Hyperspectral anomaly detection based on spectral–spatial background joint sparse representation

Lili Zhang\textsuperscript{a,b} and Chunhui Zhao\textsuperscript{a}

\textsuperscript{a}Department of Signal and Information Processing, Harbin Engineering University, College of Information and Communication Engineering, Harbin, China; \textsuperscript{b}Department of Electronic Information Engineering, Daqing Normal University, College of Mechanical and Electrical Engineering, Daqing, China

ABSTRACT

In recent years, some algorithms based on sparse representation have been proposed to improve the detection performance for hyperspectral anomaly detection. Among these algorithms, the background joint sparse representation (BJSR) algorithm adaptively selects the most representative background bases for the local region and can obtain satisfactory results. However, BJSR mainly considers spectral characteristics of hyperspectral images. In this paper, we propose a BJSR-based spectral–spatial method. BJSR is first employed to process the original hyperspectral image in spectral domain. Then, linear local tangent space alignment (LLTSA) is used to obtain the low-dimensional manifold of the hyperspectral image. Next, spatial BJSR is used to process the low-dimensional manifold obtained by LLTSA. Finally, the proposed algorithm combines spectral BJSR with spatial BJSR to detect the anomaly targets. The experimental results demonstrate that the proposed algorithm can achieve a better performance when compared with the comparison algorithms.

Introduction

Hyperspectral image (HSI) is rich in spectral information. It can span the visible, near-infrared and mid-infrared portions of the spectrum in many contiguous spectral bands, and the spectral resolution is usually less than 10 nm (Zhao, Du, & Zhang, 2014). Based on their spectral signatures, it is possible to discriminate different objects of our interest, and classification and target detection are two major applications of the discriminative capability (Camps-Valls & Bruzzone, 2005; Matteoli, Diani, & Corsini, 2010). Target detection is essentially a binary classification problem, which aims to separate specific target pixels from various backgrounds, and it attracts more attention because of many military and civilian applications in recent years (Bajorski, 2012; Manolakis & Shaw, 2002; Shaw & Manolakis, 2002). Anomaly detection (AD) is a special case of target detection, in which no a priori knowledge about the spectra of the targets of interest is available (Matteoli et al., 2010; Stein et al., 2002). In fact, AD generally has more extensive applications than target detection with a priori knowledge because it is usually very difficult to obtain the true spectral information of objects of interest.

In the past few decades, many hyperspectral AD algorithms have been proposed in the literature. Among these algorithms, the RX algorithm proposed by Reed and Yu (Reed & Yu, 1990) is acknowledged to be a benchmark AD algorithm (Li, Zhang, Zhang, & Ma, 2015). A sliding dual-window version of Reed–Xiaoli (RX) (Kwon & Nasrabadi, 2005) algorithm which is a well-known kernel-based AD algorithm is proposed. KRX assumes a Gaussian model in a higher-dimensional feature space, the so-called “Hilbert space” (Kwon & Nasrabadi, 2007). In addition to the aforementioned AD algorithms, some sparse representation-based algorithms (Chen, Nasrabadi, & Tran, 2011; Li & Du, 2015; Li et al., 2015) without any distribution hypothesis for the background have attracted attention recently, and they assume that hyperspectral signatures can be compactly represented by only a few nonzero coefficients in a certain dictionary. The algorithm proposed by Chen et al. (2011) is based on the concept that a pixel in HSI lies in a low-dimensional subspace and thus can be represented as a sparse linear combination of the training samples. The sparse representation of a test sample can be recovered by solving an $l_0$-norm minimization...
problem, and if the solution is sufficiently sparse, the problem can be relaxed to a linear programming problem which can be solved efficiently. Then, the class of the test sample can be determined by the characteristics of the sparse vector on reconstruction. Collaborative-representation-based detector (CRD) (Li & Du, 2015) exploits the concept that each pixel in the background can be approximately represented by its spatial neighborhoods, while anomalies cannot. In CRD, the representation is hypothesized to be the linear combination of neighboring pixels, and the collaboration of representation is reinforced by $\ell_2$-norm minimization of the representation weight vector. Kernel CRD which is the kernel version of CRD is also proposed and further improves the detection performance. In addition, Li et al. (2015) propose an HSI AD model using background joint sparse representation (BJRSR), and the detection result is satisfactory. BJRSR also adopts the sliding dual window as LRX and CRD. The center point of the dual window is the test point, and the pixels between the inner and outer windows are the local neighboring pixels of the test point. In addition, a local search window which is also centered at the test pixel constructs the dictionary. If the test point is an anomaly, the sub-dictionary set simultaneously representing the local neighboring pixels of the test point cannot represent the test point; otherwise, if the test point is not an anomaly, it can be represented by the sub-dictionary set. The detection result of the BJRS algorithm is judged by measuring the length of the matched projection on the orthogonal complementary background subspace that is estimated by the joint sparse representation.

All the aforementioned algorithms are mainly based on spectral information of HSI. In fact, with the development of remote sensing technology, the hyperspectral airborne sensors can also obtain the high spatial resolution, and HSI processing algorithms should exploit both the spatial and the spectral features (Plaza et al., 2009; Plaza, Martin, Plaza, Zortea, & Sanchez, 2011). In recent years, some algorithms exploiting both the spatial and the spectral information perform well for hyperspectral AD. Local sparsity divergence (LSD) (Yuan, Sun, Ji, & Li, 2014) detector is based on the fact that a target sample cannot be approximately represented by very few background samples from the local surrounding both in the spectral and in the spatial domains and gets better detection performance. However, in spatial domain, it is very time consuming to compute the LSD map at each spectral band separately. Sparsity divergence index based on locally linear embedding (SDI-LLE) (Zhang & Zhao, 2016) employs the kernel CRD both in spectral and spatial domain and in spatial domain; kernel CRD is based on the low-dimensional manifold obtained by locally linear embedding (LLE) dimensionality reduction (DR) algorithm. SDI-LLE obtains better detection results.

Although DR is lossy, it can remove interband spectral redundancy and ever present noise and optimize detection results as a result of the improved separation between anomaly and background signatures, so DR is usually employed as a preprocessing method for hyperspectral data processing (Ma, Crawford, & Tian, 2010). Among DR algorithms, principal component analysis (PCA) is the most representative one, but it as a linear projection method cannot exploit the nonlinear properties that are often intrinsic in hyperspectral data (Bachmann, Ainsworth, & Fusina, 2005). In recent years, more and more nonlinear methods based on manifold learning have been proposed. In 2000, the Isomap (isometric feature mapping) algorithm (Tenenbaum, Silva, & Langford, 2000) and LLE algorithm (Roweis & Saul, 2000) were proposed in the same issue of Science (published by the American Association for the Advancement of Science). Then, some other manifold learning algorithms appeared, such as laplacian eigenmaps (LE) (Belkin & Niyogi, 2003), local tangent space alignment (LTSA) (Zhang & Zha, 2004) and stochastic neighbor embedding (Hinton & Roweis, 2010). Among these nonlinear manifold learning algorithms, LTSA has received wide attention due to its simple geometric intuitions, straightforward implementation and global optimization (Wang, 2008). However, these nonlinear manifold learning algorithms are implemented restrictedly on the training sets and cannot show explicit maps on new testing data points (Zhang, Yang, Zhao, & Ge, 2007). The linearization versions of these manifold learning algorithms are then proposed. For example, neighborhood preserving embedding (He, Cai, Yan, & Zhang, 2005) and locality preserving projections (He, 2005) correspond to the linearization of LLE and LE, respectively. Linear local tangent space alignment (LLTSA) (Zhang et al., 2007) is the linear version of LTSA. LLTSA uses the tangent space in the neighborhood of a data point to represent the local geometry and then aligns those local tangent spaces in the low-dimensional space which is linearly mapped from the raw high-dimensional space. However, both LTSA and LLST are unsupervised learning methods and cannot deal with the sparse and non-uniformly distributed data set very well. In this paper, the alignment matrix in LLTSA is first used as a judgment to determine the anomaly regions, which are removed to obtain reliable background data. The background data are then employed as training data, and the low-dimensional manifold of the reliable background and the linear mapping matrix can be obtained by LLTSA. Finally, the low-dimensional manifold of the whole HSI can be obtained by the linear mapping.

In this paper, an algorithm based on spectral–spatial background joint sparse representation (SSBJSR) and LLTSA (LLTSA-SSBJSR) is proposed for hyperspectral
AD. In the proposed LLTSA-SSBJSR algorithm, the original hyperspectral data and the data after DR using LLTSA are employed in spectral and spatial domain, respectively, and the detection result combines spectral BJSR with spatial BJSR. The rest of the paper is organized as follows. First, introduction to LLTSA and SSBJSR is presented, and the formation of the proposed LLTSA-SSBJSR algorithm is described in detail. Next, the experiment data are described, and results obtained are commented and discussed. The final section draws our conclusion.

Related works and methodology

Brief review of LLTSA

A high-dimensional data set \( \mathbf{X} = \{ \mathbf{x}_i \in \mathbb{R}^p, i = 1, \ldots, N \} \) has \( N \) points in \( \mathbb{R}^p \). Then, for DR, LLTSA is employed to find a projection matrix \( \mathbf{A} \) to map the high-dimensional data set \( \mathbf{X} \) into a low-dimensional data set \( \mathbf{Y} = \{ y_i \in \mathbb{R}^d, i = 1, \ldots, N \} \), \( d << p \) as follows:

\[
\mathbf{Y} = \mathbf{A}^T \mathbf{X} \mathbf{H}_N,
\]

where \( T \) means transpose. Given the identity matrix \( \mathbf{I} \) and an \( N \)-dimensional column vector \( \mathbf{e} \) with all ones, we can represent \( \mathbf{H}_N = \mathbf{I} - \mathbf{e} \mathbf{e}^T / N \), where \( \mathbf{e} \mathbf{e}^T \) is a matrix with all ones. In LLTSA, local structure information in the neighborhood of each sample can be represented by the local tangent space. Thus, the local tangent spaces of all samples can be realigned in the global low-dimensional feature space to construct the projection matrix from high-dimensional input space to the low-dimensional feature space.

The implemented steps of LLTSA are as follows (Zhang et al., 2007):

Step 1: Determining the neighborhood.

According to the Euclidean distance between samples, the neighborhood \( \mathbf{X}_k = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k] \) of the sample point \( \mathbf{x}_i \), \( i = 1, \ldots, N \), is constructed, where \( k \) is the number of the nearest neighbors.

Step 2: Extracting local information.

A local transformation matrix \( \mathbf{Q} \) is used to map the neighborhood \( \mathbf{X}_k \) to local low-dimensional tangent space. Thus, the local structure of the neighborhood of sample point \( \mathbf{x}_j \) can be approximated as:

\[
\arg \min_{\mathbf{x} \in \mathbf{Q}} \sum_{j=1}^{k} \left\| \mathbf{x}_j - (\mathbf{x} + \mathbf{Q} \mathbf{b}_j) \right\|_2^2
= \arg \min_{\mathbf{Q}} \sum_{j=1}^{k} \left\| \mathbf{X}_k - \mathbf{Q} \mathbf{b}_j \right\|_2^2,
\]

where the centering matrix \( \mathbf{H}_k = \mathbf{I} - \mathbf{e} \mathbf{e}^T / k \), \( \mathbf{b}_j \) is the local low-dimensional representation of the nearest neighbor \( \mathbf{x}_j \) and \( \mathbf{Q}_j = [\mathbf{b}_{j1}, \ldots, \mathbf{b}_{jk}] \). \( \mathbf{Q}_j \) is the local low-dimensional representation of \( \mathbf{X}_k \). The optimal \( \mathbf{x} \) can be given by the mean of all \( \mathbf{x}_j \) and the optimal transformation matrix \( \mathbf{Q} \) can be given by the \( d \) eigenvectors of covariance matrix \( (\mathbf{X}_k^T \mathbf{Q}_j \mathbf{X}_k)^T \) corresponding to its \( d \) largest eigenvalues. Thus, \( \mathbf{Q}_j \) can be computed by \( \mathbf{Q}_j = \mathbf{Q}_j^T \mathbf{X}_k \).

Step 3: Constructing alignment matrix.

After local structure extraction, the obtained local tangent spaces of all samples are realigned in the global low-dimensional feature space to get the global low-dimensional representation \( \mathbf{Y} \) of \( \mathbf{X} \). Denote a selection matrix \( \mathbf{S} \) given by \( \mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_N] \), where \( \mathbf{s}_i \) is a 0–1 selecting vector, and the global low-dimensional representation of \( \mathbf{X}_i \) is given by \( \mathbf{Y}_i = \mathbf{Y} \mathbf{s}_i \). The objective function of this step can be transformed into a minimal problem as follows:

\[
\arg \min_{\mathbf{Y}} \sum_{i=1}^{N} \left\| \mathbf{e} \mathbf{e}^T \mathbf{Y}_i - \mathbf{L}_i \mathbf{Q} \right\|_2^2
= \arg \min_{\mathbf{Y}} \left\| \mathbf{Y} \mathbf{s} \right\|_2^2
= \arg \min_{\mathbf{Y}} \text{tr} (\mathbf{Y} \mathbf{S} \mathbf{W} \mathbf{W}^T \mathbf{S}^T \mathbf{Y}^T).
\]

In Equation (3), \( \mathbf{L}_i \) is a global transformation matrix and the optimal \( \mathbf{L}_i \) is given by \( \mathbf{L}_i = \mathbf{Y}_i \mathbf{H}_i \Theta_i^T \), where \( \Theta_i^T \) is the Moore–Penrose generalized inverse of \( \Theta_i \); \( \mathbf{W} = \text{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_N) \) and \( \mathbf{W}_i = \mathbf{H}_i (1 - \Theta_i \Theta_i^T) \). According to Equation (1), Equation (3) can be transformed into the following form:

\[
\begin{aligned}
\arg \min_{\mathbf{Y}} & \left( \text{tr} (\mathbf{Y} \mathbf{S} \mathbf{W} \mathbf{W}^T \mathbf{S}^T \mathbf{Y}) \right) \\
& - \text{tr} (\mathbf{Y} \mathbf{s} \mathbf{W} \mathbf{W}^T \mathbf{S}^T \mathbf{Y}) = \mathbf{L}_d
\end{aligned}
\]

where the alignment matrix \( \mathbf{B} = \mathbf{S} \mathbf{W} \mathbf{W}^T \mathbf{S} \).

Step 4: Computing the maps. The minimization problem mentioned above can be converted into solving a generalized eigenvalue problem as follows:

\[
\mathbf{X} \mathbf{H}_N \mathbf{B}_N \mathbf{X}^T \mathbf{a} = \lambda \mathbf{X} \mathbf{H}_N \mathbf{X}^T \mathbf{a}.
\]

The transformation matrix \( \mathbf{A} \) can be obtained by the \( d \) eigenvectors \( \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d \) corresponding to the eigenvalues \( \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d \) and \( \mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d) \).

Spectral BJSR

In spectral domain, BJSR (Li et al., 2015) is employed. The center point of the dual windows is the test point. The pixels between the inner and outer windows are the local neighboring pixels of the test point. The dictionary is constructed by a local search window centered at the test pixel.

For each test pixel \( \mathbf{s}_i \), the implemented steps of BJSR (Li et al., 2015) are as follows:
Step 1: Construct the neighboring pixel set $S$ and the dictionary $H$.

Step 2: Select the active atom subset $A$ from $H$ via simultaneous orthogonal matching pursuit (SOMP) (Tropp, Gilbert, & Strauss, 2006).

Step 3: Calculate the background orthogonal complementary subspace $P_B^\perp$ as follows:

$$P_B^\perp = I - \tilde{A} \left( A^T \tilde{A} \right)^{-1} A^T.$$  

(6)

Step 4: Calculate the detected value of $s_c$ according to Equation (7).

$$d_{\text{spec}} = \frac{\| P_B^\perp s_i \|^2}{\| \sum_i P_B^\perp s_i \|^2 / n} = \frac{\sum_i (s_i^T P_B^\perp s_i) / n}{\sum_i (s_i^T P_B^\perp s_i) / n}, i = 1, \ldots, n,$$

(7)

where $s_i$ is the $i$th pixel of the local neighboring pixel set $S$. The spectral BJSR detector is modeled as Equation (7).

**Spatial BJSR**

In the $j$th band, the characteristics of the reconstruction error of the local neighboring pixels $\sum_j$ can be denoted as follows:

$$\sum_j X_j - B_j^T \Theta = X_j - B_j \left( B_j^T B_j \right)^{-1} B_j^T X_j$$

$$= \left( I - B_j \left( B_j^T B_j \right)^{-1} B_j^T \right) X_j = P_{B_j}^\perp X_j,$$

(8)

where $X_j$ is the local neighboring pixel set of the test point in the $j$th band and $X_j = [x_{1j}, x_{2j}, \ldots, x_{nj}]$, $B_j$ is the sub-dictionary of the local neighboring pixels in the $j$th band; $\Theta$ is the set of all the sparse coefficient vectors; and $P_{B_j}^\perp$ represents the background orthogonal complementary subspace in the $j$th band. By the use of SOMP, we can obtain $\Theta$, $\sum_j$ and $B_j$, $P_{B_j}^\perp$ can also be calculated.

The spatial BJSR detector can be modeled as follows:

$$d_{\text{spat}} = \frac{\| P_{B_j}^\perp x_{ij} \|^2}{\| \sum_j P_{B_j}^\perp x_{ij} \|^2 / n} = \frac{\sum_j (x_{ij}^T P_{B_j}^\perp x_{ij}) / n}{\sum_j (x_{ij}^T P_{B_j}^\perp x_{ij}) / n},$$

(9)

$$i = 1, \ldots, n,$$

where $x_{ij}$ is the test point in the $j$th band and $x_{ij}$ is the $i$th local neighboring pixel of the test point in the $j$th band.

**Spectral–spatial BJSR**

The joint LSD (Yuan et al., 2014) is based on the fact that a target sample cannot be approximately represented by very few background samples from the local surrounding both in the spectral and in the spatial domains. Similar to joint LSD, SSBJSR is represented as follows:

$$d = ad_{\text{spec}} + (1 - a)d_{\text{spat}}, 0 \leq a \leq 1,$$

(10)

where $d_{\text{spec}}$ is spectral BJSR and $d_{\text{spat}}$ is spatial BJSR. Optimized detection results can be obtained by setting different weighting coefficients.

**Formation of LLTSA-SSBJSR**

In this section, we introduce the proposed LLTSA-SSBJSR, which includes four steps as follows:

1. Normalize the input hyperspectral data set and calculate $d_{\text{spec}}$ according to Equation (7).
2. Obtain the low-dimensional manifold of the normalized hyperspectral data set using LLTSA.

First, anomalous regions are selected via the alignment matrix in LLTSA.

In fact, the alignment matrix in Equation (4) is a sparse matrix, whose nonzero elements correspond to the Euclidean distances between the local neighboring pixels and the average. The anomalous regions can be obtained by the Euclidean distances. If the Euclidean distance sum of the local neighboring pixels and the average is larger than the threshold, there may be anomalous pixels in the regions of the local neighboring pixels.

Then, the anomalous regions with possible anomalous pixels are removed, and the remaining regions are employed as training data to construct the background manifold. This process is not for AD, but to obtain a reliable background data.

Next, LLTSA is used to obtain the low-dimensional manifold of the background data, and the transformation matrix can also be obtained.

Finally, low-dimensional manifold of the whole HSI is computed by the linear mapping.

3. Apply spatial BJSR to the normalized low-dimensional manifold of the whole data and compute $d_{\text{spat}}$ according to Equation (9).
4. According to Equation (10), SSBJSR is computed for the final detection result.

**Experiment data and results**

**Date description**

In this paper, we evaluate the performance of our proposed LLTSA-SSBJSR algorithm using three hyperspectral data sets. The first one is a simulated hyperspectral data set, as shown in Figure 1. The background data set collected by the ROSIS sensor (0.43–0.86 μm with 102 bands) during a flight campaign over Pavia, Northern Italy. The geometric resolution is 1.3 m. The whole data set consists of 105 × 100 pixels and 102 bands. The anomaly targets
are embedded in the different backgrounds. The synthesizing method is implanting the target signal and white noise into the background data set. The hundredth band of the synthetic data (the central wavelength is 0.85 $\mu$m) is shown in Figure 1(a). The anomaly target positions are shown in Figure 1(b). The data set has six anomaly targets, two of which have a pixel size of $4 \times 3$, another two targets have a pixel size of $2 \times 2$ and the last two targets have a pixel size of $3 \times 3$ and $2 \times 4$, respectively. The anomalous pixels are very few in number, and the percentage of anomalous pixels is about 0.47%. Figure 2 shows the spectral ranges of the background and the anomaly targets. From Figure 2, we can see that the spectral ranges of the anomalous pixels present obvious differences from those of the background. Therefore, the synthetic data is suitable for the AD experiments.

The second experiment data is the real hyperspectral data of San Diego collected by AVIRIS. The image data ranges from 400 to 1800 nm and has

Figure 1. The synthetic hyperspectral image of Pavia: (a) the hundredth band; (b) the actual distribution of anomaly targets.

Figure 2. Spectral ranges of the synthetic hyperspectral image of Pavia: (a) spectral ranges of the background; (b) spectral ranges of the anomaly targets.

Figure 3. The real hyperspectral image: (a) the hundredth band; (b) the actual distribution of anomaly targets.
126 bands after removing the bands that corresponding to the water absorption regions and low Signal Noise Ratio (SNR). The size of the image is 400 × 400 pixels, and its section with a size of 60 × 60 pixels is used for the experiment. The hundredth band of the experiment data is shown in Figure 3(a). The three planes are regarded as anomaly targets, and the actual distribution is shown in Figure 3(b).

In order to further prove the superiority of the proposed algorithm, the simulated data with more complex background is employed as the third experiment data, as shown in Figure 4. The background data was collected by the 224-band AVIRIS sensor (0.4–2.5 μm with 224 bands) over Salinas Valley, California, and is characterized by high spatial resolution (3.7-m pixels). The area covered comprises 512 lines by 217 samples. The section data set consisting of 130 × 100 pixels and 110 bands is employed for the experiment. The anomaly targets are embedded in the different backgrounds, and the synthesizing method is implanting the target signal and white noise into the background data set. The hundredth band of the synthetic data (the central wavelength is 1.3 μm) is shown in Figure 4(a). The anomaly target positions are shown in Figure 4(b). The data set has eight anomaly targets, four of which have a pixel size of 2 × 2, another two have a pixel size of 2 × 1 and the last two have a pixel size of 1 × 2 and 3 × 2, respectively. The anomalous pixels are very few in number, and the percentage of anomalous pixels is about 0.22%. As shown in Figure 5, the spectral ranges of the background and the anomaly targets are different. Therefore, this synthetic data is suitable for the AD experiments.

Results
In this paper, LRX based on PCA DR (PCA-LRX), LSD based on PCA DR (PCA-LSD) (Yuan et al., 2014), SDI-LLE (Zhang & Zhao, 2016) and BJSR (Li et al., 2015) are employed as the comparison algorithms. For a fair comparison, the test algorithms

![Figure 4. The synthetic hyperspectral image of Salinas: (a) the hundredth band; (b) the actual distribution of anomaly targets.](image)

![Figure 5. Spectral ranges of the synthetic hyperspectral image of Salinas: (a) spectral ranges of the background; (b) spectral ranges of the anomaly targets.](image)
are all with optimal parameters. Tables 1–4 are showing AUC performance of the comparison algorithms with respect to different parameters for the synthetic HSI of Pavia. For PCA-LRX, as shown in Table 1, the optimal AUC value equal to 0.9806 is obtained when the principal component $P$ is 8 and the optimal dual-window size ($w_{\text{in}}$, $w_{\text{out}}$) is (9,15). For PCA-LSD, as shown in Table 2, the optimal AUC value equal to 0.9532 is obtained when $P = 8$ and ($w_{\text{in}}$, $w_{\text{out}}$) is (5,9).

For SDI-LLE, the number of the nearest neighbor pixels is set to 30 (Zhang & Zhao, 2016), and the optimal AUC value equal to 0.9721 is obtained when the reduced band $d = 2$ and ($w_{\text{in}}$, $w_{\text{out}}$) is (5,11) as shown in Table 3. For BJSR, as shown in Table 4, the optimal AUC value equal to 0.9779 is obtained when the upper bound of the sparsity level $L = 6$, the dictionary size $D = 11$ and ($w_{\text{in}}$, $w_{\text{out}}$) is (5,7).

Table 1. AUC performance of PCA-LRX with respect to different parameters for the synthetic hyperspectral image of Pavia.

| $w_{\text{in}}$ | $w_{\text{out}}$ |
|--------------|----------|
| 9            | 11       |
| $P = 7$      |          |
| 5            | 0.9010   |
| 7            | 0.9203   |
| 9            | 0.9506   |
| 11           | 0.9648   |
| $P = 8$      |          |
| 5            | 0.9306   |
| 7            | 0.9556   |
| 9            | 0.9675   |
| 11           | 0.9768   |
| $P = 9$      |          |
| 5            | 0.9148   |
| 7            | 0.9481   |
| 9            | 0.9592   |
| 11           | 0.9733   |

Note: The AUC value in bold is optimal.

Table 2. AUC performance of PCA-LSD with respect to different parameters for the synthetic hyperspectral image of Pavia.

| $w_{\text{in}}$ | $w_{\text{out}}$ |
|--------------|----------|
| 9            | 11       |
| $P = 7$      |          |
| 3            | 0.4322   |
| 5            | 0.6302   |
| 7            | 0.5779   |
| $P = 8$      |          |
| 3            | 0.9410   |
| 5            | 0.9532   |
| 7            | 0.9459   |
| $P = 9$      |          |
| 3            | 0.6897   |
| 5            | 0.7677   |
| 7            | 0.6268   |

Note: The AUC value in bold is optimal.

Table 3. AUC performance of SDI-LLE with respect to different parameters for the synthetic hyperspectral image of Pavia.

| $w_{\text{in}}$ | $w_{\text{out}}$ |
|--------------|----------|
| 9            | 11       |
| $d = 1$      |          |
| 5            | 0.9693   |
| 7            | 0.97     |
| $d = 2$      |          |
| 5            | 0.9711   |
| 7            | 0.9705   |
| $d = 4$      |          |
| 5            | 0.9532   |
| 7            | 0.9516   |

Note: The AUC value in bold is optimal.

Table 4. AUC performance of BJSR (spectral BJSR) with respect to different parameters for the synthetic hyperspectral image of Pavia.

| $w_{\text{in}}$ | $w_{\text{out}}$ |
|--------------|----------|
| 7            | 9        |
| $L = 5$      |          |
| 5            | 0.6270   |
| 7            | 0.4703   |
| $D = 11$     |          |
| 5            | 0.9706   |
| 7            | 0.7937   |
| $D = 13$     |          |
| 5            | 0.4829   |
| 7            | 0.8729   |
| $D = 15$     |          |
| 5            | 0.4679   |
| 7            | 0.9159   |
| $L = 6$      |          |
| 5            | 0.7147   |
| 7            | 0.9530   |
| $D = 11$     |          |
| 5            | 0.9779   |
| 7            | 0.8826   |
| $D = 13$     |          |
| 5            | 0.5098   |
| 7            | 0.9269   |
| $D = 15$     |          |
| 5            | 0.4862   |
| 7            | 0.9091   |
| $L = 7$      |          |
| 5            | 0.7885   |
| 7            | 0.9344   |
| $D = 11$     |          |
| 5            | 0.5264   |
| 7            | 0.9650   |
| $D = 13$     |          |
| 5            | 0.5446   |
| 7            | 0.9481   |
| $D = 15$     |          |
| 5            | 0.5100   |
| 7            | 0.9031   |
| $L = 7$      |          |
| 5            | 0.7885   |
| 7            | 0.9344   |

Note: The AUC value in bold is optimal.

optimal parameters selection for LLTSA-SSBJSR will be discussed in the next section, and in this section, the optimal parameters are given. The optimal AUC value is obtained when $L = 6$, $D$ and ($w_{\text{in}}$, $w_{\text{out}}$) for spectral BJSR are 11 and (5,7), $d = 8$, $D$ and ($w_{\text{in}}$, $w_{\text{out}}$) for spatial BJSR are 17 and (5,13), respectively, and the weighting coefficient $a$ is 0.1.

Tables 5–8 are showing AUC performance of the comparison algorithms with respect to different parameters for the real HSI. For PCA-LRX, as shown in Table 5, the optimal AUC value equal to 0.9983 is obtained when the principal component $P$ is 3 and the optimal dual-window size ($w_{\text{in}}$, $w_{\text{out}}$) is (13,27). For PCA-LSD, as shown in Table 6, the optimal AUC value equal to 0.7796 is obtained when $P = 4$ and ($w_{\text{in}}$, $w_{\text{out}}$) is (11,13). For SDI-LLE (the number of the nearest neighbor pixels is set to 30), as shown in Table 7, the optimal AUC value equal to 0.9945 is obtained when the reduced band $d = 2$ and ($w_{\text{in}}$, $w_{\text{out}}$) is (9,19). For BJSR, as shown in Table 8, the optimal AUC value equal to 0.7637 is obtained when the upper bound of the sparsity level $L = 6$, the dictionary size $D = 21$ and ($w_{\text{in}}$, $w_{\text{out}}$) is (15,17). For LLTSA-
SSBJSR, the optimal AUC value is obtained when $L = 4$, $d = 4$, $D$ and $(w_{in}, w_{out})$ for spectral BJSR are 21 and (15,17), $D$ and $(w_{in}, w_{out})$ for spatial BJSR are 23 and (13,15), respectively, and the weighting coefficient $a$ is 0.

Tables 9–12 are showing the comparison results, with respect to different parameters for the synthetic HSI of Salinas. For PCA-LRX, as shown in Table 9, the optimal AUC value equal to 0.9774 is obtained when the reduced band $d = 8$ and $(w_{in}, w_{out})$ is (3,9). For BJSR, as shown in Table 12, the optimal AUC value equal to 0.8758 is obtained when the upper bound of the sparsity level $L = 5$, the dictionary size $D = 15$ and $(w_{in}, w_{out})$ is (3,7). For LLTSA-SSBJSR, the optimal AUC value is obtained when $L = 5$, $D$ and $(w_{in}, w_{out})$ for spectral BJSR are 15 and (3,7), $d = 6$, $D$ and $(w_{in}, w_{out})$ for spatial BJSR are 15 and (3,5), respectively, and the weighting coefficient $a$ is 0.3.

Figures 6–8 show the binary images of the detection results, and it is very intuitive to see that the pixels is set to 30), as shown in Table 11, the optimal AUC value equal to 0.9774 is obtained when the reduced band $d = 8$ and $(w_{in}, w_{out})$ is (3,9). For BJSR, as shown in Table 12, the optimal AUC value equal to 0.8758 is obtained when the upper bound of the sparsity level $L = 5$, the dictionary size $D = 15$ and $(w_{in}, w_{out})$ is (3,7). For LLTSA-SSBJSR, the optimal AUC value is obtained when $L = 5$, $D$ and $(w_{in}, w_{out})$ for spectral BJSR are 15 and (3,7), $d = 6$, $D$ and $(w_{in}, w_{out})$ for spatial BJSR are 15 and (3,5), respectively, and the weighting coefficient $a$ is 0.3.
Table 9. AUC performance of PCA-LRX with respect to different parameters for the synthetic hyperspectral image of Salinas.

| $w_m$ | $w_{out}$ |
|-------|-----------|
| 5     | 7 9 11    |

$P = 10$
- $5$ : 0.5400 0.9464 0.9916 0.9955
- $7$ : 0.8500 0.836 0.9981 0.9959
- $9$ : 0.9697 0.9965 0.9916 0.9959
- $11$ : 0.990 0.992 0.992 0.992

$P = 11$
- $5$ : 0.5498 0.5623 0.5921 0.5854
- $7$ : 0.703 0.6782 0.7061 0.7269
- $9$ : 0.6889 0.783 0.6934 0.6933
- $11$ : 0.754 0.7401 0.7558 0.6843

$P = 12$
- $5$ : 0.3588 0.3195 0.2618 0.302
- $7$ : 0.2699 0.2131 0.2352 0.156
- $9$ : 0.212 0.3653 0.2325 0.1548

Note: The AUC value in bold is optimal.

Table 10. AUC performance of PCA-LSD with respect to different parameters for the synthetic hyperspectral image of Salinas.

| $w_m$ | $w_{out}$ |
|-------|-----------|
| 5     | 7 9 11    |

$P = 10$
- $5$ : 0.5449 0.54 0.5886 0.5265
- $7$ : 0.5348 0.5208 0.781 0.5777
- $9$ : 0.5548 0.5623 0.5921 0.5854

$P = 11$
- $5$ : 0.703 0.6782 0.7061 0.7269
- $7$ : 0.6889 0.783 0.6934 0.6933
- $9$ : 0.754 0.7401 0.7558 0.6843

$P = 12$
- $5$ : 0.3588 0.3195 0.2618 0.302
- $7$ : 0.2699 0.2131 0.2352 0.156
- $9$ : 0.212 0.3653 0.2325 0.1548

Table 11. AUC performance of SDI-LLE with respect to different parameters for the synthetic hyperspectral image of Salinas.

| $w_m$ | $w_{out}$ |
|-------|-----------|
| 5     | 7 9 11    |

$d = 6$
- $5$ : 0.960 0.9322 0.9721 0.9672
- $7$ : 0.9629 0.957 0.9759 0.9498
- $9$ : 0.9581 0.9694 0.9628

$d = 7$
- $5$ : 0.960 0.9322 0.9721 0.9672
- $7$ : 0.9629 0.957 0.9759 0.9498

$d = 8$
- $5$ : 0.961 0.9531 0.9774 0.9487
- $7$ : 0.9629 0.957 0.9759 0.9498

$d = 9$
- $5$ : 0.9638 0.962 0.9716 0.9652
- $7$ : 0.966 0.9694 0.9611

Note: The AUC value in bold is optimal.

detection result of LLTSA-SSBJSR is satisfactory and LLTSA-SSBJSR achieves better results than PCA-LRD and BJSR. According to the anomaly target positions, as shown in Figures 1(b), 3(b) and 4(b), we can see that the proposed algorithm also performs better than PCA-LRX and SDI-LLE.

The experimental results for all the test algorithms are also provided through receiver operating characteristic (ROC) curves for the first two data sets, as shown in Figure 9(a,b), and the ROC curves of LLTSA-SSBJSR are always above those of the comparison algorithms. We provide the AUC bar graph for the synthetic HSI of Salinas, as shown in Figure 9(c), and the AUC value of LLTSA-SSBJSR is greater than those of the comparison algorithms. In short, the results shown in Figure 9 indicate that the proposed LLTSA-SSBJSR provides a better detection performance than the comparison algorithms.

In addition, separability maps (Lou & Zhao, 2015; Zhang, Du, Zhang, & Wang, 2016; Zhang & Zhao, 2016) evaluating the separability between anomalies and background are employed to further evaluate the performances of the test algorithms, as shown in Figure 10. The lines at the top and bottom of each column are the normalized extreme values which are 0 and 1, and the line in the middle of the box is the mean of the pixels. The position of the boxes reflects the tendency and compactness of the distribution of the pixels, and that is to say, the position reflects the separability of the anomalies and background. For the synthetic HSI of Pavia, as shown in Figure 10(a), we can see that the gap between the anomaly box and the background box for LLTSA-SSBJSR is bigger than for PCA-LRD and SDI-LLE. In addition, LLTSA-SSBJSR can suppress the background information better than PCA-LRD and BJSR. For the real HSI of three planes, as can be seen from Figure 10(b), the gap for LLTSA-SSBJSR is smaller than those of PCA-LRD and SDI-LLE, but LLTSA-SSBJSR can
effectively suppress the background information. For the synthetic HSI of Salinas, as shown in Figure 10(c), we can see that the gap for LLTSA-SSBJSR is much bigger than those of the four comparison algorithms, and LLTSA-SSBJSR can still effectively suppress the background information.
To summarize, for the three experimental data sets, in most cases, separability between anomalies and background of the proposed LLTSA-SSBJSR is not always the best in all the test algorithms, but the ability of LLTSA-SSBJSR to suppress the background information is better than the four comparison algorithms.

From the above analysis, we can conclude that the proposed LLTSA-SSBJSR generally performs better than PCA-LRX, PCA-LSD, SDI-LLE and BJSR for the three experiment data sets.

Furthermore, the corresponding execution times of the test algorithms are listed in Table 13. All experiments were carried out in MATLAB on an Intel Core i7-4770k CPU machine with 16 GB of RAM. The proposed algorithm’s computation mainly comprises two parts: spectral BJSR procedure and spatial BJSR procedure. Given a data set $X \in \mathbb{R}^{N \times B}$ ($N$ is the total number of samples and $B$ is the number of spectral bands). The computational burden of BJSR (spectral BJSR) is $O((nN+B)LBD^2)$ (Li et al., 2015), where

![Figure 10. Separability maps for the three data sets: (a) the synthetic hyperspectral image of Pavia; (b) the real hyperspectral image; (c) the synthetic hyperspectral image of Salinas.](image1)

![Figure 11. AUC performance of LLTSA-SSBJSR with respect to three data sets: (a) the synthetic data set of Pavia, (b) the real data set and (c) the synthetic data set of Salinas.](image2)

| Table 13. Execution times (in seconds) for the detection algorithms. |
|---|---|---|---|---|
| | PCA-LRX | PCA-LSD | SDI-LLE | BJSR | LLTSA-SSBJSR |
| The synthetic data set of Pavia | 9.37 | 21.58 | 1134.87 | 17.79 | 58.02 |
| The real data set | 2.76 | 3.68 | 1413.69 | 6.62 | 25.61 |
| The synthetic data set of Salinas | 4.21 | 44.17 | 4562.86 | 21.46 | 89.13 |

Note: The AUC value in bold is optimal.
Results and discussion

The proposed LLTSA-SSBJSR algorithm consists of two parts, which are spectral BJSR and spatial BJSR. In spectral BJSR, there are three main parameters. The first two are the dual-window size \((w_{\text{in}}, w_{\text{out}})\) and the dictionary size \(D\) and the last one is the upper bound of the sparsity level \(L\). Tables 4, 8 and 12 are also showing the AUC performance of spectral BJSR with respect to different parameters for three experiment data sets. The optimal parameter selections are the same as in BJSR for the three experiment data sets. From Tables 4, 8 and 12, we can see that all the three parameters have obvious influences on the detection results.

In spatial BJSR, there are also three main parameters, which are the dual-window size \((w_{\text{in}}, w_{\text{out}})\), the dictionary size \(D\) and the number of reduced bands \(d\). Tables 14–16 are showing AUC performance of spatial BJSR with respect to different parameters for the three experiment data sets. For the synthetic data set of Pavia, as shown in Table 14, when \(d = 8\), the AUC performance is better. The AUC value is not sensitive to \(D\). The optimal AUC value equal to 0.9915 is obtained when \((w_{\text{in}}, w_{\text{out}})\) is (5,13), (5,15) or (7,13). For the real HSI, as shown in Table 15, the AUC value is not sensitive to \(D\) and the optimal value 0.9992 is obtained when \(d = 4\), \((w_{\text{in}}, w_{\text{out}})\) is (13,15) or (13,17). For the synthetic HSI of Salinas, as shown in Table 16, the optimal AUC value equal to 0.9996 is obtained when \(d = 6\), \((w_{\text{in}}, w_{\text{out}})\) is (3,5) and the value is also insensitive to \(D\). From the above analysis, we can see that \(d\) and \((w_{\text{in}}, w_{\text{out}})\) have obvious influences on the detection results, but the results are not sensitive to \(D\).

The weighting coefficient \(a\) in Equation (10) is also an important parameter that has an obvious impact on the detection results. Figure 11 shows AUC per-

---

**Table 14. AUC performance of spatial BJSR with respect to different parameters for the synthetic hyperspectral image of Pavia.**

| \(d\) | \(D\) | \(w_{\text{in}}\) | \(w_{\text{out}}\) | \(w_{\text{in}}\) | \(w_{\text{out}}\) |
|------|------|----------------|----------------|----------------|----------------|
| 6    | 15   | 7   | 0.9231 0.8879 0.9894 0.9908 | 9   | 0.9868 0.9901 0.9907 0.9915 |
|      |      | 9   | 0.9905 0.9906 0.9915 0.9915 | 11  | 0.9868 0.9901 0.9907 0.9915 |
|      |      | 11  | 0.9868 0.9901 0.9907 0.9915 | 13  | 0.9868 0.9901 0.9907 0.9915 |
|      |      | 13  | 0.9868 0.9901 0.9907 0.9915 | 15  | 0.9868 0.9901 0.9907 0.9915 |

---

**Table 15. AUC performance of spatial BJSR with respect to different parameters for the real hyperspectral image.**

| \(d\) | \(D\) | \(w_{\text{in}}\) | \(w_{\text{out}}\) | \(w_{\text{in}}\) | \(w_{\text{out}}\) |
|------|------|----------------|----------------|----------------|----------------|
| 2    | 21   | 7   | 0.8577 0.8523 0.8426 | 15  | 0.8648 0.8536 0.8401 |
|      |      | 9   | 0.9231 0.9231 0.9231 | 17  | 0.8426 0.8426 0.8426 |
|      |      | 11  | 0.8577 0.8523 0.8426 | 19  | 0.8426 0.8426 0.8426 |
|      |      | 13  | 0.8648 0.8536 0.8401 | 21  | 0.8426 0.8426 0.8426 |
|      |      | 15  | 0.8426 0.8426 0.8426 | 23  | 0.8426 0.8426 0.8426 |

Note: The AUC value in bold is optimal.
formance of LLTSA-SSBJSR with respect to \( a \) for the three experiment data sets, and the other parameters are all optimal. For the synthetic data set of Pavia, as shown in Figure 11(a), the optimal AUC value equal to 0.9915 is obtained when \( a = 0 \) or 0.1, and then the value decreases with the increase of \( a \). When \( a = 1 \), the lowest value is reached. For the real HSI, the optimal AUC value is obtained when \( a = 0 \) and the AUC value is not sensitive to \( a \) when \( a \) changes from 0 to 0.2. Then, the value rapidly decreases with the increase of \( a \), and the lowest value is reached when \( a = 1 \). For the synthetic data set of Salinas, the AUC value is not sensitive to \( a \) when \( a \) changes from 0 to 0.4, and the optimal AUC value equal to 0.9997 is obtained when \( a = 0.1, 0.2 \) or 0.3. Then the AUC value decreases with the increase of \( a \), and the lowest value is reached when \( a = 1 \). From the above analysis, we can see that the optimal detection results are always achieved when \( a < 0.5 \), and that is to say, the spatial BJSR plays a major role in the optimal detection results for the three experiment data sets.

Conclusions

In this paper, a new hyperspectral AD method, called LLTSA-SSBJSR, is proposed. LLTSA-SSBJSR considers the spectral and spatial characteristics and combines the spectral BJSR with the spatial BJSR to improve the detection performance. The proposed algorithm was tested on synthetic and real hyperspectral data. The extensive experimental results show that LLTSA-SSBJSR generally achieves better detection performance than the PCA-LRX, PCA-LSD, SDI-LLE and BJSR algorithms. Parameter selection is also discussed in this paper. LLTSA-SSBJSR still has room for improvement. The parameter selection is mainly determined by experience and repeated experiments, and our future research will focus on how to quickly and effectively determine the optimal parameters.

Acknowledgements

The authors would like to thank the handling editors and the reviewers for providing valuable comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work was supported by the National Natural Science Foundation of China under Grant number 61571145; National Natural Science Foundation of China under

| \( d = 4 \) | \( D = 13 \) | \( w_{in} \) | 5 | 7 | 9 | 11 | 13 | 15 |
|---|---|---|---|---|---|---|---|---|
|\( w_{out} \) | \( w_{out} \) | \( w_{out} \) | \( w_{out} \) | \( w_{out} \) | \( w_{out} \) | \( w_{out} \) | \( w_{out} \) | \( w_{out} \) |
|3 | 0.9956 | 0.9930 | 0.9903 | 0.9876 | – | – | – | – |
|5 | – | 0.9888 | 0.9870 | 0.9850 | – | – | – | – |
|7 | – | – | 0.9844 | 0.9828 | – | – | – | – |
|9 | – | – | – | 0.9804 | – | – | – | – |
|3 | 0.9956 | 0.9930 | 0.9903 | 0.9876 | 0.9857 | – | – | – |
|5 | – | 0.9888 | 0.9870 | 0.9850 | 0.9835 | – | – | – |
|7 | – | – | 0.9844 | 0.9828 | 0.9817 | – | – | – |
|9 | – | – | – | 0.9804 | 0.9794 | – | – | – |
|3 | 0.9956 | 0.9930 | 0.9903 | 0.9876 | 0.9857 | 0.9837 | – | – |
|5 | – | 0.9888 | 0.9870 | 0.9850 | 0.9835 | 0.9818 | – | – |
|7 | – | – | 0.9844 | 0.9828 | 0.9817 | 0.9800 | – | – |
|9 | – | – | – | 0.9804 | 0.9794 | 0.9781 | – | – |
|3 | 0.9996 | 0.9995 | 0.9991 | 0.9985 | – | – | – | – |
|5 | – | 0.9989 | 0.9986 | 0.9978 | – | – | – | – |
|7 | – | – | 0.9979 | 0.9968 | – | – | – | – |
|9 | – | – | – | 0.9953 | – | – | – | – |
|3 | 0.9996 | 0.9995 | 0.9991 | 0.9985 | 0.9972 | – | – | – |
|5 | – | 0.9989 | 0.9986 | 0.9978 | 0.9962 | – | – | – |
|7 | – | – | 0.9979 | 0.9968 | 0.9950 | – | – | – |
|9 | – | – | – | 0.9953 | 0.9934 | 0.9906 | – | – |
|3 | 0.9996 | 0.9995 | 0.9991 | 0.9985 | 0.9972 | 0.9950 | – | – |
|5 | – | 0.9989 | 0.9986 | 0.9978 | 0.9962 | 0.9940 | – | – |
|7 | – | – | 0.9979 | 0.9968 | 0.9950 | 0.9927 | – | – |
|9 | – | – | – | 0.9953 | 0.9934 | 0.9906 | – | – |
|3 | 0.6361 | 0.6761 | 0.6912 | 0.6860 | – | – | – | – |
|5 | – | 0.6830 | 0.6946 | 0.6875 | – | – | – | – |
|7 | – | – | 0.6945 | 0.6845 | – | – | – | – |
|9 | – | – | – | 0.6740 | – | – | – | – |
|3 | 0.6361 | 0.6761 | 0.6912 | 0.6860 | 0.6963 | – | – | – |
|5 | – | 0.6830 | 0.6946 | 0.6875 | 0.6973 | – | – | – |
|7 | – | – | 0.6945 | 0.6845 | 0.6973 | – | – | – |
|9 | – | – | – | 0.6740 | 0.6970 | – | – | – |
|3 | 0.6361 | 0.6761 | 0.6912 | 0.6860 | 0.6963 | 0.6937 | – | – |
|5 | – | 0.6830 | 0.6946 | 0.6875 | 0.6973 | 0.6944 | – | – |
|7 | – | – | 0.6945 | 0.6845 | 0.6973 | 0.6942 | – | – |
|9 | – | – | – | 0.6740 | 0.6970 | 0.6928 | – | – |

Note: The AUC value in bold is optimal.
Grant number 61405041; and Guiding Technology Project of Daqing under Grant number zd-2016-035.

References

Bachmann, C.M., Ainsworth, T.L., & Fusina, R.A. (2005). Exploiting manifold geometry in hyperspectral imagery. *IEEE Transactions on Geoscience & Remote Sensing*, 43(3), 441–454. doi:10.1109/TGRS.2004.842292

Bajorski, P. (2012). Target detection under misspecified models in hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 5(2), 470–477. doi:10.1109/JSTARS.2012.2188095

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6), 1373–1396. doi:10.1162/0899766037641880317

Borghys, D., Käsen, I., Achard, V., & Perneel, C. (2012a). Comparative evaluation of hyperspectral anomaly detectors in different types of background. *Proceedings of SPIE – The International Society for Optical Engineering*, 8390(12), 83902J-83902J-12. doi:10.1117/12.920387

Borghys, D., Käsen, I., Achard, V., & Perneel, C. (2012b). Hyperspectral anomaly detection: Comparative evaluation in scenes with diverse complexity. *Journal of Electrical & Computer Engineering*, 2012, 1–16. doi:10.1155/2012/162106

Campos-Valls, G., & Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience & Remote Sensing*, 43(6), 1351–1362. doi:10.1109/TGRS.2005.846154

Chen, Y., Nasrabadi, N.M., & Tran, T.D. (2011). Sparse representation for target detection in hyperspectral imagery. *IEEE Journal of Selected Topics in Signal Processing*, 5(3), 629–640. doi:10.1109/JSTSP.2011.2113170

He, X. (2005). Locality preserving projections. *Advances in Neural Information Processing Systems*, 45(1), 186–197.

He, X., Cai, D., Yan, S., & Zhang, H.J. (2005). Neighborhood preserving embedding. *Tenth IEEE International Conference on Computer Vision*, 2, 1208–1213. doi:10.1109/ICCV.2005.167

Hinton, G., & Roweis, S. (2010). Stochastic neighbor embedding. *Advances in Neural Information Processing Systems*, 41(4), 833–840.

Kwon, H., & Nasrabadi, N.M. (2005). Kernel RX-algorithm: A nonlinear anomaly detector for hyper-spectral imagery. *IEEE Transactions on Geoscience & Remote Sensing*, 43(2), 388–397. doi:10.1109/TGRS.2004.841487

Kwon, H., & Nasrabadi, N.M. (2007). A comparative analysis of kernel subspace target detectors for hyperspectral imagery. *Journal on Advances in Signal Processing*, 1, 1–13. doi:10.1155/2007/29250

Li, J., Zhang, H., Zhang, L., & Ma, L. (2015). Hyperspectral anomaly detection by the use of background joint sparse representation. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 8(6), 2523–2533. doi:10.1109/JSTARS.2015.2437073

Li, W., & Du, Q. (2015). Collaborative representation for hyperspectral anomaly detection. *IEEE Transactions on Geoscience & Remote Sensing*, 53(3), 1463–1474. doi:10.1109/TGRS.2014.2343955

Lou, C., & Zhao, H. (2015). Local density-based anomaly detection in hyperspectral image. *Journal of Applied Remote Sensing*, 9(1), 095070. doi:10.1117/1.JRS.9.095070

Ma, L., Crawford, M.M., & Tian, J. (2010). Anomaly detection for hyperspectral images based on robust locally linear embedding. *Journal of Infrared Millimeter & Terahertz Waves*, 31(6), 753–762. doi:10.1007/s10762-010-9630-3

Manolakis, D., & Shaw, G. (2002). Detection algorithms for hyperspectral image applications. *IEEE Signal Processing Magazine*, 19(1), 29–43. doi:10.1109/79.974724

Matteoli, S., Diani, M., & Corsini, G. (2010). A tutorial overview of anomaly detection in hyperspectral images. *IEEE Aerospace & Electronic Systems Magazine*, 25(7), 5–28. doi:10.1109/MAES.2010.5546306

Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., … Trianni, G. (2009). Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113(9), S110–S122. doi:10.1016/j.rse.2007.07.028

Plaza, A., Martin, G., Plaza, J., Zortea, M., & Sanchez, S. (2011). Recent developments in endmember extraction and spectral unmixing. *Optical Remote Sensing*, 3, 235–267. doi:10.1007/978-3-642-14212-3_12

Reed, I.S., & 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. doi:10.1109/29.60107

Roweis, S.T., & Saul, L.K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500), 2323–2326. doi:10.1126/science.290.5500.2323

Shaw, G., & Manolakis, D. (2002). Signal processing for hyperspectral image exploitation. *IEEE Signal Processing Magazine*, 19(1), 12–16. doi:10.1109/79.974715

Stein, D.W.J., Beaven, S.G., Hoff, L.E., Winter, E.M., Schaum, A.P., & Stocker, A.D. (2002). Anomaly detection from hyperspectral imagery. *IEEE Signalling Processing Magazine*, 19(1), 58–69. doi:10.1109/79.974730

Tenenbaum, J.B., Silva, V., & Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319–2323. doi:10.1126/science.290.5500.2319

Tropp, J.A., Gilbert, A.C., & Strauss, M.J. (2006). Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3), 572–588. doi:10.1016/j.sigpro.2005.05.030

Wang, J. (2008). Improve local tangent space alignment using various dimensional local coordinates. *Neurocomputing*, 71(16–18), 3575–3581. doi:10.1016/j.neucom.2008.02.008

Yuan, Z., Sun, H., Ji, K., & Li, Z. (2014). Local sparsity divergence for hyperspectral anomaly detection. *IEEE Geoscience & Remote Sensing Letters*, 11(10), 1697–1701. doi:10.1109/LGERS.2014.2360209

Zhang, L., & Zhao, C. (2016). Sparsity divergence index based on locally linear embedding for hyperspectral anomaly detection. *Journal of Applied Remote Sensing*, 10(2), 025026. doi:10.1117/1.JRS.10.025026

Zhang, T., Yang, J., Zhao, D., & Ge, X. (2007). Linear local tangent space alignment and application to face recognition. *Neurocomputing*, 70(7–9), 1547–1553. doi:10.1016/j.neucom.2006.11.007
Zhang, Y., Du, B., Zhang, L., & 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. doi:10.1109/TGRS.2015.2479299

Zhao, R., Du, B., & Zhang, L.P. (2014). A robust nonlinear hyperspectral anomaly detection approach. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 7(4), 1227–1234. doi:10.1109/JSTARS.2014.2311995