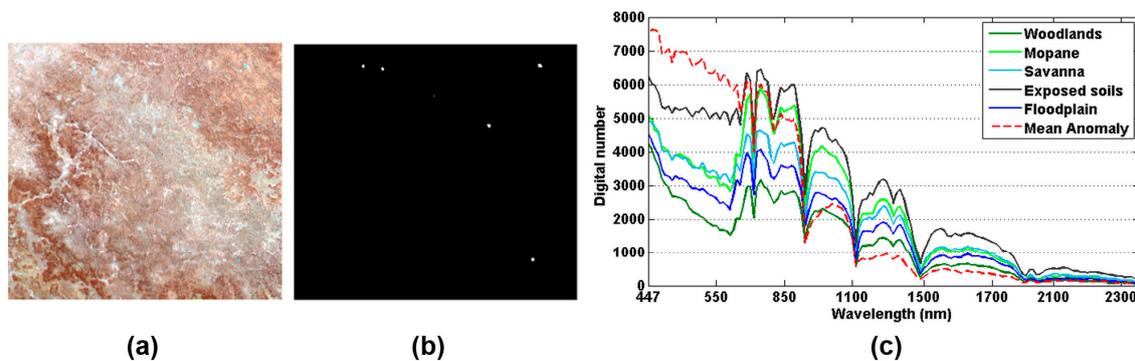

(a)　　　　　　　　　　　(b)　　　　　　　　　　　(c)

Figure 4. (**a**) The image scene of San Diego dataset by the AVIRIS sensor; (**b**) ground truth map of 58 pixels representing anomalies; and (**c**) spectral curves of background and average anomalous signature.

The third dataset includes the Botswana data acquired from Remote Sensing Group of the University of Texas at Austin (www.csr.utexas.edu/hyperspectral/index.html) [12,48]. The dataset was collected by the EO-1 Hyperion sensor. The dataset covers the area of Okavango Delta, Botswana. It was acquired on 31 May 2001 with 30 m spatial resolution and 10 nm spectral resolution, ranging the spectrum between 400 and 2500 nm. In total, 145 bands were used in the experiment: 10–55, 82–97, 102–119, 134–164, and 187–220. A smaller subset of size 235 × 255 pixels was cropped, containing five

classes, namely, woodlands, exposed soil, savanna, floodplain and mopane [44], and the digital numbers were implemented as the input data. These classes reflect the impact of flooding on vegetation in the study area. In the image scene of Figure 5a, 35 pixels were selected as anomalies since these minority pixels are spectrally different from main ground objects. Figure 5b shows the ground objects of the anomalies and Figure 5c plots spectral curves of anomalies and main ground objects.

Figure 5. (**a**) The image scene of Botswana dataset by the EO-1 Hyperion sensor; (**b**) ground truth map of 35 pixels representing anomalies; and (**c**) spectral curves of background and average anomalous signature.

The fourth dataset includes the HyMap data downloaded from the Digital Imaging and Remote Sensing (DIRS) Laboratory of the Rochester Institute of Technique (RIT) (http://dirsapps.cis.rit.edu/blindtest/information/) [49,50]. The dataset was collected around the small town of Cooke City, Montana, USA using the HyMap sensor on 4 July 2006. It has approximately 3 m spatial resolution and 126 spectral bands, ranging the spectrum from 450 to 2500 nm. After removing the bad bands [63–66, 94–95], 120 bands were left, and the digital numbers were used in our experiment. Figure 6a shows the image scene of size 200 × 800 pixels, and its main background types include houses, roads, and trees. Two small fabric targets F1 and F2 occupying 32 pixels were selected as anomalies, since these minority pixels are spectrally different from main ground objects. Figure 6b shows the ground objects of the anomalies, and Figure 6c plots spectral curves of anomalies and main ground objects.
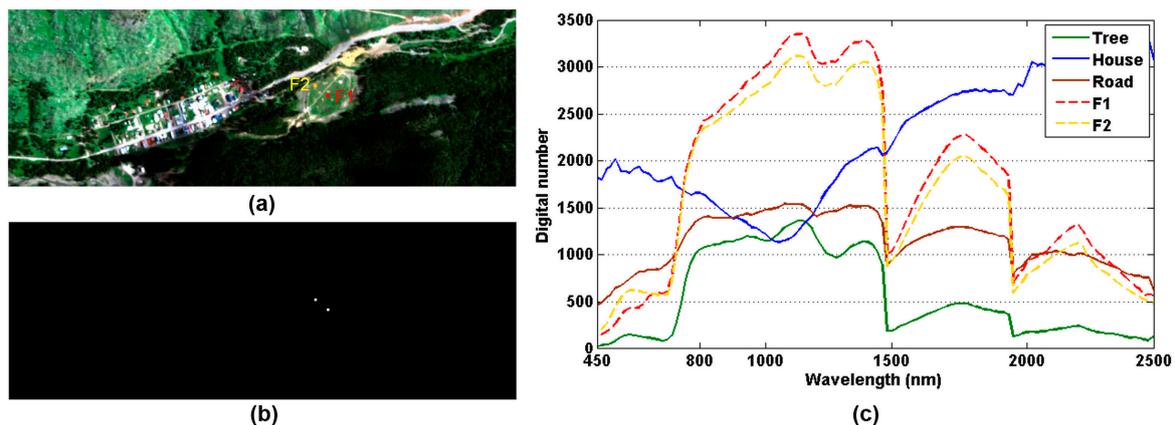

Figure 6. (**a**) The image scene of HyMap dataset; (**b**) ground truth map of F1 and F2 anomalies; and (**c**) spectral curves of main background objects and anomalous signature.

4.2. Experimental Results

In this section, we conduct four groups of experiments on the above hyperspectral datasets to testify the performance of RSLAD in anomaly detection. First, we make comparison between RSLAD and other four state-of-the-art detectors, including GRX [15], LRX [15], CRD [32] and LRaSMD [12].

Second, we investigate the performance sensitivity of RSLAD to the number of randomly sampled columns (i.e., pixels) p. Third, we explore the performance sensitivity of RSLAD to the randomly projected dimension K. Finally, we analyze the performance sensitivity of RSLAD to the residual threshold ε. The receiver operating characteristic (ROC) curve and area under curve (AUC) are utilized to evaluate the detection performance. The ROC depicts both the probability of detection and the probability of false alarm rate. In this paper, we utilize the logarithmic curve to better illustrate the details with a base 10 logarithmic scale for the false alarm rate and a liner scale for the probability of detection. The AUC quantifies the area under the ROC curve and shows how far the ROC curve from the base line (i.e., AUC = 1).

4.2.1. Detection Performance the RSLAD Method

This experiment compares the detection performance of RSLAD with four state-of-the-art methods: two classical detectors GRX and LRX, a representation based detector CRD [32] and a RPCA based detector LRaSMD [12]. Table 1 lists the parameters of all the five detectors on the four datasets. For the LRX and CRD methods, the outer window (OW) and inner window (IW) of PaviaC dataset are set to be 19 × 19 and 7 × 7, respectively; those of San Diego dataset are set to be 21 × 21 and 5 × 5, respectively; those of Botswana dataset are set to be 17 × 17 and 9 × 9, respectively; and those of HyMap dataset are set to be 15 × 15 and 7 × 7, respectively. For the LRaSMD method, the rank of background r and the sparsity level k of anomaly matrix on Pavia dataset are set to be 2 and 0.45, respectively; the r and k on San Diego dataset are set to be 2 and 0.4, respectively; the r and k on Botswana dataset are set to be 4 and 0.5, respectively; and the r and k on the on HyMap dataset are set to be 10 and 0.75, respectively. For the RSLAD method, the number of sampled pixels p, the projected dimension K and the residual threshold ε on the PaviaC dataset are manually set to be 150, 50 and 10^{-9}, respectively; the p, K and ε on the San Diego dataset are manually set to be 120, 50 and 1.7×10^{-10}, respectively; the p, K and ε on the Botswana dataset are manually set to be 100, 50 and 4×10^{-10}, respectively; and the p, K and ε on the HyMap dataset are manually set to be 200, 60 and 1.7×10^{-10}, respectively.

Table 1. The lists of parameters of all detectors on the four datasets.

Datasets	Configuration Parameters				
	GRX	LRX	CRD	LRaSMD	RSLAD
PaviaC	-	OW = 19 × 19; IW = 7 × 7	OW = 19 × 19; IW = 7 × 7	$r = 2$; $k = 0.45$	$p = 150$; $K = 50$; $\varepsilon = 1.0 \times 10^{-9}$
San Diego	-	OW = 21 × 21; IW = 5 × 5	OW = 21 × 21; IW = 5 × 5	$r = 2$; $k = 0.4$	$p = 120$; $K = 50$; $\varepsilon = 1.7 \times 10^{-10}$
Botswana	-	OW = 17 × 17; IW = 9 × 9	OW = 17 × 17; IW = 9 × 9	$r = 4$; $k = 0.5$	$p = 100$; $K = 50$; $\varepsilon = 4 \times 10^{-10}$
HyMap	-	OW = 15 × 15; IW = 7 × 7	OW = 15 × 15; IW = 7 × 7	$r = 10$; $k = 0.75$	$p = 200$; $K = 60$; $\varepsilon = 1.7 \times 10^{-10}$

Figure 7 illustrates the ROC curves and confidence intervals and regions [51] of RSLAD and other four methods on the four datasets. For the PaviaC dataset of Figure 7a, RSLAD has the lowest false alarm rate at 100% probability of detection. The false alarm rate of RSLAD is smaller than those of GRX, LRX and CRD in the ROC curve. The CRD curve behaves better in the false alarm rate than that of LRX but it is inferior to that of GRX. The LRX is the worst among all five detectors. The explanation for the worse performance of LRX than GRX is as follows. The LRX assumes homogenous background within the spatial window. However, for Pavia Center, the anomalies lie close to the edge of the bridge, and the selected window with the anomalies can be composed of different materials including water, bridge, and shadow, which may be falsely detected as the anomaly targets. For the San Diego dataset in Figure 7b, the RSLAD has higher probability of detection than GRX, LRX and CRD. Moreover, the RSLAD performs slightly better than LRaSMD, having a slightly smaller false alarm

rate at 100% probability of detection. The CRD behaves worse than LRaSMD and RSLAD in the false alarm rate but it outperforms the two RX detectors, especially LRX. The LRX has the worst performance of all the five methods. The reason for that is similar to that of PaviaC dataset in Figure 7a, i.e., the background within spatial neighborhood does not satisfy the homogeneous assumption. For the Botswana dataset of Figure 7c, the RSLAD curve is superior to those of GRX, LRX, CRD and LRaSMD, having the smallest false alarm rate at 100% probability of detection. The probability of detection in LRaSMD is higher than those of GRX, LRX and CRD. The CRD outperforms GRX and RX in the false alarm rate at 100% probability of detection, and the GRX is the worst of all the methods. For the HyMap dataset of Figure 7d, the LRX curve has the largest probability of detection with a small false alarm rate less than 0.01, but it could not fully detect all the anomalies in the HyMap imagery. The RSLAD curve has the smallest false alarm rate when obtaining 100% probability of detection. Furthermore, Figure 7 shows the 95% confidence regions drawn around each estimated ROC curves from all the five detectors. The superiority of RSLAD to GRX is statistically significant in all the four datasets, and the ROC curves of RSLAD and LRaSMD show less statistical difference from each other.
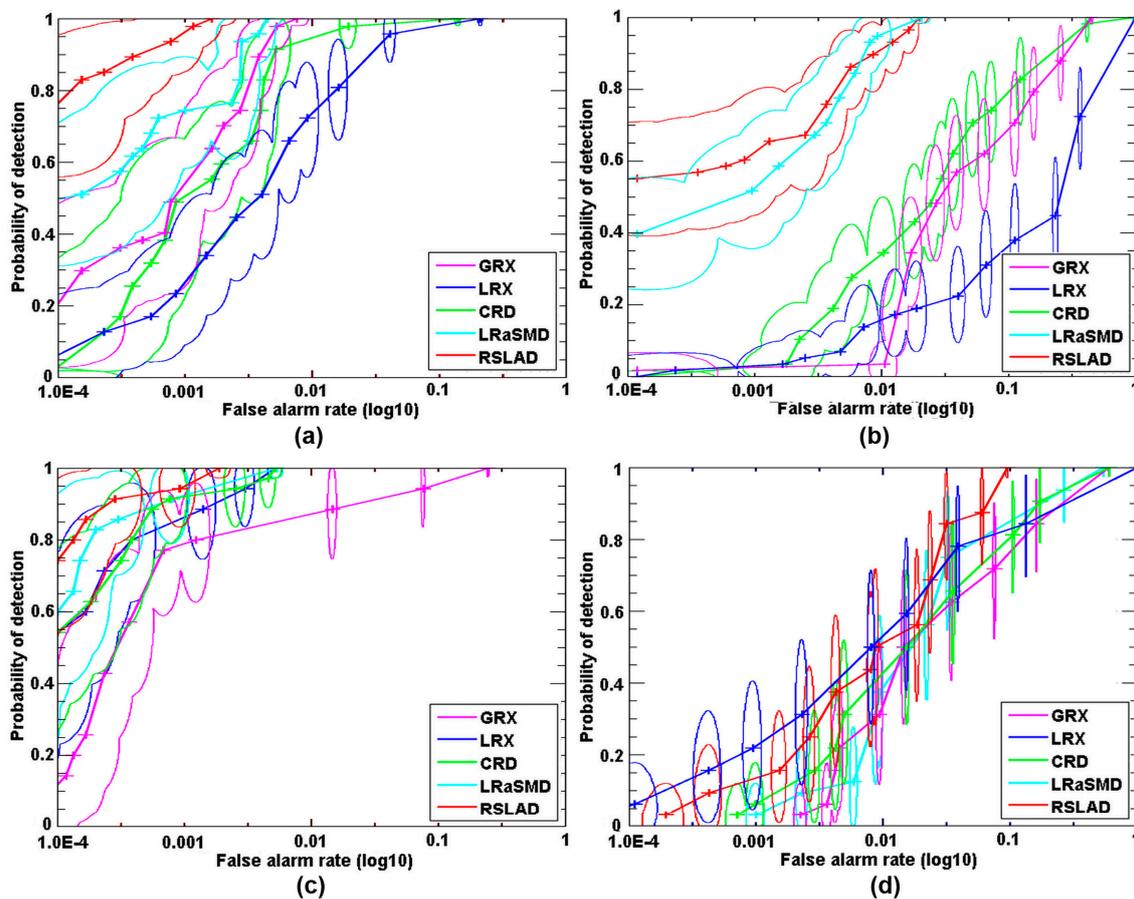

Figure 7. The logarithmic receiver operating characteristic (ROC) curves and confidence regions on the four HSI datasets: (**a**) PaviaC; (**b**) San Diego; (**c**) Botswana; and (**d**) HyMap.

Figure 8 lists the AUC results of RSLAD and other four methods on the four datasets. The observations of AUC coincide with those of Figure 7. The RSLAD has the highest AUCs on four HSI datasets among all the five methods, and the second highest is LRaSMD. The CRD performs better than LRX in the AUCs. Figures 9–11 show the detection maps of all five methods on the PaviaC, San Diego and Botswana datasets with normalized anomaly values between 0 and 1. We did not show the HyMap detection maps because of too small and even invisible anomalies with respect to the overlarge image size.
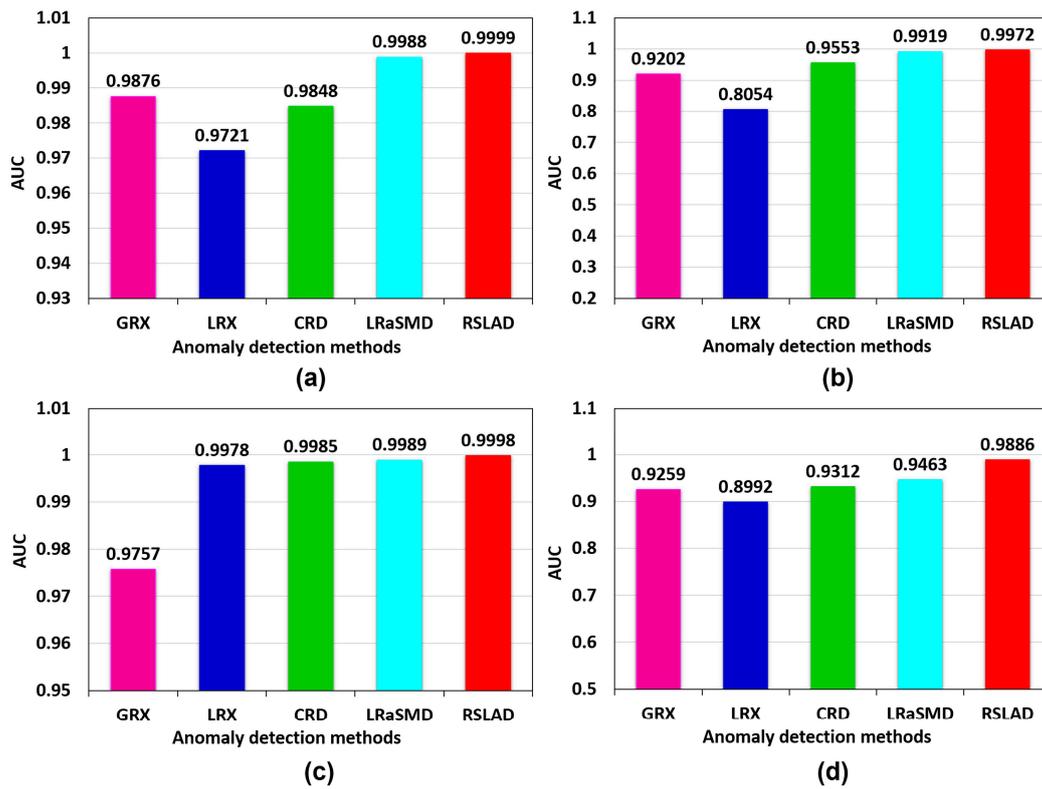
Figure 8. The comparison in area under curve (AUC) between RSLAD and four comparative methods on the four HSI datasets: (**a**) PaviaC; (**b**) San Diego; (**c**) Botswana; and (**d**) HyMap.
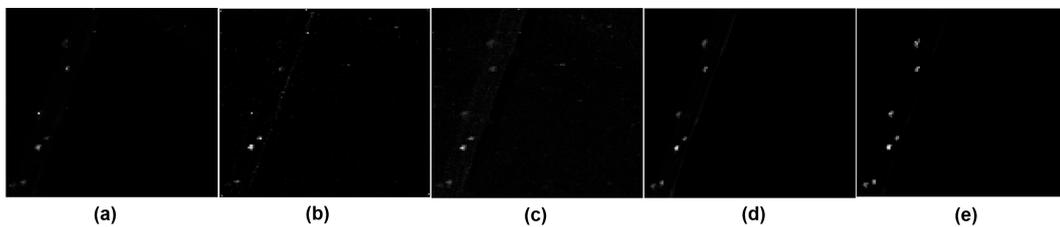

Figure 9. The normalized detection maps of PaviaC dataset before threshold segmentation: (**a**) GRX; (**b**) LRX; (**c**) CRD; (**d**) LRaSMD; and (**e**) RSLAD.
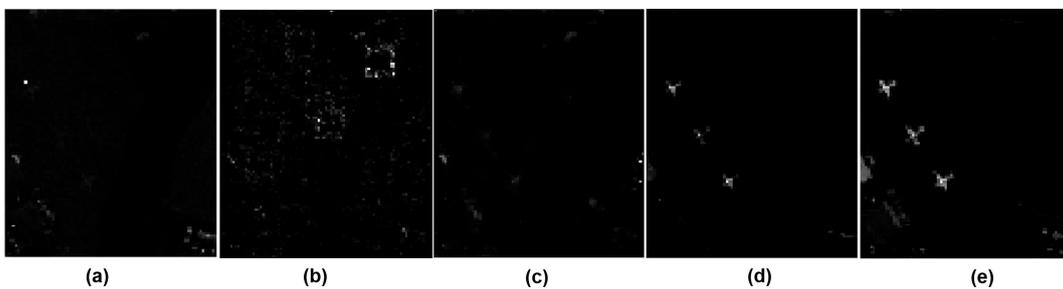

Figure 10. The normalized detection maps of San Diego dataset before threshold segmentation: (**a**) GRX; (**b**) LRX; (**c**) CRD; (**d**) LRaSMD; and (**e**) RSLAD.
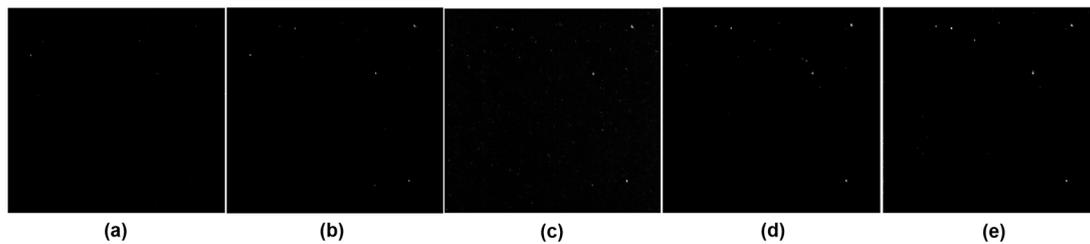
Figure 11. The normalized detection maps of Botswana dataset before threshold segmentation: (**a**) GRX; (**b**) LRX; (**c**) CRD; (**d**) LRaSMD; and (**e**) RSLAD.

Table 2 compares the computational time of all five methods on the four datasets. All the methods are implemented in Matlab 2016a and their codes are run on a WIN10 computer with Intel Core (TM) i7-6700 CPU 3.40 GHz and 32 GB of RAM. The LRX and CRD cost longer computational times than the other three methods GRX, LRaSMD and RSLAD, and the CRD takes the longest computational times of all. The LRaSMD takes shorter computational times than LRX and CRD but its computational speed is much slower than GRX and RSLAD. The RSLAD takes slightly longer time than GRX on the PaviaC dataset, but it clearly surpasses the GRX in computational speed on the San Diego, Botswana and HyMap datasets.

Table 2. The Computations Times of RSLAD and Other Four Methods on the Four Datasets.

Sensors	Datasets	Computational Time of Anomaly Detection Methods (s)				
		GRX	LRX	CRD	LRaSMD	RSLAD
ROSIS	PaviaC	0.144	34.2641	49.582	14.116	0.156
AVIRIS	San Diego	0.304	88.713	107.477	19.423	0.252
EO-1 Hyperion	Botswana	0.925	397.485	514.025	101.915	0.487
HyMap	HyMap	1.373	392.587	1223.322	151.824	0.560

4.2.2. Performance Sensitivity to the Number of Sampled Pixels p

In the experiment, the ranges of sampled pixel numbers p on the PaviaC, San Diego and Botswana datasets are manually set as [60, 120, 300, 600, 1200, 3000, 6000]. Figure 12 illustrates the ROC curves from the RSLAD on the three HSI datasets by changing number of sampled pixels p. For the PaviaC data of Figure 12a, the detection performance increases with the rising p from 60 to 6000, and the RSLAD performs the best when the sampled pixel number p equals 300. After that, the detection performance of RSLAD begins to decrease. The ROC curve with p equal to 6000 performs the worst of all, having the highest false alarm rate at 100% probability of detection. The ROC curves of San Diego data in Figure 12b have similar observations. The performance is the best when the p equals 120. For the Botswana dataset in Figure 12c, the ROC curve of p at 60 performs the best of all, having the lowest false alarm rate at 100% probability of detection.

Moreover, Figure 13a shows the result of AUC curves with the changing number of sampled pixels p on the three datasets. In the figure, the AUCs of RSLAD on three HSI datasets gradually decrease with the increasing p from 60 to 6000. Particularly, the AUC curves have the greatest fall on the San Diego dataset, decreasing from 0.9965 to 0.2157 as the p increases from 60 to 6000. The observations of AUC curves coincide with those of ROC curves in Figure 12. Moreover, Figure 13b plots the curves of computational times with respect to the increasing sampled pixel number p. The results show that computational time of RSLAD increases with the rising p. Especially, the computational times drastically increase when having a larger pixel number p over 1200. Since a large value of p degrades the detection performance and significantly increase computing time, a smaller value of p, e.g., 120, is preferred.
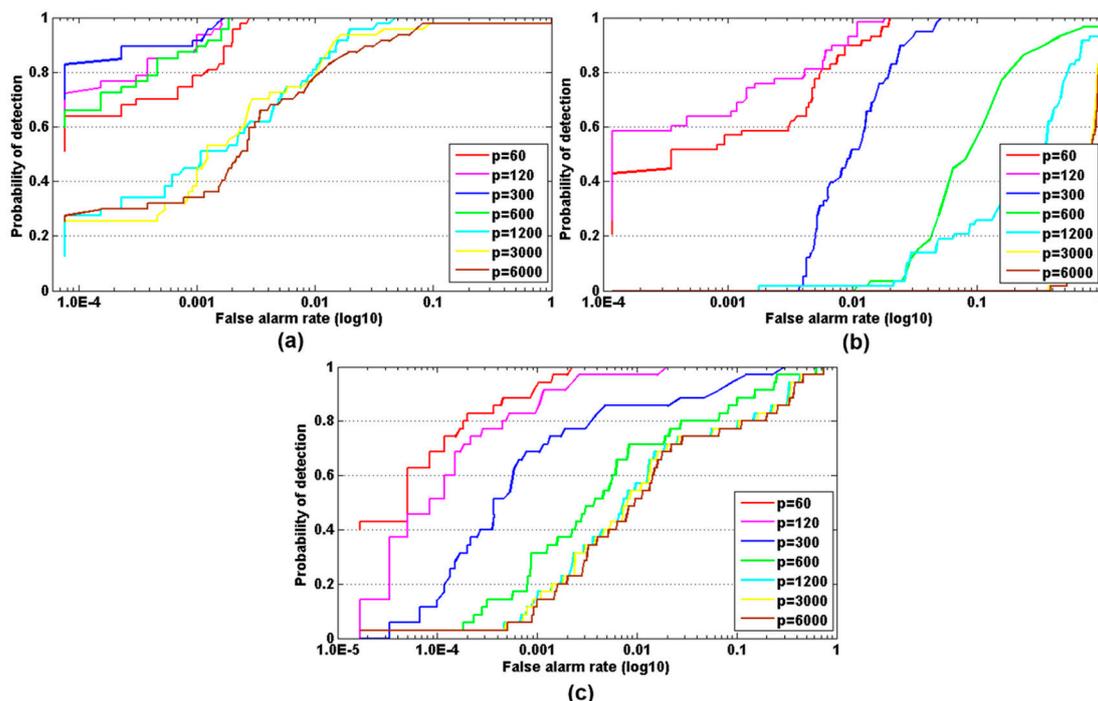
Figure 12. The logarithmic receiver operating characteristic (ROC) curves of RSLAD with different choices of sampled pixels *p* on the three datasets: (**a**) PaviaC; (**b**) San Diego; and (**c**) Botswana.
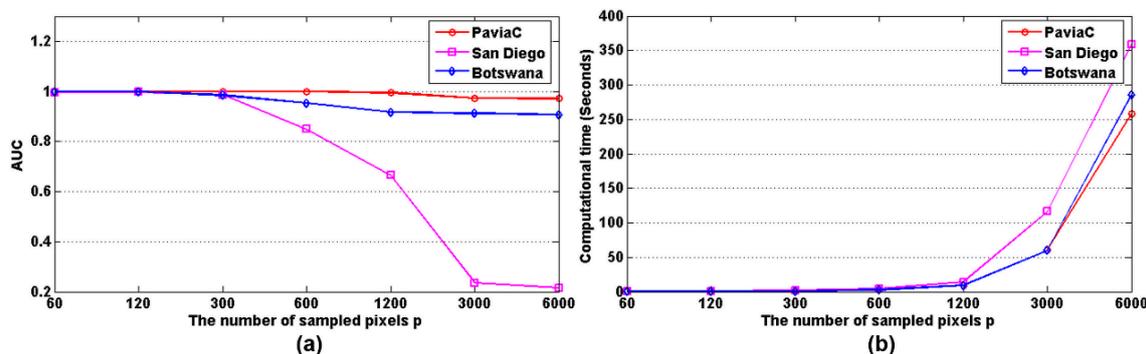

Figure 13. The curves of: (**a**) area under curve (AUC); and (**b**) computation time with different choices of sampled pixels *p* on the three HSI datasets.

4.2.3. Performance Sensitivity to the Projected Dimension *K*

The projected dimension *K* correlates with the number of random Hadamard projections in the RSLAD. In the experiment, the projected dimensions *K* on the PaviaC, San Diego and Botswana datasets are manually set between 10 and 100 with a step interval of 10. Figure 14 shows the ROC curves of RSLAD on the three datasets with the changing projected dimensions *K* from 10 to 100. For the PaviaC dataset of Figure 14a, when the *K* increases from 10 to 100, the ROC curves of RSLAD have similar trend with small fluctuations. Particularly, the false alarm rates of most ROC curves are concentrated between 0.001 and 0.003 when achieving 100% probability of detection. The ROC curves of San Diego dataset in Figure 14b and Botswana dataset in Figure 14c have similar observations with that of Figure 14a. In Figure 14b,c, the ROC curves of RSLAD are overlaid with each other, despite of the changing projected dimensions *K*.
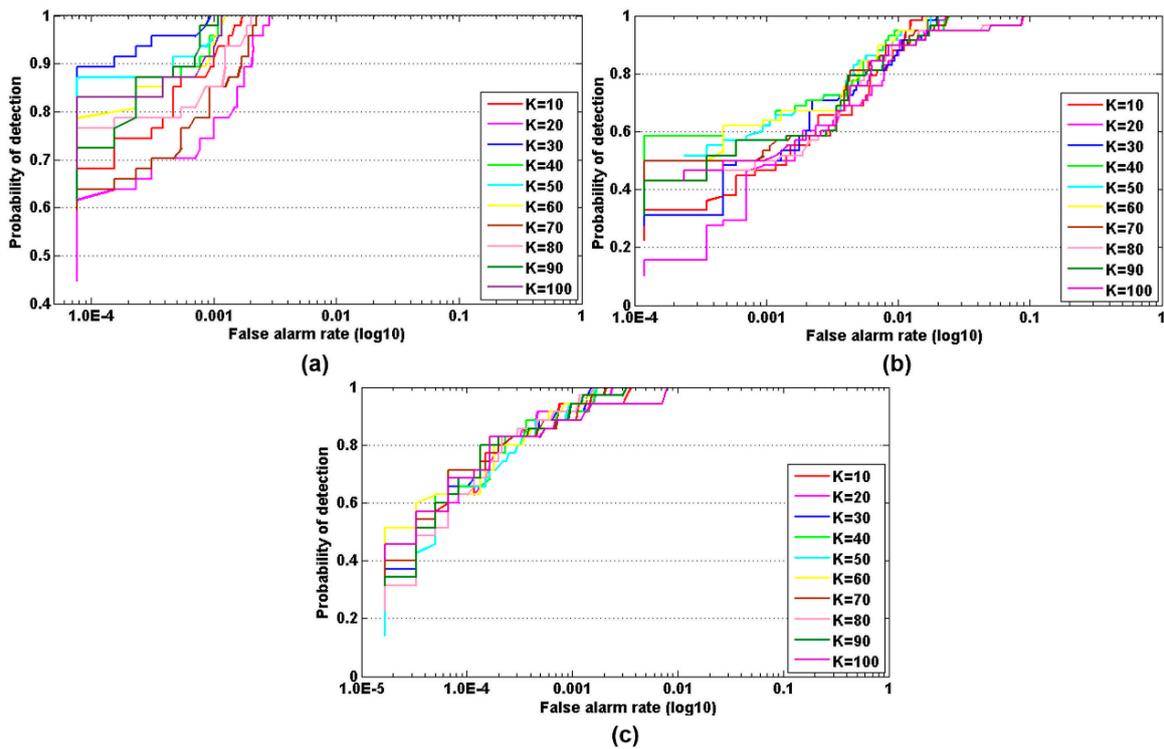
Figure 14. The logarithmic receiver operating characteristic (ROC) curves of RSLAD with different choices of projection dimensions *K* on the three datasets: (**a**) PaviaC; (**b**) San Diego; and (**c**) Botswana.

Figure 15a illustrates the AUC curves of RSLAD on the three datasets when the projected dimensions *K* change from 10 to 100. The results show that the AUC curves of RSLAD are not much affected by the changing projected dimension *K*. Among them, the AUC curves on San Diego has largest fluctuations ranging from 0.9966 to 0.9977. That further supports the above observations of ROC curves in Figure 14. Moreover, Figure 15b shows the computational times of RSLAD on the three datasets when *K* from 10 to 100, where the computational times linearly increases with the *K*. The detection performance of RSLAD is less insensitive to the projected dimension *K*. The computational times of RSLAD linearly increases with the projected dimension *K*. Therefore, we will recommend to adopt a small projected dimension *K*, say, *K* = 50.
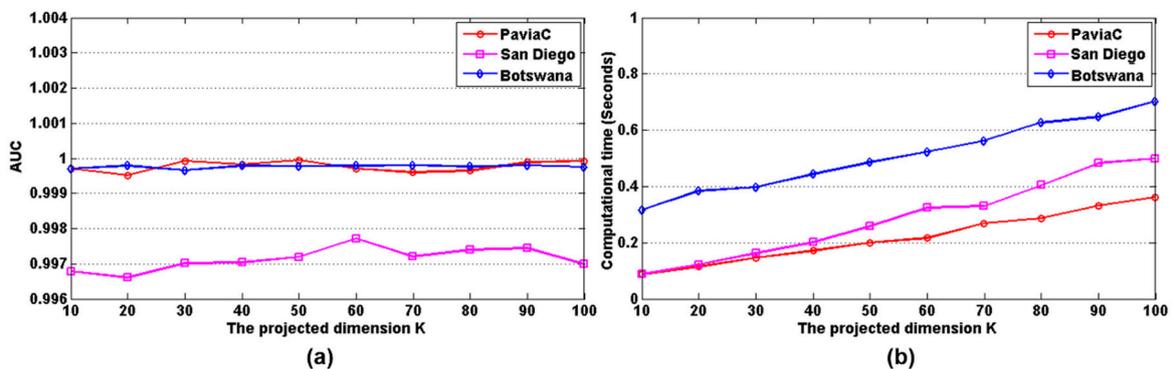

Figure 15. The curves of: (**a**) area under curve (AUC); and (**b**) computation time with different choices of projected dimension *K* on the three HSI datasets.

4.2.4. Performance Sensitivity to the Residual Threshold

This experiment investigates the impacts from the residual threshold ε in the detection performance of RSLAD on the PaviaC, San Diego and Botswana HSI datasets. With prior knowledge about the range of residual errors for all the pixels, we manually set the range of residual thresholds ε on the three datasets as $[10^{-12}, 10^{-11}, 10^{-10}, 10^{-9}, 10^{-8}]$. According to previous investigation, the sampled pixel numbers p is set to be 120, and the projected dimensions K is 50.

Figure 16a–c shows the ROC curves of RSLAD on the three datasets with different residual threshold. For the PaviaC dataset of Figure 16a, with the increase in ε from 10^{-12} to 10^{-8}, the detection performance of RSLAD first increases and then drastically decreases. The ROC curve performs best when the ε equal to 10^{-9}. The ROC curve at ε equal to 10^{-8} is worst of all, having the largest false alarm rate when achieving 100% probability of detection. The similar observations exist in ROC curves of the San Diego dataset in Figure 16b. The detection performance of RSLAD first increases and then achieves the best performance when the ε is equal to 10^{-10}. After that, further increasing the ε will lower the detection performance of RSLAD. The ROC curves of Botswana dataset in Figure 16c coincides with the observations in Figure 16b. The ROC curve at ε equal to 10^{-10} performs best, and the ROC curve at ε equal to 10^{-12} is the worst. Moreover, the AUC curves of RSLAD on three datasets in Figure 16d further support the above observations of ROC curves. With the increase in ε from 10^{-12} to 10^{-8}, the AUC first increases, achieves a maximum and then gradually decreases until the end of the range.
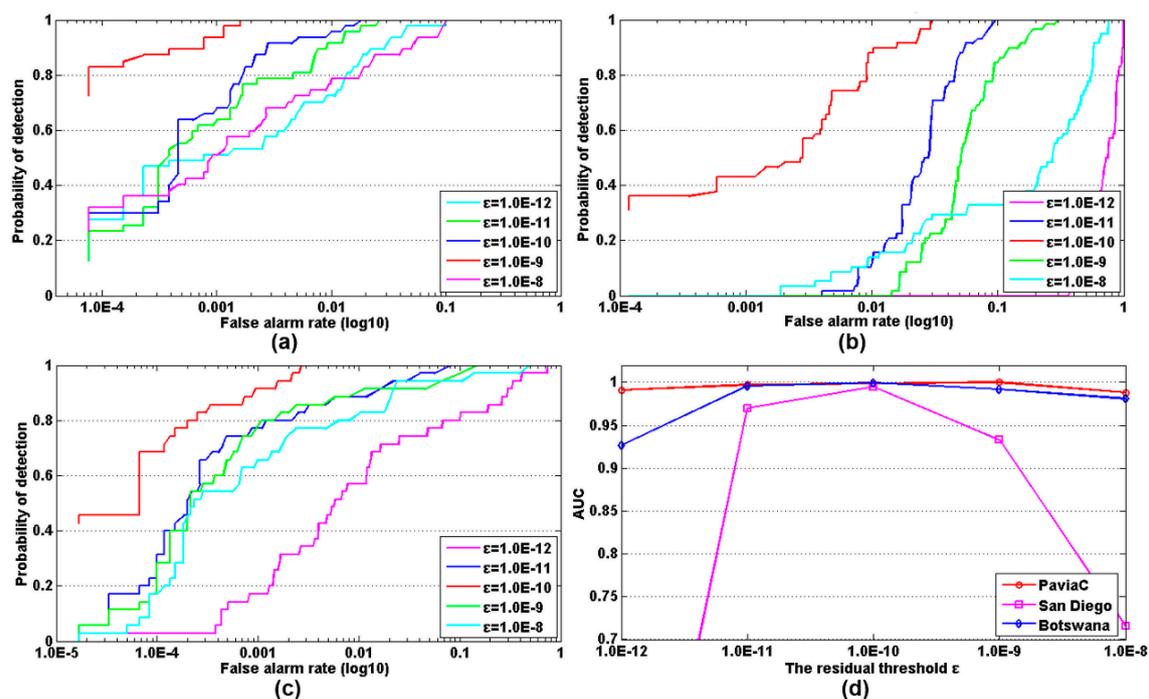

Figure 16. The logarithmic receiver operating characteristic (ROC) curves of RSLAD on: (**a**) PaviaC; (**b**) San Diego; and (**c**) Botswana, with different choices of (**d**)residual threshold .

5. Discussion

In Section 4.2, four groups of experiments are designed to analyze the performance of RSLAD. The first experiment compares the detection results of RSLAD with those of four state-of-the-art methods, i.e., GRX, LRX, CRD and LRaSMD. Experimental results show that the RSLAD has the smallest false alarm rate under 100% probability of detection on the four HSI datasets. The confidence intervals and regions are also adopted to verify the significant difference between RSLAD and other detectors. The superiority of RSLAD to GRX is statistically significant in all the four datasets, and the

ROC curves of RSLAD and LRaSMD show less statistical difference from each other. The better performance of RSLAD compared to CRD and LRX is statistically significant on the PaviaC and San Diego datasets, whereas it is less statistically significant on both Botswana and HyMap datasets. As discussed in [52], one could not guarantee a detector is always superior to other detection methods. Fortunately, the RSLAD yields better detection performance than GRX and takes less computational time than other detectors (i.e., LRaSMD, GRX, LRX and CRD). Therefore, the good detection performance and lower computational cost make the RSLAD a great alternative for hyperspectral anomaly detection.

The second experiment investigates the detection performance sensitivity of RSLAD with respect to the sampled pixels p. Experimental results show that the detection performance and computational times of RSLAD are sensitive to the number of sampled pixels p. Both the detection performance and computational speed of RSLAD decrease with the increasing pixel number p. Particularly, a small number of sample pixel could bring about good performance of RSLAD in detection and computational efficiency. In contrast, a too large sampled pixel number p severely decreases the detection performance of RSLAD and also bring about extremely high computational costs. The explanation for that is as follows. The detection performance of RSLAD relies on the orthonormal basis estimation **U** of the column subspace of background **B**. The purified randomized subspace using Equation (4) takes enough independent background columns and can be seen a basis **U** for the background matrix **B**. If a too large p pixels were randomly sampled from the original HSI dataset, the purified matrix would have too many dependent columns that represent similar background pixels. That reduces the independence among the columns of **U** and accordingly degrades the detection performance of RSLAD. From the above analysis, we recommend to adopt a small sampled pixel number p in the following experiment to guarantee good detection performance and low computational costs.

The third experiment analyzes the detection performance sensitivity of RSLAD to the projected dimension K. Experimental results show that the detection performance of RSLAD is insensitive to the projected dimension K. However, the computational times of RSLAD linearly increases with the projected dimension K. The explanation for the above conclusions is as follows. The projected dimension K could lower the computational requirements of the original HSI dataset by reducing its dimensionality from M to K. However, the projected dimension K does not affect the coarse randomized subspace $\mathbf{Y_\Phi}$ of the background. Accordingly, it shows no correlations with the detection performance of the RSLAD. From the above analysis, we recommend to adopt a small projected dimension K in further experiments to reduce the computation costs of RSLAD.

The final experiment explores the detection performance sensitivity of RSLAD to the residual threshold ε. Experimental results show that the detection performance of RSLAD is sensitive to the residual threshold ε. A too-small or too-large threshold ε from its predefined range would negatively affect the detection performance of RSLAD. The explanation for that is as follows. The residual threshold ε correlates with the estimation of anomaly columns in $\mathbf{Y_\Phi}$ and determines the accurate estimation of orthonormal basis **U** for the column space of background **B**. If a too-small threshold ε were adopted, too many columns (i.e., sampled background pixels) would be removed from $\mathbf{Y_\Phi}$. Too few independent columns of **U** could not fully describe the subspace structure of the background and accordingly negatively affect the detection performance of RSLAD. In contrast, if a too-large threshold were adopted, the anomaly columns from random sampling would be left in the columns of purified matrix **U**. That would degrade the estimation of orthonormal basis **U** and negatively affect the detection performance of RSLAD. From the above analysis from three HSI datasets, we recommend to initialize the residual threshold ε on the HSI dataset with 10^{-10}, make slight tuning around the initials, and select the proper threshold that has the best performance in ROC and AUC.

Our RSLAD still has some drawbacks and needs careful investigations in the future work. First, we could not provide an intelligent estimation scheme for the threshold ε. The automatic parameter estimation of residual threshold ε will be carefully studied in our following work. Second, the quantitative relationship between sampled pixel number p and the detection performance

(e.g., ROC and AUC) will be further explore to provide a clear mathematical expression. Third, the RSLAD will be implemented on more hyperspectral images with larger image scenes and compared with more recently proposed anomaly detectors, e.g., transferred deep convolutional neural network (CNN) [30], LRRaLD [34], Local Graph based anomaly detection (LGAD) [53] and differential attribute profile anomaly detection (DAPAD) [46]. Finally, the RSLAD will be integrated into an anomaly detection Matlab toolbox and can be downloaded later from our website to benefit practical applications of hyperspectral techniques.

6. Conclusions

In this paper, we proposed a RSLAD method to detect anomalies in hyperspectral images. The RSLAD is inspired by the RPCA model and assumes that the background is low-rank and the anomaly matrix is column-wise sparse. It finds a randomized subspace of the background where the effects from anomalies are excluded and then detects anomalies by projecting all the HSI pixels on the orthogonal subspace of the purified randomized column subspace. Experimental results on four hyperspectral datasets demonstrate that the RSLAD can offer good detection performance with lower computational cost.

Author Contributions: All coauthors made significant contributions to the paper. Weiwei Sun presented the key idea of the RSLAD method and carried on the majority of comparison experiments. Yan Xu and Long tian performed major tunings in the parameters involved in the comparison methods. Do Bo and Qian Du provided the background knowledge of hyperspectral anomaly detection and helped to revise the paper.

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

References

1. Tong, Q.; Xue, Y.; Zhang, L. Progress in hyperspectral remote sensing science and technology in china over the past three decades. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 70–91. [CrossRef]
2. 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]
3. Matteoli, S.; Diani, M.; Corsini, G. A total overview of anomaly detection in hyperspectral images. *IEEE Aerosp. Electron. Ssyt. Mag.* **2010**, *25*, 5–28. [CrossRef]
4. Nasrabadi, N.M. Hyperspectral target detection: An overview of current and future challenges. *IEEE Signal Process. Mag.* **2014**, *31*, 34–44. [CrossRef]
5. 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]
6. Matteoli, S.; Diani, M.; Theiler, J. An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2317–2336. [CrossRef]
7. Trilla, G.G.; Pratolongo, P.; Kandus, P.; Beget, M.; Di Bella, C.; Marcovecchio, J. Relationship between biophysical parameters and synthetic indices derived from hyperspectral field data in a salt marsh from Buenos Aires Province, Argentina. *Wetlands* **2016**, *36*, 185–194. [CrossRef]
8. Jay, S.; Guillaume, M.; Minghelli, A.; Deville, Y.; Chami, M.; Lafrance, B.; Serfaty, V. Hyperspectral remote sensing of shallow waters: Considering environmental noise and bottom intra-class variability for modeling and inversion of water reflectance. *Remote Sens. Environ.* **2017**, *200*, 352–367. [CrossRef]
9. Makki, I.; Younes, R.; Francis, C.; Bianchi, T.; Zucchetti, M. A survey of landmine detection using hyperspectral imaging. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 40–53. [CrossRef]
10. Wang, Q.; Yuan, Y.; Yan, P.; Li, X. Saliency detection by multiple-instance learning. *IEEE Trans. Cybern.* **2013**, *43*, 660–672. [CrossRef] [PubMed]

11. Wang, Q.; Yuan, Y.; Yan, P. Visual saliency by selective contrast. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 1150–1155. [CrossRef]

12. Sun, W.; Liu, C.; Li, J.; Lai, Y.M.; Li, W. Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery. *J. Appl. Remote Sens.* **2014**, *8*, 083641. [CrossRef]

13. Zhang, L.; Zhang, L.; Tao, D.; Huang, X.; Du, B. Hyperspectral remote sensing image subpixel target detection based on supervised metric learning. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4955–4965. [CrossRef]

14. Du, B.; Zhang, L. A discriminative metric learning based anomaly detection method. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6844–6857.

15. Reed, I.S.; Yu, X. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Trans. Acoust. Speech Signal Process.* **1990**, *38*, 1760–1770. [CrossRef]

16. Chang, C.-I.; Chiang, S.-S. Anomaly detection and classification for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1314–1325. [CrossRef]

17. Zhang, Y.; Du, B.; Zhang, L. A sparse representation-based binary hypothesis model for target detection in hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1346–1354. [CrossRef]

18. Stein, D.W.; Beaven, S.G.; Hoff, L.E.; Winter, E.M.; Schaum, A.P.; Stocker, A.D. Anomaly detection from hyperspectral imagery. *IEEE Signal Process. Mag.* **2002**, *19*, 58–69. [CrossRef]

19. Carlotto, M.J. A cluster-based approach for detecting man-made objects and changes in imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 374–387. [CrossRef]

20. Kwon, H.; Nasrabadi, N.M. Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 388–397. [CrossRef]

21. Chang, C.-I. Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 502–518. [CrossRef]

22. Kwon, H.; Der, S.Z.; Nasrabadi, N.M. Adaptive anomaly detection using subspace separation for hyperspectral imagery. *Opt. Eng.* **2003**, *42*, 3342–3351. [CrossRef]

23. Billor, N.; Hadi, A.S.; Velleman, P.F. Bacon: Blocked adaptive computationally efficient outlier nominators. *Comput. Stat. Data Anal.* **2000**, *34*, 279–298. [CrossRef]

24. Smetek, T.E.; Bauer, K.W., Jr. A comparison of multivariate outlier detection methods for finding hyperspectral anomalies. *Eff. Employ. Non-React. Sens.* **2008**, *13*, 19–43. [CrossRef]

25. Du, B.; Zhang, L. Random-selection-based anomaly detector for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1578–1589. [CrossRef]

26. Banerjee, A.; Burlina, P.; Diehl, C. A support vector method for anomaly detection in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2282–2291. [CrossRef]

27. Gurram, P.; Kwon, H.; Han, T. Sparse kernel-based hyperspectral anomaly detection. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 943–947. [CrossRef]

28. Ma, L.; Crawford, M.M.; Tian, J. Anomaly detection for hyperspectral images based on robust locally linear embedding. *J. Infrared Millim. Terahertz Waves* **2010**, *31*, 753–762. [CrossRef]

29. Ma, L.; Crawford, M.M.; Tian, J. Anomaly detection for hyperspectral images using local tangent space alignment. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Honolulu, HI, USA, 25–30 July 2010; pp. 824–827.

30. Li, W.; Wu, G.; Du, Q. Transferred deep learning for anomaly detection in hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 597–601. [CrossRef]

31. Zhang, X.; Wen, G.; Dai, W. A tensor decomposition-based anomaly detection algorithm for hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5801–5820. [CrossRef]

32. Li, W.; Du, Q. Collaborative representation for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1463–1474. [CrossRef]

33. Xu, Y.; Wu, Z.; Li, J.; Plaza, A.; Wei, Z. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1990–2000. [CrossRef]

34. Niu, Y.; Wang, B. Hyperspectral anomaly detection based on low-rank representation and learned dictionary. *Remote Sens.* **2016**, *8*, 289. [CrossRef]

35. Zhao, R.; Du, B.; Zhang, L. Hyperspectral anomaly detection via a sparsity score estimation framework. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3208–3222. [CrossRef]

36. Cui, X.; Tian, Y.; Weng, L.; Yang, Y. Anomaly detection in hyperspectral imagery based on low-rank and sparse decomposition. In Proceedings of the Fifth International Conference on Graphic and Image Processing, Hong Kong, China, 10 January 2014; p. 90690R.

37. Zhang, Y.; Du, B.; Zhang, L.; Wang, S. A low-rank and sparse matrix decomposition-based mahalanobis distance method for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1376–1389. [CrossRef]

38. Candes, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *arXiv*, 2009.

39. Zhou, T.; Tao, D. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In Proceedings of the 2011 International conference on machine learning, Bellevue, WA, USA, 28 June–2 July 2011; pp. 33–40.

40. Wright, J.; Ganesh, A.; Rao, S.; Peng, Y.; Ma, Y. *Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization*; Advances in Neural Information Processing Systems: La Jolla, CA, USA, 2009; pp. 2080–2088.

41. Xu, H.; Caramanis, C.; Sanghavi, S. *Robust PCA via Outlier Pursuit*; Advances in Neural Information Processing Systems: La Jolla, CA, USA, 2010; pp. 2496–2504.

42. Chandrasekaran, V.; Sanghavi, S.; Parrilo, P.A.; Willsky, A.S. Rank-sparsity incoherence for matrix decomposition. *SIAM J. Opt.* **2011**, *21*, 572–596. [CrossRef]

43. Rahmani, M.; Atia, G. Randomized robust subspace recovery for high dimensional data matrices. *arXiv*, **2015**, arXiv:1505.05901.

44. Menon, V.; Du, Q.; Fowler, J.E. Random hadamard projections for hyperspectral unmixing. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 419–423. [CrossRef]

45. Menon, V.; Du, Q.; Fowler, J.E. Fast svd with random hadamard projection for hyperspectral dimensionality reduction. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1275–1279. [CrossRef]

46. Taghipour, A.; Ghassemian, H. Hyperspectral anomaly detection using attribute profiles. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1136–1140. [CrossRef]

47. Liu, Y.; Gao, G.; Gu, Y. Tensor matched subspace detector for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 1967–1974. [CrossRef]

48. Ma, L.; Crawford, M.M.; Tian, J. Local manifold learning-based k-nearest-neighbor for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4099–4109. [CrossRef]

49. Snyder, D.; Kerekes, J.; Fairweather, I.; Crabtree, R.; Shive, J.; Hager, S. Development of a web-based application to evaluate target finding algorithms. In Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Boston, MA, USA, 8–11 July 2008; pp. II-915–II-918.

50. Nanceb, W.F.B.; Nance, E.; Kerekesa, J. The target implant method for predicting target difficulty and detector performance in hyperspectral imagery. *Proc. SPIE* **2011**, *2011*. [CrossRef]

51. Kerekes, J. Receiver operating characteristic curve confidence intervals and regions. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 251–255. [CrossRef]

52. Manolakis, D.; Lockwood, R.; Cooley, T.; Jacobson, J. Is there a best hyperspectral detection algorithm? *Proc. SPIE* **2009**, *7334*. [CrossRef]

53. Yuan, Y.; Ma, D.; Wang, Q. Hyperspectral anomaly detection by graph pixel selection. *IEEE Trans. Cybern.* **2016**, *46*, 3123–3134. [CrossRef] [PubMed]