Sparsity Score Estimation for Hyperspectral Anomaly Detection

Rui Zhao^a, Xianquan Han^b, Bo Du^{c,*}, Liangpei Zhang^a

^aState Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, P.R. China

^bChangjiang River Scientific Research Institute, Wuhan, China

^cSchool of Computer, Wuhan University, P.R. China

*Corresponding Author: gunspace@163.com

Abstract

Hyperspectral image usually possesses complicated conditions of land-cover distribution, which brings challenge to achieve an effective background representation for hyperspectral anomaly detection. Sparse learning gives a way to overcome this obstacle. A novel sparsity score estimation framework for hyperspectral anomaly detection (SSEAD) is proposed in this paper. Firstly, an overcomplete dictionary and corresponding sparse code matrix are obtained from the hyperspectral data. Then, the frequency of each dictionary atom for reconstruction, which is also called the atom usage probability (AUP), are estimated from the sparse code matrix, from which the strength of each atom for reconstruction is obtained. Finally, the estimated frequencies are transformed to the sparsity score for each pixel. In the proposed detection framework, two operations which aim to enhance the learned dictionary to be more effective for anomaly detection are implemented: 1) dictionary based background feature transformation, and 2) dictionary iterative reweighting. Two real-world hyperspectral datasets are utilized to evaluate the performance of the proposed framework. The experimental results show that the proposed framework achieves superior performance relative to some of the other state-of-the-art anomaly detection methods.

Keywords: hyperspectral, anomaly detection, sparsity score estimation, dictionary enhancement, negative log atom usage probability.

1. Introduction

Hyperspectral imagery spans the visible, near-infrared, and mid-infrared portions of the electromagnetic spectrum

in hundreds of continuous and narrow spectral bands, with a high spectral resolution of generally less than 10 nm ⁽¹⁾. The discriminative spectral information provides hyperspectral data with great potential for either land-cover analysis or man-made object identification. Anomaly detection, which aims at identifying the pixels that are significantly different from the background land covers ⁽²⁾, greatly benefits from this kind of discriminative spectral information.

Anomaly detection needs no prior knowledge about the spectral signature of background and target land-covers and has been applied in many civil or military applications, such as mine detection, search and rescue missions, border surveillance, etc.⁽³⁾. Pixels that have significantly different spectral signatures from their surrounding background are defined as anomaly targets. In most literatures of hyperspectral anomaly detection, for the sake of extruding the anomalies, background estimation is the most important issue ⁽⁴⁾⁻⁽⁶⁾. A variety of anomaly detection algorithms, such as Reed-Xiaoli (RX) detector ⁽⁷⁾, kernel-RX (KRX) ⁽⁸⁾, and the cluster-based anomaly detector (CBAD) ⁽⁹⁾, aim at suppressing the background land-covers via a statistical background estimation. However, these statistical based anomaly detection methods are limited in that they incorporate the anomaly targets in the background estimation when undertaking a global or clustered background statistics estimation. Due to this imperfection, some robust anomaly or outlier detectors, such as the random-selection based anomaly detector (RSAD) ⁽¹⁰⁾, the blocked adaptive computationally efficient outlier nominator (BACON) ⁽¹¹⁾, and the probabilistic anomaly detector (PAD) ⁽¹²⁾ have been developed. These methods utilize different robust statistical strategies to avoid anomaly contamination in the background estimation, but

they usually result in a huge computational burden. Another kind of anomaly detection method, such as the subspace-based RX (SSRX) ⁽¹³⁾ and the selective kernel principal component analysis (KPCA) based anomaly detector ⁽¹⁴⁾, search for a suitable subspace within the hyperspectral data space to extract the representative hyperspectral background features. Nevertheless, the feature extraction based anomaly detectors may find that the complex and redundant spectral mixture ⁽¹⁵⁾ in hyperspectral images makes it difficult for them to extract accurate and appropriate background features for anomaly detection.

The obstacles occurred in statistical based or feature extraction based anomaly detectors are generally derived from the facts that: 1) pixels in hyperspectral images usually possess complexly cluttered spectral signals due to the complicated conditions of land-cover distribution, which brings challenge to achieve an effective background representation to distinguish anomaly targets in the processing of detection, and 2) the statistical based or feature extraction based anomaly detection methods usually explore the background analysis directly with the hyperspectral pixels. Is there a way to construct a hyperspectral anomaly detector that will avoid implementing the background analysis directly from the hyperspectral pixels? The sparse learning technique gives a way to indirectly explore the background and anomaly characteristics with the learned dictionary and corresponding sparse code. Sparse learning ⁽¹⁶⁾ is a powerful tool and has been applied in many signal and image processing applications. The sparse learning theory assumes that most signals can be approximately represented by a linear combination of only a few coefficients (sparse code) over a certain basis or dictionary. And this approximate representation can be solved by compressed sensing and sparse coding techniques. The dictionary obtained in the sparse learning processing is actually a compressive representation of the original signals. One of the advantages of the sparse representation is that there is no explicit assumption for the statistical distribution or feature spanning of the observed data. In this paper, a novel sparsity score estimation framework for hyperspectral anomaly detection (SSEAD), which is inspired by the analysis of the frequency and strength of learned dictionary atoms for hyperspectral image reconstruction, is proposed. In the proposed framework, we aim to indirectly reveal the background and anomaly information from the learned dictionary. For each dictionary atom, respective different

frequency occurs for it to reconstruct the hyperspectral image. And for a hyperspectral pixel, some dictionary atoms play a stronger role than others in generating the pixel vector. From this view of point, we can analyze the background and anomaly characteristics of the dictionary atoms. It is assumed that the frequently used dictionary atoms for reconstruction are more related to the background land-covers and contain more background information. Oppositely, the rarely used dictionary atoms for reconstruction are more related to the anomalies and contain more anomaly information. If a pixel consists of one or more atoms with significant strength that rarely occur in reconstructing other pixels, then it can be seen with a higher anomaly degree than others. In other words, if an atom rarely plays a strong role in reconstructing the hyperspectral image, as captured by its low frequency for reconstruction, its strong involvement in reconstructing a particular pixel renders this pixel an anomaly. Then, an analysis with the combination of the frequency and strength of the learned dictionary atoms for the hyperspectral image reconstruction, is undertaken in our proposed SSEAD framework to reveal the background or anomaly information that each dictionary atom occupies.

The rest of the paper is organized as follows. Section II briefly reviews the sparse learning theory, and then proposes the SSEAD framework for hyperspectral anomaly detection. Section III describes the extensive experiment results of the proposed algorithm. Finally, Section IV draws our conclusions.

2. Proposed Framework

2.1 Sparse Learning Theory

For a hyperspectral image, which is denoted as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{b \times n}$, there are two components crucial for the sparse representation: the overcomplete dictionary and the sparse code. Here, n is the volume of the hyperspectral image, and b is the number of hyperspectral bands. It is assumed that each pixel $\mathbf{x}_i$, $1 \leq i \leq n$ can be approximated by a sparse representation $\mathbf{a}_i = [a_{i1}, a_{i2}, \dots, a_{ik}]^T$ over a dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_k] \in \mathbb{R}^{b \times k}$ of k atoms, demonstrated as:

$$\mathbf{x}_i \approx \mathbf{D}\mathbf{a}_i = a_{i1}\mathbf{d}_1 + a_{i2}\mathbf{d}_2 + \dots + a_{ik}\mathbf{d}_k, \quad (1)$$

here, $\mathbf{d}_i$, $1 \leq i \leq k$ is the i th atom in the dictionary. And the sparse representation admits that $\mathbf{x}_i$ can be approximated as a linear combination of k_0 atoms in $\mathbf{D}$ with $k_0 \ll k$. In other words, only a few entries in $\mathbf{a}_i$ are non-zero. And the sparse vector $\mathbf{a}_i$ can be obtained by solving the optimization problem in the following:

$$\min_{\mathbf{A}, \mathbf{D}} \frac{1}{2} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2 \quad \text{s.t.} \quad \|\mathbf{a}_i\|_0 \leq k_0 \quad \forall i, \quad (2)$$

here, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ is constructed as the sparse code matrix, $\|\cdot\|_0$ denotes the l_0 -norm, the number of non-zeros elements. k_0 is the maximum sparsity level for each $\mathbf{a}_i$ in $\mathbf{A}$. The absolute value of the sparse code $|a_{ji}|$ corresponding to the j th atom $\mathbf{d}_j$ for reconstructing the i th pixel $\mathbf{x}_i$ in the hyperspectral image can be seen as a way to reveal the strength of the j th atom for reconstruction. The optimization problem (2) can be minimized by alternating between $\mathbf{A}$ and $\mathbf{D}$ until reach a stable state⁽¹⁷⁾. This leads to two alternative optimization operations: sparse coding and dictionary update. In sparse coding, given a fixed dictionary $\mathbf{D}$, corresponding sparse code matrix $\mathbf{A}$ can be obtained by the use of the Orthogonal Matching Pursuit (OMP) method⁽¹⁸⁾. In dictionary learning, given the sparse code matrix $\mathbf{A}$, the optimal updated dictionary $\mathbf{D}$ can be obtained by the use of the K-SVD algorithm. Note that, the sparse coding technique is usually much more important than the quality of the learned dictionary⁽¹⁹⁾. Hence, we start the optimization with randomly-chosen pixels from the hyperspectral image as the initial dictionary. In our proposed SSEAD framework, the K-SVD and OMP methods are utilized for dictionary learning and sparse coding.

Fig. 1. The relationship between NLAUP score and the frequency of dictionary atom used for reconstruction.

2.2 Sparsity Score Estimation

After the overcomplete dictionary $\mathbf{D}$ and corresponding sparse code matrix $\mathbf{A}$ have been obtained, an analysis with the combination of the frequency and strength of the learned dictionary atoms for hyperspectral image reconstruction, is undertaken in our proposed SSEAD framework to reveal the background or anomaly information that each dictionary atom occupies.

For a hyperspectral pixel vector $\mathbf{x}_i$, which is the effect of a linear combination of a few atoms in the learned dictionary, some atoms play a stronger role than others in generating the pixel vector. Hence, for each dictionary atom, respective different frequency occurs for it to reconstruct the hyperspectral image. The frequency of the j th dictionary atom $\mathbf{d}_j$ for reconstruction, which is also called its atom usage probability (AUP), can be revealed with an analysis based on its reconstruction strength ($|a_{ji}|$ in the sparse code matrix $\mathbf{A}$). The AUP denoted as p_j for the j th atom $\mathbf{d}_j$ ($j = 1, 2, \dots, k$) is defined as:

$$p_j = \left(\sum_{i=1}^n |a_{ji}| \right) / \sum_{m=1}^k \sum_{i=1}^n |a_{mi}|. \quad (3)$$

From (3), it can be seen that, the AUP will be lower for rarely used atoms, and higher for frequently used ones for reconstruction. If a pixel vector consists of one or more atoms with significant strength that rarely occur in reconstructing other pixels, then it can be seen with a higher

anomaly degree than others. In other words, if an atom rarely plays a strong role in reconstructing the hyperspectral image, as captured by its low frequency for reconstruction, its strong involvement in reconstructing a particular pixel renders this pixel an anomaly. Then, we can define the rarity of a dictionary atom that it is inversely proportional to the frequency of the dictionary atom for reconstruction. The rarity of dictionary atoms is computed by the use of the negative log atom usage probability (NLAUP). The NLAUP of the j th atom $\mathbf{d}_j$, which is denoted as θ_j , is defined as:

$$\theta_j = -\log_2(p_j). \quad (4)$$

NLAUP of a dictionary atom is actually the Shannon information content for the usage of the atom. Use of a rarely-used atom in reconstructing a hyperspectral pixel is improbable and evokes non-routine when it is actually used. Hence, the NLAUP Shannon information is a function of the probability of the atom usage. The smaller the atom's usage, the larger its information content. It is noteworthy that this information content is a proper score of the atom usage, i.e. its objective is to track the true probability distribution of the atom usage. From the relationship between the NLAUP score and the atom usage, the usage characteristic of the learned dictionary will reveal the connection between the dictionary and hyperspectral background/anomaly representation. Fig. 1 illustrates the relationship between the NLAUP score and the frequency of dictionary atom used for reconstruction (AUP). It can be seen that the atom, which occupies low frequency for reconstruction, obtains high NLAUP. This can be defined as a useful manner to identify the background and anomaly characteristics for the learned dictionary atoms. The lower the NLAUP a dictionary atom obtains, the more the background information the atom occupies. On the contrary, the higher the NLAUP a dictionary atom obtains, the more the anomaly information the atom occupies. And here, the background/anomaly information that an atom occupies can be seen as an evaluation index to assess this atom more "background related" or "anomaly related".

After calculating the NLAUP score, it is normalized to convert to a probability distribution function by dividing the NLAUP of each element by the sum of NLAUPs of all the elements:

$$\theta_j^* = \theta_j / \sum_{i=1}^k \theta_i. \quad (5)$$

Then, the normalized NLAUP score vector of the learned dictionary is $\boldsymbol{\theta}^* = [\theta_1^*, \theta_2^*, \dots, \theta_k^*]^T$. Finally, by the association of the background/anomaly information (NLAUP score) that a dictionary atom occupies and the strength of the atom for reconstructing a pixel, the anomaly degree of a hyperspectral pixel can be revealed. Then, with the sparse code matrix $\mathbf{A}$ of the hyperspectral image $\mathbf{X}$ with respect to the learned dictionary $\mathbf{D}$, the sparsity score is computed as (6) to obtain the anomaly degree for each pixel in the hyperspectral image.

$$\delta_{SSEAD}(\mathbf{x}_i) = |\mathbf{a}_i|^T \boldsymbol{\theta}^*. \quad (6)$$

2.3 Dictionary Enhancement

Another analysis with the learned dictionary is also undertaken in our proposed SSEAD framework. This analysis consists of two implementations based on the learned dictionary: 1) dictionary based background feature transformation and 2) dictionary iterative reweighting. These two implementations are both for the sake of adjusting the dictionary to enlarge the diversity between the anomaly and background information in the learned dictionary.

2.3.1 Background Feature Transformation

As aforementioned, the frequently used dictionary atoms for reconstructing the hyperspectral image contains higher background information. Meanwhile, due to the fact that the background land-covers occupies most of the pixels in the hyperspectral image, there would be more background information content that be possessed in the learned dictionary than the anomaly. By the representative features of the learned dictionary, the background information in the dictionary can be represented. Hence, we could find a subspace which is spanned by the representative features of the learned dictionary to suppress the background information in the learned dictionary. The sparse learning process can be transformed to the subspace in our proposed SSEAD framework.

We denote the representative features of the learned dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_k]$ as a series feature vectors

$\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m] \in \mathbb{R}^{b \times m}$, m is the number of representative features and $m \ll k$. And the mean vector $\mathbf{u}_D$ and the covariance matrix Σ_D of the dictionary are

firstly computed, as shown in (7) and (8):

$$\mathbf{u}_D = \left(\sum_{i=1}^k \mathbf{d}_i \right) / k, \quad (7)$$

$$\Sigma_D = \left(\sum_{i=1}^k (\mathbf{d}_i - \mathbf{u}_D)^T (\mathbf{d}_i - \mathbf{u}_D) \right) / (k-1). \quad (8)$$

With the covariance matrix Σ_D of the learned dictionary, the representative features can be obtained by solving the following eigen-decomposition problem:

$$\Sigma_D \mathbf{p}_i = \lambda_i \mathbf{p}_i. \quad (9)$$

Here, λ_i is the i th eigen-value of the covariance matrix Σ_D . The representative features of the learned dictionary can be captured with the first m eigen-vectors corresponding to m maximum eigen-values of the covariance matrix.

After the representative features $\mathbf{P}$ captured, the sparse learning processing will be transformed to:

$$\min_{\mathbf{A}} \frac{1}{2} \sum_{i=1}^n \|\mathbf{P}^T (\mathbf{x}_i - \mathbf{D} \mathbf{a}_i)\|_2 \quad \text{s.t.} \quad \|\mathbf{a}\|_0 \leq k_0. \quad (10)$$

Then, the corresponding sparse matrix $\mathbf{A}$ will be reobtained by the use of the OMP methods. Due to that the representative features represent the background information in the learned dictionary, then the diversity of the background and anomaly information will be enlarged in the transformed subspace spanned by these features.

2.3.2 Dictionary Iterative Reweighting

As aforementioned, it can be concluded that if the total strength $\sum_{i=1}^n |a_{ji}|$ of the j th dictionary atom $\mathbf{d}_j$

for reconstruction is higher than others, then the j th atom

$\mathbf{d}_j$ contains more background information. Due to this fact, we can implement a processing of dictionary iterative reweighting to enhance the NLAUP score analysis, then the diversity of the “background related” and “anomaly related” atoms will be enlarged.

In each iteration of the processing of dictionary iterative reweighting, if we want to enlarge the diversity of the “background related” and “anomaly related” atoms in the learned dictionary, we should enhance the diversity of the NLAUP scores for these two kinds of atoms. That is the

high NLAUP scores should be furthermore highlighted and the low NLAUP scores should furthermore be suppressed. We can implement the processing of dictionary reweighting on the learned dictionary from an aspect of enhancing the diversity of the total strength for reconstruction between the “background related” and “anomaly related” atoms.

If we want to enhance the strength of a dictionary atom for reconstruction, due to the linear combination (1), we should give this atom a small weight. Oppositely, if we want to suppress the strength of a dictionary atom for reconstruction, we should give this atom a large weight. Hence, the normalized NLAUP scores of the learned dictionary $\boldsymbol{\theta}^*$ is an appropriate reweighting vector in the processing of dictionary reweighting. Then, the reweighted dictionary is computed as:

$$\mathbf{D}^r = \boldsymbol{\theta}^{*T} \mathbf{D} = [\theta_1^* \mathbf{d}_1, \theta_2^* \mathbf{d}_2, \dots, \theta_k^* \mathbf{d}_k]. \quad (11)$$

Moreover, in the j th iteration, reweighted dictionary

$\mathbf{D}_{re,j}$ as computed as:

$$\mathbf{D}_j^r = \boldsymbol{\theta}_{j-1}^{*T} \mathbf{D}_{j-1}^r. \quad (12)$$

Note that, $\mathbf{D}_0^r$ is the original learned dictionary

solved from the optimization problem (2), and the $\boldsymbol{\theta}_0^*$ is the corresponding NLAUP score vector computed via (3), (4) and (5). Then by substituting (12) to (2), in the i th iteration, the sparse coding problem will be transformed as:

$$\begin{aligned} & \min_{\mathbf{A}_j} \frac{1}{2} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{D}_j^r \mathbf{a}_i\|_2 \\ & = \min_{\mathbf{A}_j} \frac{1}{2} \sum_{i=1}^n \|\mathbf{x}_i - \boldsymbol{\theta}_{j-1}^{*T} \mathbf{D}_{j-1}^r \mathbf{a}_i\|_2 \quad \text{s.t.} \quad \|\mathbf{a}_i\|_0 \leq k_0 \end{aligned} \quad (13)$$

Furthermore, by combining the representative feature transformed sparse coding problem in (10), the sparse coding problem in (2) in the i th iteration will be rewritten as:

$$\min_{\mathbf{A}} \frac{1}{2} \sum_{i=1}^n \|\mathbf{P}^T (\mathbf{x}_i - \boldsymbol{\theta}_{j-1}^{*T} \mathbf{D}_{j-1}^r \mathbf{a}_i)\|_2 \quad \text{s.t.} \quad \|\mathbf{a}\|_0 \leq k_0. \quad (14)$$

By solving the sparse coding problem in (14) with the OMP method, a new sparse code matrix $\mathbf{A}_j$ will be obtained. And the reweighting vector $\boldsymbol{\theta}_{j-1}^*$ will be updated

to a new one θ_j^* with the new sparse code matrix $\mathbf{A}_j$ via (3), (4) and (5). Then, the final sparsity score δ_{SSEAD} will be obtained with the final θ^* and $\mathbf{A}$.

3. Experiments and Analyses

Hyperspectral datasets from Hyperspectral Digital Imagery Collection Experiment (HYDICE) and Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) remote sensors are utilized in our experiments. The HYDICE dataset, which covers an urban area with an approximately 3-m spatial resolution, is publicly available ⁽²⁰⁾. This HSI consists of 210 spectral channels from 0.4–2.5 μm . After removing the water absorption bands, 162 bands remained. 10 man-made vehicles considered as the anomaly targets as shown in Fig. 2. The AVIRIS dataset covers an urban scene at the location of the airport in San Diego, CA, USA, with a spatial resolution of 3.5m per pixel ⁽²¹⁾. The raw hyperspectral image consists of 224 spectral channels spanning the wavelength range of 0.37–2.51 μm . In our experiment, 189 bands of the raw dataset were retained after removing the band corresponding the low signal-to-noise ratio and water absorption region. The main background materials in this image are ground, road, shadow, and roof. There are 3 aircrafts that consist of 58 pixels and account for 0.58% of the whole image as shown in Fig. 3. These 3 aircrafts are considered as anomaly targets to be detected in this dataset.

Fig. 2. HYDICE real hyperspectral image utilized in our experiments. (a) RGB pseudocolor image (b) Reference.

Fig. 3. AVIRIS real hyperspectral image utilized in our experiments. (a) RGB pseudocolor image (b) Reference.

Seven state-of-the-art methods are compared: (i) Reed-Xiaoli (RX), (ii) cluster-based anomaly detector (CBAD), (iii) subspace-based RX (SSRX), (iv) block adaptive computationally efficient outlier nominators (BACON), (v) probabilistic based anomaly detector (PAD), (vi) kernel RX (KRX), and (vii) collaborative representation-based detector (CRD). In order to quantitatively evaluate the detection performance, receiver operating characteristic (ROC) curves ⁽²²⁾ and background-anomaly separation analysis are used to compare the different methods.

Fig. 4. The ROC curves of the detection methods for the experimental datasets. (a) The HYDICE dataset. (b) The AVIRIS dataset.

Fig. 5. Separability analyses of the detection methods. (a) The HYDICE dataset. (b) The Hyperion dataset.

Fig. 4 shows the ROC curves of the detection methods for the experimental datasets. If the ROC curve indicates that a detection method acquires a higher detection rate than the other methods at the same FAR, it indicates that the detection method outperforms the other methods. It can be seen that, with the empirical parameter values, the proposed SSEAD method achieves better ROC curves than the other detection methods on these two hyperspectral datasets. Fig. 5 shows the separability analyses of the detection methods. It can be seen that the proposed SSEAD method obtains the best background-anomaly separation on these two hyperspectral datasets.

4. Conclusions

With the sparse representation of a learned dictionary and corresponding sparse code matrix, a novel sparsity score estimation framework for hyperspectral anomaly detection (SSEAD) has been proposed in this paper. Our proposed SSEAD framework indirectly reveal the background and anomaly characteristics of the hyperspectral image by the analysis of the frequency and strength of learned dictionary atoms for reconstruction. Furthermore, two dictionary enhancement implementations enlarge the diversity between the anomaly and background

information in the learned dictionary. Extensive experiments with a large amount of hyperspectral image datasets confirmed that the proposed anomaly detection framework outperforms the other state-of-the-art anomaly detection methods. And the empirical parameter ranges were summarized and verified in our experiments. This proves that the proposed SSEAD framework has practical sense of usability.

References

- (1) Rajendran, S., *Hyperspectral Remote Sensing and Spectral Signature Applications*. New India Publishing, 2009
- (2) D. W. J. Stein, S. G. Beaven et al., "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, Vol. 19, No. 1, pp. 58–69, 2002
- (3) S. Matteoli, M. Diani et al., "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, Vol. 25, No. 7, pp. 5–28, 2010
- (4) S. Matteoli, T. Veracini et al., "Models and Methods for Automated Background Density Estimation in Hyperspectral Anomaly Detection," *IEEE Trans. Geosci. Remote Sens.*, Vol. 51, No. 5, pp. 2837–2852, 2013
- (5) R. Zhao, B. Du, and L. Zhang, "A Robust Nonlinear Hyperspectral Anomaly Detection Approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1227–1234, 2014.
- (6) R. Zhao, B. Du et al., "Beyond Background Feature Extraction: An Anomaly Detection Algorithm Inspired by Slowly Varying Signal Analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1–18, 2016.
- (7) I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics Speech Signal Process.*, Vol. 38, No. 10, pp. 1760–1770, 1990
- (8) K. Heesung and N. M. Nasrabadi, "Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, Vol. 43, No. 2, pp. 388–397, 2005
- (9) M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, Vol. 43, No. 2, pp. 374–387, 2005
- (10) B. Du and L. Zhang, "Random-Selection-Based Anomaly Detector for Hyperspectral Imagery," *IEEE*

Trans. Geosci. Remote Sens., Vol. 49, No. 5, pp. 1578–1589, 2011

- (11) N. Billor, A. S. Hadi et al., “BACON: blocked adaptive computationally efficient outlier nominators,” *Comput. Stat. Data An.* Vol. 34, No. 3, pp. 279–298, 2000
- (12) Gao, L., Guo, Q., Plaza, A. J., Li, J., Zhang, B. “Probabilistic anomaly detector for remotely sensed hyperspectral data,” *Journal of Applied Remote Sensing*, Vol. 8, No. 1, 2014
- (13) A. P. Schaum, “Hyperspectral anomaly detection beyond RX,” in *Proc. Int. Soc. Optics and Photonics Defense and Security Symp.*, 2007
- (14) Y. Gu, Y. Liu, and Y. Zhang, “A Selective KPCA Algorithm Based on High-Order Statistics for Anomaly Detection in Hyperspectral Imagery,” *IEEE Geosci. Remote Sens. Lett.*, Vol. 5, No. 1, pp. 43–47, Jan. 2008
- (15) N. Keshava and J. F. Mustard, “Spectral unmixing,” *IEEE Signal Process. Mag.*, Vol. 19, No. 1, pp. 44–57, 2002
- (16) Elad, M, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer Publishing Company, Incorporated, 2010
- (17) Aharon, M., M. Elad, et al, "K-SVD : An algorithm for designing overcomplete dictionaries for sparse representation." *IEEE Transactions on Signal Processing*, Vol. 54, No. 11, pp: 4311-4322, 2006
- (18) Tropp, J. A. and A. C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit." *Information Theory, IEEE Transactions on*, Vol, 53, No. 12, pp: 4655-4666, 2007
- (19) A. Coates and A. Y. Ng, “The importance of encoding versus training with sparse coding and vector quantization,” in *Intl. Conf. Machine Learning*, pp: 921–928, 2011
- (20) US Army Corps of Engineers [Online]. Available: <http://www.tec.army.mil/Hypercurbe>
- (21) Du, B., Zhao, R. and Zhang, L., “A spectral-spatial based local summation anomaly detection method for hyperspectral images,” *Signal Processing*, Vol. 124, pp: 115-131, 2015
- (22) C-I. Chang, “Multiparameter receiver operating characteristic analysis for signal detection and classification,” *IEEE Sensors J.*, Vol. 10, No. 3, pp. 423–442, 2010