Anomaly Detection for Hyperspectral Images Based on Anisotropic Spatial-Spectral Total Variation and Sparsity Constraint

Huihui Ju¹, Zhigang Liu¹, Jiangjun Jiang¹ and Yang Wang¹

¹ Xi'an Research Institute of Hi-Tech, Tongxin Street, Hongqing Town, Xi'an

Abstract: A novel anomaly detection method for hyperspectral images (HSIs) is proposed based on anisotropic spatial-spectral total variation and sparse constraint. HSIs are assumed to be not only smooth in spectral dimension but also piecewise smooth in spatial dimension. The proposed method adopts the anisotropic spatial-spectral total variation model which combines 2D spatial total variation and 1D spectral variation to explore the spatial-spectral smooth property of HSIs. Meanwhile, the sparse property of anomalies is exploited for its low probability in the image. To utilize both the spatial and spectral information of HSIs, we preserve the original cubic form of HSIs and divide the HSIs into three 3D arrays, each representing the background, the anomaly, and the noise respectively. By using anisotropic spatial-spectral total variation regularization on the background component and sparse constraint on the anomaly component, this anomaly detection problem has therefore been formulated as a constraint optimization problem whose solution has been derived by alternately using Split Bregman Method and Go Decomposition (GoDec) Method. Experimental results on hyperspectral datasets illustrate that our proposed method has a better detection performance than state-of-the-art hyperspectral anomaly detection methods.

1 INTRODUCTION

Hyperspectral sensors can collect a spectral vector with hundreds or thousands of elements from each pixel in a given scene (Bioucas-Dias et al, 2013), and the data product, which is generally termed hyperspectral image (HSI), is a three-dimensional array or “cube” with the width and length of the array corresponding to spatial dimension and the spectrum of each point as the third dimension (Stein et al, 2002). With abundant spatial and spectral information, HSIs possess a strong diagnostic capability to discriminate different materials and a unique advantage in target detection. With available targets spectra, target detection methods for HSIs are generally called spectral matching algorithms (Matteoli et al, 2010). Before applying spectral matching algorithms, atmosphere calibration is needed to convert the raw radiance data collected from the hyperspectral sensors into reflectance or convert the targets spectra into radiance at the sensor. However, these involve many complicated factors in practical applications. Besides, different species of objects may have the same spectrum, and the same object may have different spectra (Matteoli et al. 2014). These restrictions of spectral matching methods have therefore led to the development of anomaly detection (AD). Anomaly detection means

detecting objects without any reference to target information (Matteoli et al, 2010). Usually, “anomalies” refer to two aspects: (1) its spectral signatures are different from background and (2) they have a low probability in the whole image (Stein et al, 2002).

In the literature, many AD algorithms have been put forward. Among them, Reed-Xiaoli (RX) (Reed and Yu, 1990) algorithm is considered as the benchmark for HSIs AD. RX assumes that the spectral signatures of the background pixels follow a multivariate Gaussian distribution. However, the background is usually so complex that it does not follow a multivariate Gaussian distribution. And the existence of anomalies and noise often makes the background estimation inaccurate. Therefore, some methods focus on making background closer to a multivariate Gaussian distribution and the background covariance matrix estimation more accurate. BACON (Billor et al, 2000), RSAD (Du and Zhang, 2011), and WRXD (Guo et al, 2014) refine the background estimation by removing anomalies or reducing the weight of anomalies in the background samples. Cluster based anomaly detection (CBAD) (Carlotto, 2005) algorithm holds that a global Gaussian mixture model (GMM) is more appropriate to characterize the complex background. Moreover, some methods try to detect anomalies by suppressing the background. Based on the linear mixture model (LMM)

* Corresponding author: juhuihui95@gmail.com, dannlzg@163.com, jiangjiangjun85@foxmail.com, 1982450862@qq.com

that each pixel is a linear combination of endmembers, an unsupervised orthogonal subspace projection (OSP) (Chang, 2005) method detects the anomalies by projecting the data onto the subspace orthogonal to the background subspace to suppress background components. Usually the background subspace is unknown, and it has to be adaptively estimated from the image itself, which is quite difficult.

In addition to aforementioned methods, recently, low-rank and sparse matrix decomposition-based methods for HSIs AD have gained much attention. This idea comes from that the background has the low-rank property due to the high correlation of spectral information (Zhang et al, 2014) and the anomalies are sparse for its low probability (Cui et al, 2014). By resizing the hyperspectral cube into 2D matrix, the image can be decomposed into a low-rank matrix plus a sparse matrix, representing the background and the anomaly respectively (Chen and Chang, 2013). Candes presents the robust principal component analysis (RPCA) (Candes et al, 2009) to recover the low-rank matrix and the sparse matrix. Considering the universal noise in the image, Zhou and Tao (Zhou and Tao, 2011) develop the model of RPCA by adding a Gaussian noise term and propose Go Decomposition (GoDec) to recover the low-rank matrix and the sparse matrix from a high-dimensional data matrix corrupted by noise. Applying GoDec model to detect anomalies is generally called low-rank and sparse matrix decomposition-based anomaly detection (LRaSMD). More recently, considering the ubiquitous mixed pixels in HSIs, Liu (Liu et al, 2010) assumes that the low-rank matrix representing the background should lie in multiple subspaces instead of a single subspace proposed in RPCA model. This low-rank representation (LRR) model is consistent with the LMM. Based on this idea, AD methods based on low-rank and sparse representation (LRASR) (Xu et al, 2016) have been proposed. Besides, taking the spatial correlation into consideration, a collaborative representation-based anomaly detection (CRD) (Li and Du, 2015) algorithm has been proposed recently. It is based on the concept that each pixel in the background can be approximately represented by its spatial neighborhoods, whereas anomalies cannot. However, this method fails to work well when anomalies are placed in close proximity.

By rearranging the 3D hyperspectral data into 2D matrix, the above low-rank-based anomaly detection methods destroy the spatial correlation of the pixels in neighboring locations and only exploit spectral correlation of HSIs. Although CRD has taking the spatial correlation into consideration, it ignores the spectral correlation of HSIs. In this paper, we intend to integrate the spectral correlation and spatial correlation to explore the piecewise spatial and spectral smooth property of HSIs (Bioucas-Dias et al, 2013). Total variation (TV) (Rudin et al, 1992) model is an effective tool to impose the spatial constraint (Sun et al, 2017) and has been widely used to keep the spatial smoothness of neighboring pixels. It has been very active in image denoising and restoration (Osher et al, 2006; He et al, 2015; Aggarwal and Majumdar, 2016). Anisotropic

spatial-spectral total variation (ASSTV) (Sun et al, 2017; Chang et al, 2015; Wang et al, 2017), which is defined as a combination of 2D spatial TV and 1D spectral variation of the HSIs, has recently been proposed to keep the spatial-spectral smoothness of HSIs. Meanwhile, considering the low probability of anomalies in the image, the anomalies are still assumed to have the sparse property. To utilize both the spatial and spectral information of HSIs, we preserve the original cubic form of HSIs and divide HSIs into three 3D arrays, each representing the background component, the anomaly component and the noise component respectively. Combining with the above analysis, a novel anomaly detection method based on anisotropic spatial-spectral total variation and sparse constraint (ASSTVSC) is proposed with the assumption that the background has the spatial-spectral smooth property, the anomaly is sparse and the noise is independent and identically distributed Gaussian noise.

2 PROBLEM FORMULATION

2.1 HSIs Decomposition Model

Let $\tilde{X} \in R^{m \times n \times b}$ and $\tilde{A} \in R^{m \times n \times b}$ represent the background component and the anomaly component respectively, where m, n represent the rows and columns of the image and b is the number of spectral bands. Then, the observed HSI cube $\tilde{Y} \in R^{m \times n \times b}$, corrupted by Gaussian noise $\tilde{N} \in R^{m \times n \times b}$, can be formulated as:

$$\tilde{Y} = \tilde{X} + \tilde{A} + \tilde{N} \quad (1)$$

2.2 ASSTV Model

In most natural hyperspectral images, spatial neighboring pixels are usually correlated and neighboring bands also exhibit high spectral correlation. This prior information can be exploited using spatial-spectral TV model in an anisotropic way. The ASSTV can be formulated as:

$$\text{ASSTV}(\tilde{X}) = \alpha \|D_m \tilde{X}\|_1 + \beta \|D_n \tilde{X}\|_1 + \gamma \|D_b \tilde{X}\|_1 \quad (2)$$

where D_m and D_n denote the horizontal and vertical difference operators along the spatial dimension, D_b is a linear difference operator along the spectral dimension, and α, β , and γ are the weights that control the proportion to horizontal, vertical and spectral correlation. And in this paper $\alpha = \beta = \gamma = 1$.

2.3 ASSTVSC-Based Anomaly Detection Model for HSIs

In this paper, we integrate the sparse constraint and ASSTV regularization into a unit framework and propose a novel anomaly detection method, which can be written as:

$$(\tilde{\mathbf{X}}, \tilde{\mathbf{A}}) = \arg \min_{\tilde{\mathbf{X}}, \tilde{\mathbf{A}}} \lambda \text{ASSTV}(\tilde{\mathbf{X}}) + \frac{1}{2} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{X}} - \tilde{\mathbf{A}}\|_F^2 \quad (3)$$

$$\text{s.t. } \text{card}(\tilde{\mathbf{A}}) \leq kN$$

where λ is the scalar parameter trading off between the decomposition error and regularization term, kN is the upper bound of the cardinality of $\tilde{\mathbf{A}}$. The cardinality k of $\tilde{\mathbf{A}}$ reflects the sparse energy in the image scene, which is usually defined as the l_0 -norm of $\tilde{\mathbf{A}}$ (Zhang et al, 2016), and $N=m \times n \times b$. $\|\cdot\|_F$ generally represents the Frobenius norm of matrix and is extended to 3D arrays, which can be formulated as:

$$\|\tilde{\mathbf{P}}\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n \sum_{l=1}^b |P_{i,j,l}|^2 \right)^{\frac{1}{2}} \quad (4)$$

3 PROBLEM OPTIMIZATION

To recover the components of ASSTVSC model, we first convert the 3D array optimization problem into vector optimization problem and then use Split Bregman Method (Goldstein, 2009) and GoDec Method (Zhou and Tao, 2011) to solve the vector optimization problem. Finally, the approach of detecting anomalies with the sparse anomaly component is presented.

3.1 Problem Vectorization

Suppose a hyperspectral cube $\tilde{\mathbf{X}} \in R^{m \times n \times b}$, and $X_{i,j,l}$ denotes a pixel at the spatial location (i, j) in the l th band. Define $\mathbf{x} = \mathbf{V}(\tilde{\mathbf{X}})$ as the vector representation of any cube and $\tilde{\mathbf{X}} = \mathbf{C}(\mathbf{x})$ as the cube representation of any vector. $\mathbf{x} = \mathbf{V}(\tilde{\mathbf{X}})$ means using vector $x \in R^{mn b}$ to represent cube $\tilde{\mathbf{X}} \in R^{m \times n \times b}$ and $\tilde{\mathbf{X}} = \mathbf{C}(\mathbf{x})$ means using cube $\tilde{\mathbf{X}} \in R^{m \times n \times b}$ to represent vector $x \in R^{mn b}$. We use small bold letters for vectors, capital bold letters for matrices and capital bold letters with symbol $\sim$ on top for arrays. $\mathbf{x} = \mathbf{V}(\tilde{\mathbf{X}})$ is represented as follows:

$$x_{i+(j-1)m+(l-1)mn} = X_{i,j,l} \quad (5)$$

$$1 \leq i \leq m, 1 \leq j \leq n, 1 \leq l \leq b$$

This defines a linear mapping $\mathbf{V}$ from $R^{m \times n \times b}$ to $R^{mn b}$. And, the inverse linear mapping $\mathbf{C}$ from $R^{mn b}$ to $R^{m \times n \times b}$, making $\tilde{\mathbf{X}} = \mathbf{C}(\mathbf{x})$, is represented as:

$$X_{i,j,l} = x_{i+(j-1)m+(l-1)mn} \quad (6)$$

$$1 \leq i \leq m, 1 \leq j \leq n, 1 \leq l \leq b$$

$\mathbf{V}$ and $\mathbf{C}$ are respectively called “vectorization mapping” and “cubic mapping” in this paper, and have:

$$\mathbf{V}\mathbf{C} = \mathbf{I}_{R^{mn b}}, \mathbf{C}\mathbf{V} = \mathbf{I}_{R^{m \times n \times b}} \quad (7)$$

where $\mathbf{I}_{R^{mn b}}$ is an identity mapping in $R^{mn b}$ and $\mathbf{I}_{R^{m \times n \times b}}$ is an identity mapping in $R^{m \times n \times b}$.

According to (5), the ASSTV can be converted from (8) to (9):

$$\begin{aligned} \text{ASSTV}(\tilde{\mathbf{X}}) = & \sum_{i=1}^{m-1} \sum_{j=1}^n \sum_{l=1}^b |X_{i+1,j,l} - X_{i,j,l}| \\ & + \sum_{i=1}^m \sum_{j=1}^{n-1} \sum_{l=1}^b |X_{i,j+1,l} - X_{i,j,l}| \\ & + \sum_{i=1}^m \sum_{j=1}^n \sum_{l=1}^{b-1} |X_{i,j,l+1} - X_{i,j,l}| \end{aligned} \quad (8)$$

$$\begin{aligned} \text{ASSTV}(\tilde{\mathbf{X}}) = & \sum_{i=1}^{m-1} \sum_{j=1}^n \sum_{l=1}^b |x_{i+1+(j-1)m+(l-1)mn} - x_{i+(j-1)m+(l-1)mn}| \\ & + \sum_{i=1}^m \sum_{j=1}^{n-1} \sum_{l=1}^b |x_{i+jm+(l-1)mn} - x_{i+(j-1)m+(l-1)mn}| \\ & + \sum_{i=1}^m \sum_{j=1}^n \sum_{l=1}^{b-1} |x_{i+(j-1)m+lm} - x_{i+(j-1)m+(l-1)mn}| \end{aligned} \quad (9)$$

Then, how to represent (9) in a matrix linear transform form becomes the key to vectorize problem (3). Suppose a first-order difference operator ∇ in R^L making $\boldsymbol{\beta} = \nabla \boldsymbol{\alpha}$ where

$$\begin{aligned} \beta_1 &= 0 \\ \beta_i &= \alpha_i - \alpha_{i-1} \quad (2 \leq i \leq L, L \geq 2) \end{aligned} \quad (10)$$

This first-order difference operator ∇ actually defines a linear transformation in R^L and it is easy to get that

$$\nabla_L = \begin{pmatrix} 0 & 0 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & \cdots & \cdots & \cdots & 0 \\ 0 & -1 & 1 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & -1 & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & -1 & 1 \end{pmatrix} \quad (11)$$

where the subscript L in ∇_L indicates the order of the matrix. With the first-order difference operator ∇ , $\mathbf{D}_m, \mathbf{D}_n, \mathbf{D}_b$ are represented as follows:

$$\begin{aligned} \mathbf{D}_m &= \mathbf{E}_b \otimes \mathbf{E}_n \otimes \nabla_m \\ \mathbf{D}_n &= \mathbf{E}_b \otimes \nabla_n \otimes \mathbf{E}_m \\ \mathbf{D}_b &= \nabla_b \otimes \mathbf{E}_n \otimes \mathbf{E}_m \end{aligned} \quad (12)$$

where $\mathbf{E}$ represents the identity matrix, the subscript represents the order of the matrix and $\otimes$ represents the Kronecker product of matrices. Let $\mathbf{D} = [\mathbf{D}_m^T, \mathbf{D}_n^T, \mathbf{D}_b^T]^T$, combining (9) and (12), we can get:

$$\text{ASSTV}(\tilde{\mathbf{X}}) = \|\mathbf{D}\mathbf{x}\|_1 \quad (13)$$

Let $\mathbf{y} = \mathbf{V}(\tilde{\mathbf{Y}}), \mathbf{x} = \mathbf{V}(\tilde{\mathbf{X}}), \mathbf{a} = \mathbf{V}(\tilde{\mathbf{A}})$, the equivalent formulation of (3) becomes as follows:

$$\begin{aligned}
 (\mathbf{x}, \mathbf{a}) = \arg \min_{\mathbf{x}, \mathbf{a}} \lambda \|\mathbf{D}\mathbf{x}\|_1 \\
 + \frac{1}{2} \|\mathbf{y} - \mathbf{x} - \mathbf{a}\|_2^2 \quad (14) \\
 \text{s.t. } \text{card}(\mathbf{a}) \leq kN
 \end{aligned}$$

3.2 Recovery Of The Vectorized Background And Anomaly Components

Since problem (14) is quite difficult to solve directly, we use Split Bregman Method and GoDec Method to recover the background and anomaly components. Problem (14) can be optimized by alternately solving the following two subproblems: updating the background component $\mathbf{x}$ and updating the anomaly component $\mathbf{a}$.

The subproblem for updating the background component $\mathbf{x}$ is formulated as:

$$\min_{\mathbf{x}} \lambda \|\mathbf{D}\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{y} - \mathbf{x} - \mathbf{a}\|_2^2 \quad (15)$$

Split Bregman Method (Goldstein, 2009) is adopted to update the background component $\mathbf{x}$. According to Split Bregman Method, two auxiliary variables $\mathbf{d}$ and $\mathbf{b}$ need to be introduced into this subproblem optimization and added to (15) in the following way:

$$\begin{aligned}
 (\mathbf{x}^{w+1}, \mathbf{d}^{w+1}) = \arg \min_{\mathbf{x}, \mathbf{d}} \lambda \|\mathbf{d}\|_1 \\
 + \frac{1}{2} \|\mathbf{y} - \mathbf{x} - \mathbf{a}\|_2^2 \quad (16) \\
 + \frac{1}{2} \|\mathbf{d} - \mathbf{D}\mathbf{x} - \mathbf{b}^w\|_2^2 \\
 \mathbf{b}^{w+1} = \mathbf{b}^w + \mathbf{D}\mathbf{x}^{w+1} - \mathbf{d}^w \quad (17)
 \end{aligned}$$

The optimization of problem (16) can be finished by alternately solving the following two subproblems:

$$\begin{aligned}
 \mathbf{x}^{w+1} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{x} - \mathbf{a}\|_2^2 \\
 + \frac{1}{2} \|\mathbf{d}^w - \mathbf{D}\mathbf{x} - \mathbf{b}^w\|_2^2 \quad (18) \\
 \mathbf{d}^{w+1} = \arg \min_{\mathbf{d}} \lambda \|\mathbf{d}\|_1 \\
 + \frac{1}{2} \|\mathbf{d} - \mathbf{D}\mathbf{x}^{w+1} - \mathbf{b}^w\|_2^2 \quad (19)
 \end{aligned}$$

And the solutions to problem (18) and (19) are respectively:

$$\mathbf{x}^{w+1} = (\mathbf{E} + \mathbf{D}^T \mathbf{D})^{-1} (\mathbf{y} - \mathbf{a} + \mathbf{D}^T (\mathbf{d}^w - \mathbf{b}^w)) \quad (20)$$

$$\mathbf{d}^{w+1} = \text{shrink}(\mathbf{D}\mathbf{x}^{w+1} + \mathbf{b}^w, \lambda) \quad (21)$$

Where

$$\text{shrink}(x, \gamma) = \text{sign}(x) * \max(|x| - \gamma, 0) \quad (22)$$

However, it is extremely difficult to acquire the inverse matrix of $\mathbf{E} + \mathbf{D}^T \mathbf{D}$ for its high dimension. Actually, due to $\mathbf{E} + \mathbf{D}^T \mathbf{D}$ is a sparse matrix, we use Conjugate Gradients Squared (CGS) (Sonneveld, 1989) algorithm to solve the following sparse linear equation to obtain $\mathbf{x}^{w+1}$:

$$(\mathbf{E} + \mathbf{D}^T \mathbf{D})\mathbf{x}^{w+1} = (\mathbf{y} - \mathbf{a} + \mathbf{D}^T (\mathbf{d}^w - \mathbf{b}^w)) \quad (23)$$

By alternately updating $\mathbf{x}$ and $\mathbf{d}$, the subproblem for updating the background component $\mathbf{x}$ can be solved.

The subproblem for updating the anomaly component $\mathbf{a}$ is formulated as:

$$\begin{aligned}
 \min_{\mathbf{a}} \frac{1}{2} \|\mathbf{y} - \mathbf{x} - \mathbf{a}\|_2^2 \quad (24) \\
 \text{s.t. } \text{card}(\mathbf{a}) \leq kN
 \end{aligned}$$

We adopt part of GoDec Method to update anomaly component $\mathbf{a}$, which can be solved via entry-wise hard thresholding of $\mathbf{y} - \mathbf{x}$ (Zhou and Tao, 2011):

$$\begin{aligned}
 \mathbf{a} = \mathbf{P}_{\Omega}(\mathbf{y} - \mathbf{x}), \Omega: |(y - x)_{i \in \Omega}| \neq 0 \\
 \text{and } \geq |(y - x)_{i \in \bar{\Omega}}|, |\Omega| \leq kN \quad (25)
 \end{aligned}$$

where Ω means the nonzero subset of the first kN largest entries of $|\mathbf{y} - \mathbf{x}|$.

By alternately updating the background subproblem and the anomaly subproblem, problem (14) can be optimized. And the background component $\mathbf{x}$ and the anomaly component $\mathbf{a}$ are then recovered.

3.3 Detection Of Anomalies With The Anomaly Component

By mapping anomaly component $\mathbf{a}$ and the background component $\mathbf{x}$ into the cube, the anomaly component $\tilde{\mathbf{A}}$ and the background component $\tilde{\mathbf{X}}$ are recovered:

$$\begin{aligned}
 \tilde{\mathbf{A}} = \mathbf{C}(\mathbf{a}) \\
 \tilde{\mathbf{X}} = \mathbf{C}(\mathbf{x}) \quad (26)
 \end{aligned}$$

$\tilde{\mathbf{A}}$ is a hyperspectral cube with structured sparsity, and the nonzero entries in each band indicate the position of anomalies. According to the sparsity structure of $\tilde{\mathbf{A}}$, the final detection result $\mathbf{S}$ is defined as follows:

$$S_{i,j} = \sum_{l=1}^b |A_{i,j,l}| \quad (1 \leq i \leq m, 1 \leq j \leq n) \quad (27)$$

4 EXPERIMENTAL RESULTS AND ANALYSIS

4.1 HYPERSPECTRAL DATASETS DESCRIPTION

In this section, three real hyperspectral images data are used to validate the effectiveness of our proposed algorithm. The first hyperspectral data, named PaviaC, were collected by the Reflective Optics System Imaging Spectrometer sensor. In this experiment, a small subset is segmented from the original image, with 108×120 pixels and 102 bands after removing low signal-to-noise ratio (SNR) bands, as shown in figure 1(a). Some vehicles on the bridge and bare soil are regarded as anomalies, occupying 43 pixels and accounting for 0.33% of the image scene, as shown in figure 1(b).

The second and the third hyperspectral data come from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over San Diego, CA, USA. The whole image has a 3.5m resolution with 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, 189 available bands of the data were retained in our experiments. A region with size of 120×120 pixels in the upper-left corner is selected as the second hyperspectral data and is called upper-left San Diego data, thereafter, as shown in figure 2(a). Three planes in the image are selected as anomalies, which consist of 90 pixels and account for 0.625% of the image, as shown in figure. 2(b). And the third hyperspectral data lie in the central part of the whole image and is named central San Diego data, thereafter, with a size of 100×100 pixels, as shown in figure 3(a). Three planes which consist of 114 pixels and account for 1.15% of the image are regarded as anomalies, as shown in figure 3(b).

Figure 1: PaviaC data. (a) band 100 of PaviaC. (b) anomaly location.

Figure 2: Upper-left San Diego data. (a) band 100 of Upper-left San Diego. (b) anomaly location.

Figure 3: Central San Diego data. (a) band 100 of Central San Diego. (b) anomaly location.

4.2 Parameter Analysis and Determination

For ASSTVSC, there are four parameters need to be set: the stopping criteria $\varepsilon_1, \varepsilon_2$, the sparsity parameter k and the regularization parameter λ . In all experiments, $\varepsilon_1 = 10^{-2}$, $\varepsilon_2 = 10^{-2}$. What we need to discuss is the sparsity parameter k and the regularization parameter λ . One thing needs to be clarified in advance, that is, all the raw hyperspectral data are normalized before experiments to make the selected parameters more universal.

Referring to the most recent related publications (Zhang et al, 2016; Cui et al, 2014; Sun et al, 2014), the parameter k is set according to the ratio of the anomalies

in the image scene (RAI). To test whether this prior is suitable for our algorithm, the performance of ASSTVSC with respect to the sparsity parameter k are investigated on the aforementioned three hyperspectral data. Interestingly to find that, when k is two and a half times RAI, the detection performance reaches the best for all the data. Specifically, for PaviaC whose backgrounds are relatively simple, the detection performance achieves the best when k is smaller than two and a half times RAI and stays unchanged when k increases. For two San Diego data whose background are much complex, the detection performance reaches the best only when k reaches two and a half times RAI and drops a little when k increases. A closer analysis may account for this phenomenon. For relatively complex images, there may exist more than one kind of anomalies and a larger k will assure the detection of interested anomalies. For relatively simple data, such as PaviaC, only one kind of anomalies exist, so a smaller k can guarantee a good performance. It reveals that the parameter k may probably not only be related with RAI but also the complexity of image.

The regularization parameter λ can be set in an empirical method. Experiments show parameter λ is relatively robust. For simplicity, in our experiments, we empirically set $\lambda = 0.5$, achieving satisfactory results in all cases.

4.3 Detection Performance

The detection performance of ASSTVSC-based anomaly detection method is evaluated and compared with three state-of-the-art detectors: Global RX (GRX) (Reed and Yu, 1990), LRASR (Xu et al, 2016), and LSMAD (Zhang et al, 2016). Based on the aforementioned experiments, in the following comparative experiments, we set $k = 2.5RAI, \lambda = 0.5$ for our proposed algorithm. For LRASR, there are four parameters need to be set. We adopt the fixed parameters mentioned in (Xu et al, 2016) for the following experiments. For LSMAD, two parameters, respectively the rank r of the low-rank matrix and the cardinality k of the sparse matrix, have to be set in an empirically way. Through abundant experiments, we determine the best parameters for each hyperspectral data, as follows: $k = 0.003, r = 1$ for PaviaC data, $k = 0.004, r = 2$ for upper-left San Diego data, $k = 0.003, r = 2$ for central San Diego data. The experimental results for all the detectors are provided through receiver operating characteristic (ROC) curves, AUC values and 2D plots of the detection statistic results.

Table 1: AUC values by different methods for three hyperspectral data.

AUC	PaviaC	Upper-left San Diego	Central San Diego
GRX	0.9983	0.8043	0.9338
LRASR	0.6019	0.8120	0.9130
LSMAD	0.9998	0.8393	0.8252

ASSTVSC	0.9997	0.9920	0.9878
---------	--------	--------	--------

For PaviaC data, as shown in figure 4 and table 1, ASSTVSC and LSMAD demonstrate a much better performance than the other two detection algorithms, achieving a high detection rate with a low false alarm rate. From figure 5, we find the differences between ASSTVSC and LSMAD are subtle and both can suppress most of the background effectively. By a further observation on figure 4, we find that the GRX outperforms LRASR because LRASR fails to suppress the background, resulting in many false alarms.

Figure 4: ROC curves obtained by different methods on PaviaC data.

Figure 5: Detection results obtained by different methods on PaviaC data: (a) GRX. (b) LRASR. (c) LSMAD. (d) ASSTVSC.

For the upper-left San Diego data, the ROC curves and AUC values are shown in figure 6 and table 1, respectively. It reveals that ASSTVSC performs much better than the other three algorithms. The 2D plots of the obtained detection results are shown in figure 7. The planes are pretty obvious using ASSTVSC accompany with a few kinds of uninterested anomalies, whereas the planes detected using LSMAD and GRX are not obvious with naked eyes. Although the planes detected are evident using LRASR, its inability to suppress the background incurs too many false alarms.

Figure 6: ROC curves obtained by different methods on upper-left San Diego data.

Figure 7: Detection results obtained by different methods on upper-left San Diego data: (a) GRX. (b) LRASR. (c) LSMAD. (d) ASSTVSC.

Figure 8: ROC curves obtained by different methods on central San Diego data.

Figure 9: Detection results obtained by different methods on central San Diego data: (a) GRX. (b) LRASR. (c) LSMAD. (d) ASSTVSC.

As we can see from ROC curves and AUC values in figure 8 and table 1, ASSTVSC still performs the best among the four algorithms for the central San Diego data. From figure 9, we can see that although both ASSTVSC and LSMAD can suppress the background, ASSTVSC can distinguish all the anomalies from the background whereas LSMAD has difficulty in detecting the wings of the planes, thus affecting the performance. LRASR and RX are much better than LSMAD since they can detect all the anomalies.

Based on the aforementioned experimental results, we find that ASSTVSC performs much better than GRX and LRASR for all the data, since ASSTVSC can suppress the background effectively and remove the noise from the anomalies. This can be explained as follows: the proposed ASSTVSC method has divided the HSI data into three parts: the background, the anomaly and the noise. This guarantees that the detected anomalies do not be bothered by the background and the noise. Both LSMAD and ASSTVSC can achieve satisfactory detection performance on PaviaC data and have nearly no distinct differences as they can suppress the background and separate the noise from the anomalies. But for relatively complex scenes, LSMAD sometimes has difficulty in detecting all the anomalies. However, our proposed method does not be bothered by this problem and it can achieve good performance regardless of the complexity of images. The reasons may be as follows: the ASSTV model in our algorithm not only explores the spectral correlation but also the spatial correlation whereas the low-rank property of background used in LSMAD only considers the spectral correlation. So some marginal pixels of the anomalies, such as the wing of planes in central San Diego data, which cannot be detected using LSMAD, can easily be detected using ASSTVSC.

5 CONCLUSIONS

In this paper, ASSTVSC has been proposed from a novel perspective to employ total variation model for hyperspectral anomaly detection. Different from the low-

rank-based models which only explore the spectral correlation of HSIs, we use ASSTV model to exploit the spatial correlation and spectral correlation of HSIs. To the best of our knowledge, it is the first time that the ASSTV model is adopted for HSIs anomaly detection. Besides, the sparse property of the anomaly is utilized and experiments reveal that the sparsity parameter k is not only related with the ratio of the anomalies in image but also the complexity of image. Experiments on hyperspectral datasets indicate that ASSTVSC has a superior performance in suppressing background and detecting anomalies than state-of-the-art anomaly detection methods.

But, there are some aspects that need a further research. In our proposed algorithm, parameter k needs to be determined as a prior. Although we have pointed that k is related with RAI and $k = 2.5RAI$ can achieve satisfactory results for all data, it is actually hard to acquire RAI for different data as we usually don't know the size of anomalies. So the selection of parameter k becomes a limitation in reality application. Besides, the runtime of ASSTVSC needs to be optimized, which will be the focus of our future work.

ACKNOWLEDGEMENTS

The authors would like to thank the handling editor and anonymous reviewers for their careful reading, insightful and constructive remarks. They would also like to thank the researchers for providing the hyperspectral datasets.

REFERENCES

1. Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., and Scheunders, P., 2013. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience & Remote Sensing Magazine* 1(2):6-36.
2. Stein, D. W. J., Beaven, S. G., Hoff, L. E., and Winter, E. M., 2002. Anomaly detection from hyperspectral imagery. *Signal Processing Magazine IEEE* 19(1):58-69.
3. Matteoli, S., Diani, M., and Corsini, G., 2010. A tutorial overview of anomaly detection in hyperspectral images. *Aerospace & Electronic Systems Magazine IEEE* 25(7):5-28.
4. Matteoli, S., Veracini, T., Diani, M., and Corsini, G., 2014. Background density nonparametric estimation with data-adaptive bandwidths for the detection of anomalies in multi-hyperspectral imagery. *IEEE Geoscience & Remote Sensing Letters* 11(1):163-167.
5. Reed, I. S., and Yu, X., 1990. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Transactions on Acoustics Speech & Signal Processing* 38(10):1760-1770.
6. Billor, N., Hadi, A. S., and Velleman, P. F., 2000. BACON: blocked adaptive computationally efficient outlier nominators. *Computational statistics & data analysis* 34(3):279-298.

7. Du, B., and Zhang, L., 2011. Random-selection-based anomaly detector for hyperspectral imagery. *IEEE Transactions on Geoscience & Remote Sensing* 49(5):1578-1589.
8. Guo, Q., Zhang, B., Ran, Q., Gao, L., Li, J., and Plaza, A., 2014. Weighted-RXD and linear filter-based RXD: improving background statistics estimation for anomaly detection in hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 7(6):2351-2366.
9. Carlotto, M. J., 2005. A cluster-based approach for detecting man-made objects and changes in imagery. *IEEE Transactions on Geoscience & Remote Sensing* 43(2):374-387.
10. Chang, C. I., 2005. Orthogonal subspace projection (OSP) revisited: a comprehensive study and analysis. *IEEE Transactions on Geoscience & Remote Sensing* 43(3):502-518.
11. Zhang, H., He, W., Zhang, L., Shen, H., and Yuan, Q., 2014. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Transactions on Geoscience & Remote Sensing* 52(8):4729-4743.
12. Cui, X., Tian, Y., Weng, L., and Yang, Y., 2014. Anomaly detection in hyperspectral imagery based on low-rank and sparse decomposition. *Proceedings of SPIE - The International Society for Optical Engineering* 9069(4).
13. Chen, S. Y., and Chang, C. 2013. Low-rank decomposition-based anomaly detection. In *SPIE Defense, Security, and Sensing*.
14. Candes, E. J., Li, X., Ma, Y., and Wright, J. 2009. Robust principal component analysis? *J. ACM*, 58(3):11.
15. Zhou, T., and Tao, D., 2011. GoDec: randomized low_rank & sparse matrix decomposition in noisy case. In *International Conference on Machine Learning, ICML, Bellevue, Washington*.
16. Zhang, Y., Du, B., Zhang, L. and Wang, S., 2016. A low-rank and sparse matrix decomposition-based mahalanobis distance method for hyperspectral anomaly detection. *IEEE Transactions on Geoscience & Remote Sensing* 54(3):1376-1389.
17. Sun, W., 2014. Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery. *Journal of Applied Remote Sensing* 8(1):83641.
18. Liu, G., Lin, Z., and Yu, Y., 2010. Robust subspace segmentation by low-rank representation. In *International Conference on Machine Learning*, pp. 663-670.
19. Xu, Y., Wu, Z., Li, J., Plaza, A., and Wei, Z., 2016. Anomaly detection in hyperspectral images based on low-rank and sparse representation. *IEEE Transactions on Geoscience & Remote Sensing* 54(4):1990-2000.
20. Li, W., and Du, Q., 2015. Collaborative representation for hyperspectral anomaly detection. *IEEE Transactions on Geoscience & Remote Sensing* 53(3):1463-1474.
21. Rudin, L. I., Osher, S. and Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. In *Eleventh International Conference of the Center for Nonlinear Studies on Experimental Mathematics*.
22. Sun, L., Zheng, Y., and Jeon, B., 2017. Hyperspectral restoration employing low rank and 3D total variation regularization. In *International Conference on Progress in Informatics and Computing*.
23. Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W., 2006. An iterative regularization method for total variation-based image restoration. *Siam Journal on Multiscale Modeling & Simulation* 4(2):460-489.
24. He, W., Zhang, H., Zhang, L., and Shen, H., 2015. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Transactions on Geoscience & Remote Sensing* 54(1):178-188.
25. Aggarwal, H. K., and Majumdar, A., 2016. Hyperspectral image denoising using spatio-spectral total variation. *IEEE Geoscience & Remote Sensing Letters* 13(3):442-446.
26. Chang, Y., Yan, L., Fang, H., and Luo, C., 2015. Anisotropic spectral-spatial total variation model for multispectral remote sensing image destriping. *IEEE Transactions On Image Processing* 24(6):1852-1866.
27. Wang, Y., Peng, J., Zhao, Q., Leung, Y., Zhao, X., and Meng, D., 2017. Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* pp(99):1-19.
28. Goldstein, T., 2009. The Split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences* 2(2):1-21.
29. Sonneveld, P., 1989. CGS: a fast Lanczos type solver for nonsymmetric linear systems. *Siam J.sci.stat.comput* 10(1):36-52.