Hyperspectral anomaly change detection with slow feature analysis

Chen Wu, Liangpei Zhang, Bo Du

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, PR China
School of Computer, Wuhan University, Wuhan, PR China

Article history:
Received 22 January 2014
Received in revised form 19 June 2014
Accepted 29 September 2014
Communicated by X. Li
Available online 7 November 2014

Keywords:
Anomaly change detection
Slow feature analysis
Hyperspectral image
Multi-temporal images
Hyperion

1. Introduction

Change detection is the process of identifying the differences in the state of an object or phenomenon by observing it at different times [1]. Remote sensing provides a large-scale view and consistent monitoring of landscape conditions over a long period of time [2,3]. It has been demonstrated to be an efficient method for change detection in many applications, such as land-cover/land-use change, city expansion, and ecosystem monitoring [4–7]. In the past two decades, hyperspectral remote sensing technology has advanced significantly and has attracted increasing attention due to the wealth of information in the data and the wide range of potential applications [8–14]. The development of hyperspectral sensors onboard airborne and spaceborne platforms has made it feasible to obtain multi-temporal hyperspectral images covering the same scene, for use in change detection [15].

There are two main topics in hyperspectral change detection. One topic is general change detection, where real landscape transitions are determined by measuring the spectral differences [16–18]. The other topic is anomaly change detection, where small, rare, and anomalous changes are searched for and distinguished from the background, which consists of non-changes and pervasive changes in the scene [19–22]. In this paper, we focus on anomaly change detection. The motivation for anomaly change detection is that the small but anomalous changes, which may be important in practice, could easily be omitted by a human analyst. What anomaly change detection can do is provide a way to highlight the unusual changes that the analyst needs to notice and examine [19].

Eismann et al. [20] proposed a basic diagram for hyperspectral anomaly change detection, which is shown in Fig. 1. The multi-temporal images are used as the input for the predictor to predict the variation in the background. The difference between original image 1 and predicted image 2 is then calculated to obtain the change residuals. Finally, anomaly detection algorithms can be performed on the residual image to detect the anomalous changes. In the basic diagram, the predictor is the key process since it needs to reduce the spectral differences of the background pixels so that the target changes in the residual image are anomalous and can be easily distinguished [20]. One effective predictor is the chronochrome (CC) predictor, which is a global linear predictor [23]. CC predicts changes by least squares linear regression, and the large residuals indicate the anomalous changes. Another conventional predictor is the covariance equalization (CE) predictor, which assumes that the distributions of multi-temporal data after whitening are similar [24]. This method is based on a whitening/dewhitting calculation [25]. In summary, the predictor methods should minimize the spectral differences of the background pixels to highlight the anomalous changes in the change residuals. Meanwhile, they should also extract features from the high-dimension data to improve the performance of the anomaly detection algorithms.

Slow feature analysis (SFA) is a new feature learning algorithm that extracts invariant and slowly varying features from input signals [26,27]. Research found that the function trained from image sequences by slow feature analysis has a good qualitative and quantitative match with the population of complex cells in the
primary visual cortex (V1) [28]. In multi-temporal hyperspectral image pair, the invariant features correspond to the transformed bands that the pixels have very low spectral variance. The SFA algorithm can minimize the spectral difference of multi-temporal pixel pairs and sort the output bands by their importance and effectiveness. Thus, SFA is a potential method of both prediction and feature extraction in hyperspectral anomaly change detection.

In this paper, the SFA-based hyperspectral anomaly change detection algorithm is proposed to effectively distinguish anomalous changes. SFA is applied in multi-temporal hyperspectral images to extract the invariant features. Several of the top transformed difference features are then selected as the input of the RX anomaly detection algorithm, which is an effective and robust anomaly detector to find out anomalous targets from the background. Finally, the anomalous changes are distinguished in the final results.

Section 2 introduces SFA for multi-temporal images and details the process of anomaly change detection. The experimental results and discussions are provided in Section 3, and, finally, the conclusion is drawn in Section 4.

2. SFA for hyperspectral anomaly change detection

Slow feature analysis (SFA) is an unsupervised algorithm that learns functions to extract invariant features from a quickly varying input signal. Mathematically, it can be described as an
Table 1
Maximum AUC values of all the methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>AUC</th>
<th>SFA (7)</th>
<th>Dif</th>
<th>CC</th>
<th>CE</th>
<th>Dif-PCA (4)</th>
<th>CC-PCA (8)</th>
<th>CE-PCA (9)</th>
<th>SegCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFA (7)</td>
<td>0.9995</td>
<td>0.9714</td>
<td>0.9629</td>
<td>0.8641</td>
<td>0.9814</td>
<td>0.9960</td>
<td>0.9796</td>
<td>0.9700</td>
<td></td>
</tr>
<tr>
<td>SegCE</td>
<td>0.8985</td>
<td>0.9934</td>
<td>0.9738</td>
<td>0.9448</td>
<td>0.9980</td>
<td>0.9963</td>
<td>0.9929</td>
<td>0.6775</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Anomaly change detection results of: (a) SFA (7); (b) Dif; (c) CC; (d) CE; (e) Dif-PCA (4); (f) CC-PCA (8); (g) CE-PCA (9); (h) SegCC; (i) SegCE; (j) SegCC-PCA (5); (k) SegCE-PCA (7); (l) stackRX; (m) subHyper; (n) MAD; (o) IRMAD; (p) SAM.
optimization problem: given a multi-dimensional input signal \( s(t) = [s_1(t), \ldots, s_n(t)] \), where \( n \) is the number of the dimensions, we want to find functions \( g_j(s) \) such that the output signal \( z_j(t) = g_j(s(t)) \) satisfies the objective [26]:

\[
\Delta_j = \Delta(z_j) = \left\langle z_j^2 \right\rangle \quad \text{is minimal}
\]  
(1)

under the constraints:

\[
\text{mean zero: } \left\langle z_j \right\rangle = 0 \quad (2)
\]

\[
\text{unit variance: } \left\langle z_j^2 \right\rangle = 1 \quad (3)
\]

where \( \langle \cdot \rangle \) is the mean of the signal over time, and \( z \) is the derivative of \( z \) with respect to time.

SFA has been successfully applied in many applications, such as human action recognition and blind source separation [29, 30]. However, the original SFA algorithm was proposed for continuous temporal signals, whereas multi-temporal remote sensing images are discrete and do not have such a time-series structure. Therefore, SFA theory needs to be improved for the discrete case of multi-temporal images.

Consider two hyperspectral images \( X \) and \( Y \) covering the same scene but acquired at two different times. The corresponding spectral vectors in the same location in the multi-temporal images are presented as \( x' = [x_1', \ldots, x_n'] \) and \( y' = [y_1', \ldots, y_n'] \), where \( i \) denotes the pixel index and \( N \) is the band number. We first centralize and standardize the spectral vectors to zero mean and unit variance, which are expressed as:

\[
x'_j = \frac{x'_j - \mu_{x_j}}{\sigma_{x_j}} \quad \text{and} \quad y'_j = \frac{y'_j - \mu_{y_j}}{\sigma_{y_j}}
\]  
(5)

where \( \mu_{x_j} \) and \( \sigma_{x_j} \) are the mean and standard deviation of band \( j \) of image \( X \). The centralization and standardization are employed for two purposes: satisfying the constraint of SFA and preprocessing for the reduction of radiometric variance.

Based on the fundamental theory of SFA [28,31], we want to find a set of functions \( g_1(x), \ldots, g_N(x) \) to satisfy the optimization objective:

\[
\Delta_j = \frac{1}{P} \sum_{i=1}^{P} (g_j(x'_i) - g_j(y'_i))^2 \quad \text{is minimal}
\]  
(6)

under the constraints:

\[
\frac{1}{2P} \left[ \sum_{i=1}^{P} g_j(x'_i)^2 + \sum_{i=1}^{P} g_j(y'_i)^2 \right] = 1 \quad (7)
\]

\[
\frac{1}{2P} \left[ \sum_{i=1}^{P} g_j(x'_i)^2 + \sum_{i=1}^{P} g_j(y'_i)^2 \right] = 1 \quad (8)
\]

\[
\forall i < j : \quad \frac{1}{2P} \left[ \sum_{i=1}^{P} g_i(x'_i)g_j(x'_i) + \sum_{i=1}^{P} g_i(y'_i)g_j(y'_i) \right] = 0 \quad (9)
\]

where \( P \) is the number of pixels in the image.

The objective (6) of the optimization problem ensures the temporal invariance of the transformed multi-temporal features by minimizing the mean of their squared difference. In anomaly change detection, the objective of the SFA algorithm is to minimize the spectral differences of the background pixels, which include non-changes and pervasive changes. Because the anomalous changes are small and rare, we can assume that the statistical characteristic of the background can be estimated by the whole image. Thus, in hyperspectral anomaly change detection, the objective of the SFA algorithm is to minimize the mean of their squared difference. In anomaly change detection, the objective of the SFA algorithm is to minimize the spectral differences of the background pixels, which include non-changes and pervasive changes. Because the anomalous changes are small and rare, we can assume that the statistical characteristic of the background can be estimated by the whole image. Thus, in hyperspectral anomaly change detection, the whole multi-temporal images are the input of the calculation.

![Fig. 7. ROC curves for the proposed method and the comparative methods.](image)

![Fig. 8. Runtime performance of our proposed method and the comparative methods.](image)
Constraint (7) is presented to normalize all the transformed features to a common scale and simplify the problem, which has been satisfied by centralizing and standardizing the input data by (5). Constraint (8) makes sure that the transformed features must contain some information and avoids a trivial constant solution. Constraint (9) guarantees that each transformed feature is decorrelated from the others and contains different types of change information.

In this problem, when the transformation is linear, the function can be expressed as \( g_j(x) = w_j^T x \), where \( w \) is the transformation vector. Thus, the objective (6) and constraints (8) and (9) can be rewritten as:

\[
\frac{1}{p} \sum_{i=1}^{p} (w_j^T \hat{x}_i - w_j^T \hat{y}_i)^2 = w_j^T \left[ \frac{1}{p} \sum_{i=1}^{p} (\hat{x}_i - \hat{y}_i) (\hat{x}_i - \hat{y}_i)^T \right] w_j = w_j^T A w_j
\]  

(10)

Fig. 9. Binary change map obtained by (a) OTSU and (b) \( k \)-means.

Fig. 10. ROC curves of SFA and SFA with LCRA with three window sizes with the mis-registration range of (a) one pixel, (b) two pixels, (c) five pixels and (d) eight pixels.
\[
\sum_{i=1}^{p} \left( w_i^T x_i \right)^2 + \sum_{i=1}^{p} \left( w_i^T y_i \right)^2
\]
\[
= w_i^T \left[ \frac{1}{2p} \left( \sum_{i=1}^{p} \left( x_i \right)^T + \sum_{i=1}^{p} \left( y_i \right)^T \right) \right] w_i = w_i^T B w_i
\]
(11)

\[
\sum_{i=1}^{p} \left( w_i^T x_i \right)^2 + \sum_{i=1}^{p} \left( w_i^T y_i \right)^2
\]
\[
= w_i^T \left[ \frac{1}{2p} \left( \sum_{i=1}^{p} \left( x_i \right)^T + \sum_{i=1}^{p} \left( y_i \right)^T \right) \right] w_i = w_i^T B w_i
\]
(12)

If we integrate constraint (8) into the objective (6), we can obtain [28]:
\[
\Delta(x_j) = \frac{\left( z_j^2 \right)}{\left( z_j \right)^T} = \frac{1}{2p} \left( \sum_{i=1}^{p} \left( g_i(x_i) - g_i(y_i) \right)^2 \right) = \frac{w_i^T A w_i}{w_j^T B w_j}
\]
(13)

where A and B can be written as:
\[
A = \frac{1}{p} \sum_{i=1}^{p} (x_i - y_i)(x_i - y_i)^T = \Sigma_x
\]
(14)

\[
B = \frac{1}{2p} \left[ \sum_{i=1}^{p} (x_i)^T + \sum_{i=1}^{p} (y_i)^T \right]^2 = \frac{1}{2}(\Sigma_x + \Sigma_y)
\]
(15)

It can be seen that A is the covariance matrix \( \Sigma_x \) of the difference image, and B is the average of the covariance matrix \( \Sigma_x \) of image X and the covariance matrix \( \Sigma_y \) of image Y.

The optimization problem is equal to the generalized problem:
\[
A W = B W A
\]
(16)

where A is a diagonal matrix of the generalized eigenvalues, and W is the corresponding generalized eigenvector matrix.

The transformation vector \( w_j \) can be normalized to fulfill constraints (8) and (9):
\[
\tilde{w}_j = \frac{w_j}{\sqrt{w_j^T B w_j}}
\]
(17)

Unlike the basic diagram in Fig. 1, SFA does not predict an image to get the difference with the other original image. Instead, it transforms both multi-temporal images to obtain the change residuals by the transformed difference, as follows:
\[
r_j = w_j^T x - w_j^T y
\]
(18)

The residual image bands are sorted by their importance according to the SFA algorithm. By (13) and (18), the eigenvalue is equal to the temporal variance \( \Delta_j = \lambda_j \). Thus, the eigenvectors are sorted by the corresponding eigenvalues \( \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N \) to

\[
\begin{array}{c}
\text{AUC values with different numbers of top bands and the corresponding} \\
1/\text{eigenvalue. }
\end{array}
\]

Fig. 12. AUC values with different numbers of top bands and the corresponding 1/eigenvalue.
determine the order of the output bands $\Delta_1 \leq \Delta_2 \leq \cdots \leq \Delta_N$. The more important residual image bands, which contain more anomalous change information, have lower index values. In hyperspectral anomaly change detection, to improve the performance, we select the top $m$ ($m$ is smaller than $N$) residual image bands as the input for the next detection.

The goal of anomaly change detection is to detect temporal anomalous changes in a multi-temporal hyperspectral image pair. After the process of SFA, the background, which consists of non-changes and pervasive changes, will have a very low difference value in the residual image because the difference in the background is minimized by the proposed SFA method. In this way, the anomalous changes are highlighted from the background by the high residuals. The anomaly change detection problem in the image pair is thus transferred to an anomaly detection problem in the residual image. As a result of the rareness of the anomalous changes, the anomaly detection algorithm is more effective than a general distance measurement such as the Euclidean distance.

Generally speaking, anomaly detection can be employed by measuring the similarity between the pixel under test and the background. The RX anomaly detector is a famous and effective anomaly detection method [32]. It measures the similarity with the squared Mahalanobis distance between the pixel under test and the distribution of the background. Since the anomalous changes are rare and small, the mean and the covariance matrix of the distribution of the background can be estimated globally by using all the pixels in the residual image. In this paper, we apply the RX algorithm to detect the anomalous changes from the residual image. The formula of the RX algorithm is expressed as follows [33]:

$$RX(r) = (r - \mu_r)\Sigma^{-1}(r - \mu_r)$$  \hspace{1cm} (19)

where $r$ is the spectral vector of the change residual, $\mu_r$ is the mean, and $\Sigma$ is the covariance matrix of the residual image.

Finally, the detection result obtained by RX on the residual image highlights the presence of anomalous changes in the study scene. The anomalous changes have high values and are well separated from the background. The threshold can be selected manually or automatically to get the binary anomalous change map.

Fig. 2 shows the flowchart of the proposed hyperspectral anomaly change detection with SFA. The multi-temporal images are first centralized and standardized by (5). The transformation matrix is then learned and calculated by (14)–(16). The residual image is obtained by (18). Finally, RX anomaly detection by (19) is applied to the residual image to identify the anomalous changes.

Based on description in this section, the computational complexity of our proposed SFA anomaly detection contains five parts, as follows. First, the centralization and standardization of the multi-temporal images are applied, and the computation complexity of this part is $O(p \times n)$, where $p$ is the number of pixels and $n$ is the number of bands. Second, matrix $A$ and $B$ are computed according to the covariance matrix with a computational complexity of $O(p \times n^2)$. Then, the SFA transformation matrix is obtained by a generalized eigenvalue decomposition task, which has the computational complexity of $O(n^3)$. In the fourth part, we get the change residuals with the transformation of original images, and the computational complexity is $O(p \times n^2)$. Finally, the RX detector is employed and its computational complexity is composed of covariance matrix computation, matrix inversion, and image transformation, which is $O(p \times m^2) + O(m^2 \times m^2)$. Therefore, the total computational complexity of the algorithm is $O(p \times n^2 + n^3 + p \times m^2 + m^2 \times m^2)$. Although it costs time to apply SFA before anomaly detector, it will obviously increase the accuracy for anomaly change detection.

3. Experiments

3.1. Hinggan league Hyperion dataset

For our first experiment, in order to evaluate the performance of the proposed method, we selected multi-temporal Hyperion data covering farmland in Hinggan League, Inner Mongolia, China. These multi-temporal hyperspectral images have a scene size of 400 x 400 and were acquired on April 24, 2010, and June 25, 2010, respectively, as shown in Fig. 3. The level of the Hyperion data is L1Gst. A total of 155 of the 242 bands were selected to avoid the zero-value and noise bands. Accurate manual relative geometric registration was applied to reduce the mis-registration errors. Because the two images were acquired in spring and summer, there are obvious phenological variations in the scene. We extracted five city spectral signatures from other parts of the original Hyperion image on June 25 and embedded them into the study scene, as shown in Fig. 3(b). Each kind of spectral signature was embedded in one row, and each column had different embedding proportions, from 100% to 20%. Compared with the pervasive vegetation changes and non-changes, the embedded pixels are the anomalous changes. The advantage of this simulated dataset is that the anomalous changes are known, and the spectral variance of the non-changes caused by environmental differences and the pervasive phenological variations are realistic.

![Fig. 13. AUC values with different numbers of top bands for SFA, Dif-PCA, CC-PCA, and CE-PCA.](image)

Table 2

<table>
<thead>
<tr>
<th></th>
<th>SFA(5)</th>
<th>Dif</th>
<th>CC</th>
<th>CE</th>
<th>Dif-PCA(2)</th>
<th>CC-PCA(2)</th>
<th>CE-PCA(1)</th>
<th>SegCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.9937</td>
<td>0.9447</td>
<td>0.9078</td>
<td>0.9160</td>
<td>0.9822</td>
<td>0.9895</td>
<td>0.9798</td>
<td>0.7412</td>
</tr>
<tr>
<td></td>
<td>SegCE</td>
<td>SegCC-PCA (2)</td>
<td>SegCC-PCA (20)</td>
<td>stackRX</td>
<td>subHyper</td>
<td>MAD</td>
<td>IRMAD</td>
<td>SAM</td>
</tr>
<tr>
<td>AUC</td>
<td>0.9480</td>
<td>0.9531</td>
<td>0.9786</td>
<td>0.9063</td>
<td>0.9845</td>
<td>0.9898</td>
<td>0.9711</td>
<td>0.9900</td>
</tr>
</tbody>
</table>
Fig. 14. Anomaly change detection results of: (a) SFA (5); (b) Dif; (c) CC; (d) CE; (e) Dif-PCA (2); (f) CC-PCA (2); (g) CE-PCA (1); (h) SegCC; (i) SegCE; (j) SegCC-PCA (2); (k) SegCE-PCA (20); (l) stackRX; (m) subHyper; (n) MAD; (o) IRMAD; (p) SAM.
The eigenvalues in SFA correspond to the content of the information. Because the top bands have smaller eigenvalues, we used the 1/eigenvalue to measure the change information. The area under the curve (AUC) is the area under the receiver operating characteristic (ROC) curve, which indicates the ability of the detection [34]. Fig. 4 shows the AUC values with different numbers of top bands and the 1/eigenvalue of the corresponding top band. The maximum AUC value, which is indicated by the solid green line, was obtained with seven top bands. It can be seen that the seven top bands contain most of the information. Behind the solid green line, the curve of the AUC decreases, which illustrates that the other bands contain little change information and most of the noise. In order to automatically select the number of top bands for anomaly change detection, we recommend a criterion according to the 1/eigenvalue. The information content can be measured by the 1/eigenvalue. Therefore, the bands which have a 1/eigenvalue of greater than 1 can be selected as the input of the RX algorithm. In Fig. 4, the horizontal dashed line indicates the 1/eigenvalue of 1, and the vertical dashed line indicates the AUC value obtained with the eight top bands selected by the criterion. Although the criterion did not lead to the maximum AUC value, it was very close to the best result.

In this paper, the comparative methods include RX anomaly detection on the simple difference image (Dif), on the residual image by CC, and on the residual image by CE [20]. Furthermore, in order to compare the performance of the feature extraction methods, the difference images and the residual images by CC and CE were processed by principal component analysis (PCA). Several of the top bands of PCA were used as the input of RX. Segmentation-based CC and CE were also employed, while the class map was obtained by k-means and the stack dataset [20,35]. In addition, we used the spectral angle mapper (SAM) [36], stack RX detection (stackRX) [37], sub-pixel hyperbolic anomalous change detection (subHyper) [37], MAD, and IRMAD [38–40] for comparison.

Because the detection ability is affected by the number of top feature bands, we compared the performance of the proposed method and the PCA-based state-of-the-art methods with different numbers of top bands. Fig. 5 shows the AUC values of the anomaly change detection results by SFA and three PCA-based state-of-the-art methods. It can be seen that the AUC curve of SFA is always above those of the comparative methods in Fig. 5. This confirms that the proposed method has a better detection ability. Among the three comparative methods, CC with PCA showed a better performance than the other two methods. It is worth noting that the performance of SFA drops when the number of bands is 2. It is because SFA can concentrate the anomalous change information on the top feature bands, whereas there may be a single band containing other non-targeted change information. So the result may be unstable with only 2 or 3 feature bands.

The maximum AUC values of the proposed method and the other state-of-the-art methods are shown in Table 1. The number in parentheses is the number of top bands with which the best AUC value was obtained. The bold number indicates the highest AUC value. In Table 1, the highest AUC value, which is 0.9995, was obtained by SFA. This illustrates that the proposed method outperformed the other methods. All the methods with PCA obtained higher AUC values than those without PCA, which shows that the feature extraction can effectively improve the performance of anomaly change detection. The segmentation-based CC and CE obtained a slightly higher accuracy than the original CC and CE, but they did not perform as well as the original methods with PCA. SubHyper outperformed the other methods, with an AUC value of 0.9980. The AUC value of MAD was higher than IRMAD because IRMAD is more effective in detecting large changes with the iterative reweighting process. SAM only measures the spectral difference between multi-temporal images, so it cannot separate anomalous changes from the pervasive changes.

In this paper, the comparative methods include RX anomaly detection on the simple difference image (Dif), on the residual image by CC, and on the residual image by CE [20]. Furthermore, in order to compare the performance of the feature extraction methods, the difference images and the residual images by CC and CE were processed by principal component analysis (PCA). Several of the top bands of PCA were used as the input of RX. Segmentation-based CC and CE were also employed, while the class map was obtained by k-means and the stack dataset [20,35]. In addition, we used the spectral angle mapper (SAM) [36], stack RX detection (stackRX) [37], sub-pixel hyperbolic anomalous change detection (subHyper) [37], MAD, and IRMAD [38–40] for comparison.

Because the detection ability is affected by the number of top feature bands, we compared the performance of the proposed method and the PCA-based state-of-the-art methods with different numbers of top bands. Fig. 5 shows the AUC values of the anomaly change detection results by SFA and three PCA-based state-of-the-art methods. It can be seen that the AUC curve of SFA is always above those of the comparative methods in Fig. 5. This confirms that the proposed method has a better detection ability. Among the three comparative methods, CC with PCA showed a better performance than the other two methods. It is worth noting that the performance of SFA drops when the number of bands is 2. It is because SFA can concentrate the anomalous change information on the top feature bands, whereas there may be a single band containing other non-targeted change information. So the result may be unstable with only 2 or 3 feature bands.

The maximum AUC values of the proposed method and the other state-of-the-art methods are shown in Table 1. The number in parentheses is the number of top bands with which the best AUC value was obtained. The bold number indicates the highest AUC value. In Table 1, the highest AUC value, which is 0.9995, was obtained by SFA. This illustrates that the proposed method outperformed the other methods. All the methods with PCA obtained higher AUC values than those without PCA, which shows that the feature extraction can effectively improve the performance of anomaly change detection. The segmentation-based CC and CE obtained a slightly higher accuracy than the original CC and CE, but they did not perform as well as the original methods with PCA. SubHyper outperformed the other methods, with an AUC value of 0.9980. The AUC value of MAD was higher than IRMAD because IRMAD is more effective in detecting large changes with the iterative reweighting process. SAM only measures the spectral difference between multi-temporal images, so it cannot separate anomalous changes from the pervasive changes.
Fig. 6 shows the detection results of all the methods, with the maximum AUC values. All the images were stretched by the square root stretch method, according to the reference pixels, with ENVI software. Since the temporal difference of the background is minimized after the process of SFA, the anomalous changes in the hyperspectral pair are transferred to anomalous targets in the residual image. In this way, RX is employed in the proposed method to highlight the anomalous changes from the non-changes and pervasive changes. It can be seen that when the most anomalous changes are detected, the result of the proposed method has few false alarms. The detection results after PCA in Fig. 6(e)–(g) suffer less from the striping noise than those without PCA shown in Fig. 6(b)–(d). SubHyper obtained a better performance with fewer false alarms than the other comparative methods, as shown in Fig. 6(m). Fig. 6(p) shows that SAM could not detect anomalous changes from the pervasive changes in this case.

Fig. 7 shows the ROC curves for SFA with three and seven top bands, along with the other state-of-the-art methods. Among the comparative methods, CC with PCA presents the best performance. The curve of SFA with seven top bands is below that of CC-PCA when the detection rate is not high. However, when the detection rate is higher than 0.5, SFA with seven top bands shows a much lower false alarm rate, which is why SFA with seven top bands obtained the highest AUC value. Although the AUC value of SFA with three top bands is less than that of SFA with seven bands, its ROC curve is always above the other curves, except for that of subHyper. SubHyper shows a very similar performance to the proposed method. Fig. 7 illustrates that the proposed method shows a better detection ability than most of the comparative methods.

The runtime of our proposed method and the comparative methods are shown in Fig. 8. Green bar indicates the runtime of the predictor, such as SFA and PCA, while red bar indicates that of the detector, which is the RX anomaly detector combined with the transformed features. In order to be comparable, the methods which need to select the features are all detected with the top 7 bands. StackRX, subHyperRX and SCD don’t contain predictors. Fig. 8 illustrates that the methods with feature selection can reduce the runtimes of detector, while the predictor may cost some time. Segment based methods are too time-consuming because of their clustering. IRMAD cost much time in the iteration, and SCD runs very slowly since it needs to calculate the subspace projection for every pixel. Although the runtime of our proposed SFA anomaly change detection is longer than some state-of-the-art methods, it is still acceptable.

Threshold is a very important issue to obtain the final binary anomalous change map. In this research, we tested two automatic threshold methods: OTSU [41] and k-means [42]. The input data were the gray-scale anomaly change detection results of SFA with seven output bands. The result is shown in Fig. 9. The detection rate (DR) and false alarm rate (FAR) obtained by OTSU were 0.4800 and 0.0001, respectively. Meanwhile, k-means obtained the results with a DR of 1.0000 and a FAR of 0.1721. This indicates that OTSU tends to get a result with few false alarms but a low DR, while k-means is the opposite.

Co-registration is an essential process for change detection, since the performance is strongly affected by the residual mis-registration error [43,44]. As a pixel-based method, SFA is inevitably influenced by this error. Local co-registration adjustment (LCRA) is an effective approach to deal with this problem and improve the anomaly change detection method [19,36,45,46]. In this research, we evaluated the performances of the proposed method and the improved version with LCRA with different

![Binary change map obtained by (a) OTSU and (b) k-means.](image-url)
window sizes. The experimental data were manually shifted vertically and horizontally by one, two, five and eight pixels. The ROC curves of these four mis-registration experiments are shown in Fig. 10. The number in parentheses indicates the output band number for the highest AUC value. The corresponding AUC value is shown behind the legend. Fig. 10 illustrates that all the methods obtained similar AUC values, since in the simulated dataset, the changes are very anomalous when compared to the other landscape types, and the background situation is comparatively simple. Despite this, for the ROC curves shown in Fig. 10(a) and (b), SFA with LCRA obtained a higher DR than the original SFA when the false alarm rate was not high in the cases when the mis-registration is not very large. In Fig. 10(c) and (d), LCRA didn’t improve the performance obviously and even performed worse. It is because although LCRA can deal with the mis-registration errors, it will reduce the detection ability at the same time. When the mis-registration is very large, LCRA may not be always so effective. However, in practical situations, after preprocessing the mis-registration error will mostly be limited into 2 pixel. So LCRA can be useful in most applications.

3.2. Manaus Hyperion dataset

In the second experiment, we selected two Hyperion images acquired on August 14, 2004, and August 23, 2004, covering a river area near the city of Manaus, Brazil. The multi-temporal images have a scene size of 200 × 100, with 155 spectral bands. The level of the Hyperion data is L1R. Manual relative geometric correction was applied in the pre-processing. It was such a short time interval between August 14 and August 23 that we could assume there were not any land-cover changes in the scene, except for the ships in the Negro River. The anomalous changes corresponding to the ships, which were in the white circles and contained 47 pixel, were selected as the reference for the visual and spectral analysis, as shown in Fig. 11(c).

Fig. 12 shows the AUC values with different numbers of top bands and the corresponding 1/eigenvalue of each band. The AUC curve increases to the maximum with the five top bands and then decreases slowly, since the bands with a high index contain little change information and most of the noise. The band number determined by the criterion of the 1/eigenvalue being greater than 1 was nine. Its AUC value was also very close to the maximum, as indicated by the dashed line.

The AUC values for SFA, Dif-PCA, CC-PCA, and CE-PCA, with different numbers of top bands, are shown in Fig. 13. The AUC curve of SFA increases to the maximum value with the five top bands and then decreases slowly. In most cases, the curve of SFA lies above those of the comparative methods. Among the three methods, CC-PCA always performs better than Dif-PCA and CE-PCA.

Table 2 shows the maximum AUC values of the proposed method and the other state-of-the-art methods. The highest AUC values were obtained by the proposed method with five top bands. It can be seen that the anomaly change detection results after PCA feature extraction are better than those without PCA, since the anomalous changes have a better separability in the feature space. SAM gave a good performance in this experiment, since there were not any pervasive changes in the study scene.

The anomaly change detection results of SFA and the comparative methods are shown in Fig. 14, where the anomalous changes are circled by white lines. All the resulting images were stretched

---

Fig. 18. ROC curves of SFA and SFA with LCRA with three window sizes with the mis-registration range of (a) one pixel, (b) two pixels, (c) five pixels and (d) eight pixels.
by the square root stretch method, according to the reference pixels, with ENVI software. It can be seen that SFA clearly out-performed most of the comparative methods. The results of SFA show fewer false alarms in the scene. The PCA-based methods show better results than those without the PCA process. SubHyp, MAD, and SAM also obtained good performances.

The quantitative assessment of the anomaly change detection results in Fig. 14 is employed through ROC curves, which are shown fewer false alarms in the scene. The PCA-based methods by the square root stretch method, according to the reference runtime compared with the other methods. Fig. 16 illustrates that our proposed method takes an acceptable multi-temporal hyperspectral images. It is worth noting that IRMAD costs too much time, which depends on the times of iteration. All the methods with feature selection were detected with the top 5 bands for a fair comparison. Fig. 16 illustrates that our proposed method takes an acceptable runtime compared with the other methods.

Fig. 17 shows the binary change map obtained by OTSU and k-means for SFA with five output bands. The DR and FAR of OTSU were 0.5319 and 0.0054, and those of k-means were 1 and 0.0494, respectively. This demonstrates that OTSU tends to find a threshold with a low DR and FAR, while k-means tend to obtain high DR and FAR values.

Fig. 18 shows the ROC curves of SFA and SFA with LCRA when dealing with mis-registration errors. It can be observed that when the mis-registration is not very large, the original SFA was strongly negatively affected by the mis-registration error. The AUC decreased sharply when the mis-registration magnitude increased. LCRA clearly improved the performance of SFA, giving much higher AUC values. When the window size was higher than the mis-registration magnitude, SFA with LCRA obtained a satisfactory performance. It should be noted that the optimal window size was not the manual mis-registration magnitude, because the original dataset was not perfectly co-registered. When the mis-registration error is as large as five or eight pixels, the improvement will not be obviously, as discussed in experiment I. It is worth noting that with the increase of mis-registration errors from Fig. 18(a) to (d), the ROC curves are lower and lower. It means that even though LCRA can improve the robustness of anomaly change detection algorithm for mis-registration, the error will also decrease the accuracy of the results.

4. Conclusion

In this paper, we propose a slow feature analysis (SFA) based anomaly change detection method for multi-temporal hyperspectral remote sensing images. SFA is employed on the whole multi-temporal images to obtain the change residual image. Several of the top bands of the residual image are selected as the input for the RX anomaly detector to get the final anomaly change detection result. Two sets of experiments using multi-temporal Hyperion data sets were employed to evaluate the proposed method. The results indicated that the proposed method outperformed the other state-of-the-art anomaly change detection methods, giving higher AUC values and better ROC curves. In addition, the feature extraction can clearly improve the performance of the anomaly change detection.

References


Chen Wu received B.S. degree in surveying and mapping engineering from Southeast University, Nanjing, China, in 2010. He is currently pursuing the Ph.D. degree at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China. His research interests include multitemporal remote sensing image change detection and analysis in multispectral and hyperspectral images.

Liangpei Zhang received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998. He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a “Chang-Jiang Scholar” Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 310 research papers. He is the holder of five patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.


Bo Du received the B.S. degree from Wuhan University, Wuhan, China, in 2005, and received the Ph.D. degree in Photogrammetry and Remote Sensing from State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2010. He is currently an associate professor with the School of Computer, Wuhan University, Wuhan, China. His major research interests include pattern recognition, hyperspectral image processing, and signal processing.