

Article

Spectrally-Spatially Regularized Low-Rank and Sparse Decomposition: A Novel Method for Change Detection in Multitemporal Hyperspectral Images

Zhao Chen ^{1,*} and Bin Wang ^{2,3}

¹ School of Computer Science and Technology, Donghua University, Shanghai 201620, China

² Key Laboratory for Information Science of Electromagnetic Waves (MoE), Fudan University, Shanghai 200433, China; wangbin@fudan.edu.cn

³ Research Center of Smart Networks and Systems, School of Information Science and Technology, Fudan University, Shanghai 200433, China

* Correspondence: chenzhao@dhu.edu.cn

Received: 30 August 2017; Accepted: 9 October 2017; Published: 12 October 2017

Abstract: Change detection (CD) for multitemporal hyperspectral images (HSI) can be approached as classification consisting of two steps, change feature extraction and change identification. This paper is focused on binary classification of the changed and the unchanged samples, which is the essential case of change detection. Meanwhile, it is challenging to extract clean change features from heavily corrupted spectral change vectors (SCV) of multitemporal HSI. The corruptions can be characterized as gross sample-specific errors, i.e., outliers, and small entry-wise noise following Gaussian distribution. To address the issue, this paper proposes a novel Spectrally-Spatially (SS) Regularized Low-Rank and Sparse Decomposition (LRSD) model, denoted by LRSD_SS. It decomposes the SCV into three components, a locally smoothed low-rank matrix for the clean change features, a sparse matrix for the outliers and an error matrix for the small Gaussian noise. The proposed method is effective in change feature extraction and robust to noise corruptions as it exploits the underlying data structures of the SCV, especially local spectral-spatial smoothness. It is also efficient since there is a closed-form solution for the feature component in the optimization problem of LRSD_SS. The experimental results in the paper show that the proposed method outperforms several classic methods which only deal with the spectral domain of image samples, as well as some state-of-the-art methods which use both spectral and spatial information.

Keywords: change detection; classification; feature extraction; low-rank and sparse decomposition; spectral-spatial regularization; outlier; robustness

1. Introduction

Detection of land changes in multitemporal remote sensing images is useful to various areas, such as disaster monitoring, resource management and urban development [1–4]. Change detection (CD) is mostly studied in two ways [5–15]. One is to design methods specifically for CD, e.g., the well-established framework for automatic and unsupervised detection [5–7] using the polar-coordinates of spectral change vectors (SCV) of multitemporal images. Another exemplary method of this type proposes a novel concept, change endmember, for unsupervised identification of hierarchical changes [8]. The other way is to adapt existing image pattern recognition techniques to change detection. A typical method in this category relies on fusion and classification to increase change detection accuracy [9]. A family of subspace-based CD (SCD) methods regards CD as target detection [10]. When approached as a classification task, CD can be accomplished in two steps, change feature extraction and change identification [9,11]. The first step is to obtain features of changes, usually

from the SCV of multitemporal images. The second step is to identify changed samples by these enhanced features, perhaps also specifying different kinds of changes, including the unchanged class, i.e., background [16–19]. In this way, many conventional feature extraction (e.g., Principal Component Analysis (PCA) [11,19]) and classification methods (e.g., K-Means clustering [5,20]) have been directly applied to CD. Since this approach can be very inclusive and flexible, the classification framework is adopted for CD in this paper.

Recently, multitemporal hyperspectral images (HSI) have begun to attract attentions by exhibiting great potentials for CD [8,10,16,21,22]. Produced as the subtraction result of bitemporal HSI, the SCV are basically a set of HSI, featuring in high spectral resolution. Thus, it is natural to detect changes by making use of the rich spectral details of the SCV as did by most existing methods [8,10,22]. However, these methods neglect spatial information, which is proven very useful in analysis of HSI [23–28]. Then, it is natural to expect that advanced spectral-spatial classification algorithms could have some effects on the increase of change detection accuracy. Unfortunately, there are potential hazards in directly applying these methods to HSI change detection, since they are not able to deal with the problems that particularly exist in the SCV.

One of the most serious problems is that the SCV are usually corrupted. Apart from the actual noise in HSI [29–35], there are many other causes for corruptions, such as misregistration, spectral variability over the spatial domain and spectral deviation over the sampling time. Take misregistration as an example [36–40]. There might be a great spectral dissimilarity between two misaligned pixels of two different images, where it should not be if the images were properly registered. The spectral difference could be treated as change feature by a pixel-based method, thus creating false alarms and hurting detection accuracy. An adaptive SCD (ASCD) method [10] managed to deal with the misregistrations but neglected corruptions of other sources. What is worse, the amount of corruptions in one original set of temporal HSI is likely to be doubled in SCV, since the SCV is the difference between the bitemporal HSI.

To solve the problem, this paper proposes a Spectrally-Spatially Regularized Low-Rank and Sparse Decomposition (LRSD_SS) method for change feature extraction. It addresses the most essential case in HSI change detection, which is binary classification of the changed and the unchanged samples [5–7,10]. The proposed model is built upon the reasonable assumption that the desired change features and the unwanted corruptions are separable due to their intrinsically different data structures in SCV [41–48]. The motivation is analyzed in three aspects as follows.

First, being high-dimensional data, the SCV exhibit intrinsic low-rank features [41–48]. This assumption is based on two facts: the number of endmembers (i.e., spectral signatures of substances) in the temporal HSI is far less than the number of bands [49,50]; and the number of real changes is far less than the number of bands [8,10]. Therefore, it is safe to infer that the features of real changes, including the background, lie in some low-dimensional subspace. The background data are naturally low-rank because the unchanged samples have approximately zero SCV.

Second, the corruptions of various sources in the SCV can be characterized as gross sample specific noise or small entry-wise noise added to the clean low-rank data [42–44]. The former is known as outliers, e.g., impulse noise and gross Gaussian noise. Since they only corrupt specific pixels, the outliers exhibit sparsity in high-dimensional data such as SCV. The latter is regarded as thermal noise, being widely spread in HSI while following the Gaussian distribution with a relatively low intensity, compared to that of the outliers. The underlying structures of the high-dimensional noisy data, i.e., SCV in this case, can be exploited by the well-established LRSD model [43] to simultaneously separate the low-rank features, the sparse outliers and the small Gaussian noise. LRSD is much more robust than its original model, PCA, as the former considers the outliers [42–44]. Given the two assumptions above, LRSD is chosen as the basic model for designing the proposed change feature extraction method.

Third, samples of real changes in the SCV exhibit local smoothness, as ground objects within a small spatial region are likely to belong to the same class and share similar spectral signatures. Meanwhile, noise is very unlikely to be locally smooth due to its sparsity or spatial randomness. The

spectral-spatial information can be crucial in differentiation of low-rank change features from low-rank noise, e.g., dead lines that are simultaneously low-rank and sparse [44,45]. Therefore, the novel SS regularization is designed. It characterizes the local patterns by averaging pixel-wise similarity measurements within each square neighborhood centered at one sample of SCV. In this way, the proposed method, LRSD_SS, maintains the change features by enhancing the local spectral-spatial smoothness in the low-rank data and further suppressing the noise that could not be completely contained or removed in the noise components yielded by the original LRSD.

Apart from effectiveness, efficiency is also considered by the proposed method in the following aspects. Firstly, the SS regularization is designed to process all the spectral bands of the SCV simultaneously, which is less time-consuming than processing each band image separately, as done by a recently published total-variation (TV) [51,52] regularized low-rank model [44] denoted by LRSD_TV in this paper. Secondly, the SS regularization is formulated based on the Frobenius norm, which results in a closed-form solution for the feature data when LRSD_SS is solved using the Augmented Lagrange Multiplier (ALM) [44,53].

After the low-rank change features are extracted, they are fed to some classifier to be identified. As the temporal changes are mostly unpredictable, prior knowledge is usually scarce and unsupervised classification is adopted [54,55]. In fact, many unsupervised classifiers can be used, as long as they are able to perform binary classification. This paper chooses the classic K-Means [20] since it is quite efficient while being plain enough to fully expose the effects of the previous change feature extraction methods on detection accuracy. Our experiments show that LRSD_SS has outperformed several classic algorithms (e.g., Spectral Angle Mapping (SAM) [10], PCA [11], and LRSD [43]), which only consider spectral information, as well as some of the state-of-the-art ones, which consider both spectral and spatial information (e.g., ASCD [10] and LRSD_TV [44]).

The main contributions of this paper are summarized as follows:

1. It deals with a critical but not well solved problem with CD in HSI, which is to recover clean change data from noisy SCV of bitemporal HSI. Moreover, it addresses the issue from a relatively new angle for CD, which is to extract change features by exploiting inherent data structures exhibited in the SCV. To do so, this paper proposes a novel method, LRSD_SS, based on LRSD. Although the original LRSD has been applied in some areas of non-temporal HSI processing [38,49], it has not been used in temporal HSI change detection, to the best of our knowledge. This paper tries to fill the void by construction of LRSD_SS and demonstration of its capacity in solving the aforementioned problem of CD in multitemporal HSI.
2. For better characterization of the underlying data structures, this paper designs a novel SS regularization superimposed on LRSD. The regularization enhances the local spectral-spatial smoothness, which normally exhibits in real change data, but is seldom observed from noise. Thus, the proposed LRSD_SS can further suppress the noise, especially those being as low-rank as the change features, which, however, has not been considered by the original LRSD. Experimental results show that LRSD_SS is robust to noise of various forms and intensities while being able to extract change features and increase change detection accuracy.
3. This paper offers an implementation-friendly method for change detection in temporal HSI. The SS regularization processes all the bands of the SCV simultaneously and yields a closed form solution for the smoothed low-rank data matrix, thus ensuring efficiency.

The rest of the paper is organized as follows. Section 2 provides background knowledge on low-rank decomposition. Section 3 presents the proposed method, LRSD_SS, in full details. Section 4 analyzes experimental results. Finally, Section 5 concludes the paper.

2. Background Knowledge

2.1. The LRSD Model

LRSD [43] is sometimes referred to as Low-Rank Matrix Recovery (LRMR) [42]. The model is derived from robust PCA (RPCA), which originates from PCA [19,41–47]. It is designed to separate both gross sample-specific outliers and small entry-wise Gaussian noise from clean low-rank data, thus being more robust than PCA in low-rank approximation and denoising [19,42,43,47]. Suppose $\mathbf{Y} \in R^{M \times B}$ ($B \leq M$) is an observed high-dimensional matrix corrupted by noise, where M and B are the number of samples and the dimension of each sample, respectively. The LRSD model of \mathbf{Y} is formulated as

$$\mathbf{Y} = \mathbf{L} + \mathbf{S} + \mathbf{N} \quad (1)$$

where $\mathbf{L} \in R^{M \times B}$ is the low-rank data, $\mathbf{S} \in R^{M \times B}$ holds the outliers which exhibit sparsity, and $\mathbf{N} \in R^{M \times B}$ represents the small entry-wise noise which are i.i.d. Gaussian.

In the case of change detection, \mathbf{Y} is SCV of bitemporal HSI $\mathbf{T}_1, \mathbf{T}_2 \in R^{M \times B}$ while M and B are the number of pixels and the number of bands, respectively. \mathbf{L} maintains the change features. The reasons for choosing LRSD instead of other low-rank decomposition models for change detection are manifold. First, LRSD accurately models the intrinsic data structures of the SCV. Second, LRSD deals with both small noise and gross outliers, which widely exist in the SCV. Plus, it can be solved efficiently by bilateral random projections (BRP) [43,45,56,57].

2.2. Optimization for LRSD

The basic optimization problem [19,42–47] of the LRSD model is described as follows

$$\min_{\mathbf{L}, \mathbf{S}} \text{rank}(\mathbf{L}) + \lambda \|\mathbf{S}\|_0 \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{L} + \mathbf{S} + \mathbf{N}, \quad (2)$$

where λ is a constant for the trade-off between $\text{rank}(\mathbf{L})$ and $\|\mathbf{S}\|_0$. Equation (2) can be rewritten as

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{Y} - \mathbf{L} - \mathbf{S}\|_F^2 \quad \text{s.t.} \quad \text{rank}(\mathbf{L}) \leq r, \quad \text{card}(\mathbf{S}) \leq k, \quad (3)$$

where r and k stand for the upper bound of the rank of \mathbf{L} and the cardinality of \mathbf{S} , respectively [37]. Their values are set a priori.

Since Equation (2) is highly nonconvex with no efficient solution, a tractable optimization problem is obtained by relaxing Equation (2), i.e., replacing the $\text{rank}(\bullet)$ and the ℓ_0 -norm in Function (2) with ℓ_* -norm (nuclear norm) and ℓ_1 -norm, respectively [43,58,59],

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{L} - \mathbf{S}\|_F^2 \leq \varepsilon, \quad (4)$$

where ε is a constant related to the standard deviation of the random noise \mathbf{N} .

2.3. Spatial Regularization

Let $\|\bullet\|_{spa}$ be an arbitrary spatial regularization. By the maximum a posteriori (MAP) estimation theory [35,44], $\|\bullet\|_{spa}$ can be superimposed on Equation (4), resulting in a spatially regularized optimization function as follows

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 + \tau \|\mathbf{X}\|_{spa} \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{L} - \mathbf{S}\|_F^2 \leq \varepsilon, \quad \mathbf{L} = \mathbf{X}, \quad (5)$$

where $\mathbf{X} \in R^{M \times B}$ is a set of spatially smoothed low-rank data and τ is a constant for trade-off and \mathbf{X} is introduced to solve \mathbf{L} conveniently by iterations.

3. Methodology

When approached as a classification problem, change detection can be accomplished in two steps: change feature extraction and identification. This paper addresses the most essential case: obtaining the change features from bitemporal HSI and clustering them into two classes, the changed and the unchanged. The framework for change detection in HSI is illustrated by Figure 1. In this paper, a spectrally-spatially regularized LRSD model, LRSD_SS, is proposed for change feature extraction, which is explicitly explained in the next subsection.

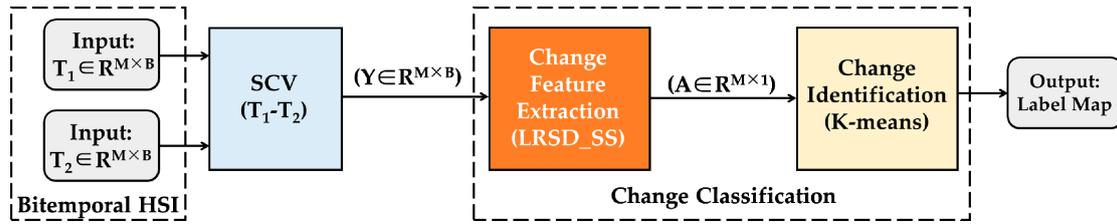


Figure 1. A framework for change detection in HSI, where $T_1 \in R^{M \times B}$ and $T_2 \in R^{M \times B}$ are bitemporal hyperspectral images acquired at time 1 and time 2, respectively; $Y \in R^{M \times B}$ is SCV of T_1 and T_2 ; and $A \in R^{M \times 1}$ is amplitude vector of change features.

3.1. The Proposed Model: LRSD_SS

The main idea of the proposed method, LRSD_SS, is to extract change features by exploiting intrinsic data structures in SCV of bitemporal HSI. The features of real changes and background are assumed to be lying in a low-dimensional subspace, thus being low-rank, compared to the original high-dimensional SCV. Meanwhile, image corruptions in SCV may lead to false changes and must be removed from the low-rank change features of interest. They can be characterized as gross sample-specific corruptions, i.e., outliers, or small entry-wise noise added to the clean, low-rank data. The former exhibits sparsity while the latter behaves as thermal noise and follows Gaussian distribution. Therefore, we use the LRSD model in Equation (1), which is able to simultaneously extract low-rank features (L) and remove sparse outliers (S) and Gaussian noise (N).

However, the spectral data structures of the change features and those of the corruptions may not always be totally different. Some noise can be simultaneously low-rank and sparse, for instance, dead lines located at the same row or column in all the band images. In that case, conventional LRSD algorithms may fail to recover the clean data [44,45]. Since the spectral information alone is not enough for clean data recovery, spatial information is engaged. It is assumed that the real change features exhibit local spectral-spatial smoothness while the noise does not. This data structure can help in eliminating the aforementioned noise. We design a spectral-spatial (SS) regularization for LRSD to exploit the local patterns and enhance the change features, thus proposing the LRSD_SS model. The SS regularization is defined as follows,

$$\|\mathbf{X}\|_{SS} = \sum_{i=1}^I \sum_{j=1}^J \sum_{w=1}^{W^2-1} \omega_{ij}^w \|\mathbf{x}_{ij} - \mathbf{x}_{ij}^w\|_F^2 = \sum_{i=1}^I \sum_{j=1}^J \sum_{w=1}^{W^2-1} \omega_{ij}^w d_{ij}^w, \quad (6)$$

where $d_{ij}^w = \|\mathbf{x}_{ij} - \mathbf{x}_{ij}^w\|_F^2$, i.e., the squared Euclidean distance between pixel \mathbf{x}_{ij} and pixel \mathbf{x}_{ij}^w ($i = 1, 2, \dots, I, j = 1, 2, \dots, J$). \mathbf{x}_{ij} is located at the i th row and the j th column of the feature data $\Theta \mathbf{X} \in R^{I \times J \times B}$, whereas Θ is an operator that reshapes a $IJ \times 1$ vector into a $I \times J$ matrix. \mathbf{x}_{ij}^w is a sample within the $W \times W$ neighborhood centered at \mathbf{x}_{ij} while $\mathbf{x}_{ij}^w \neq \mathbf{x}_{ij}$. Let W be an odd number

and (i_n, j_n) be the spatial location of \mathbf{x}_{ij}^w in $\Theta\mathbf{X}$, whereas $i_n \in \{1, 2, \dots, I\}$, $j_n \in \{1, 2, \dots, J\}$ and $(i_n, j_n) \neq (i, j)$. Hence,

$$w = \begin{cases} (j_n - j)W + i_n - i + (W^2 + 1)/2, & \text{if } (j_n < j) \vee (j_n = j \ \& \ i_n < i) \\ (j_n - j)W + i_n - i + (W^2 - 1)/2, & \text{otherwise} \end{cases} \in \{1, 2, \dots, W^2 - 1\}. \quad (7)$$

For fixed i, j and w , ω_{ij}^w is a constant weighing the dissimilarity d_{ij}^w between \mathbf{x}_{ij}^w and \mathbf{x}_{ij} , which can be adjusted a priori so as to avoid over-smoothness. Figure 2 illustrates the generation of the neighborhood and the weights for computation of Equation (6) when $W = 3$. As shown in the figure, \mathbf{x}_{ij} can be denoted by \mathbf{x}_{ij}^0 and no weight is needed for d_{ij}^0 as $d_{ij}^0 = \|\mathbf{x}_{ij} - \mathbf{x}_{ij}^0\|_F^2 = 0$.

With the proposed SS regularization in (6) being substituted for $\|\bullet\|_{spa}$ in Equation (5), the optimization problem of LRSD_SS is obtained as follows,

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \quad & \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 + \tau \|\mathbf{X}\|_{SS} \\ \text{s.t.} \quad & \|\mathbf{Y} - \mathbf{L} - \mathbf{S}\|_F^2 \leq \varepsilon, \quad \mathbf{L} = \mathbf{X} \end{aligned} \quad (8)$$

Since $\|\bullet\|_{SS}$ is particularly designed based on the Frobenius norm, the smoothed images \mathbf{X} in Equation (8) can be solved in a closed form within the ALM framework, as explained in the next subsection. It will be demonstrated in Section 4 that with the help of $\|\bullet\|_{SS}$, the proposed LRSD_SS can improve change detection results.

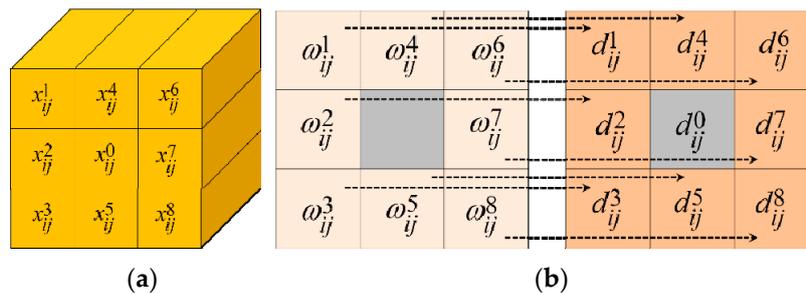


Figure 2. An illustration of the proposed Spectral-Spatial (SS) regularization, with: (a) $W \times W$ neighborhood centered by pixel \mathbf{x}_{ij} (\mathbf{x}_{ij}^0); and (b) weights superimposed on d_{ij}^w , whereas $i = 1, 2, \dots, I, j = 1, 2, \dots, J, w = 0, 1, \dots, W^2 - 1$ and $W = 3$.

3.2. Optimization for LRSD_SS

The optimization problem formulated by Equation (8) is solved by the ALM method in this paper. ALM is adopted mainly because of its excellent convergence property and simple implementation [53], which are not typical with other methods used for solving low-rank decomposition models, such as the Alternating Direction Method of Multipliers (ADMM) [60,61]. To be specific, the inexact ALM algorithm is implemented since the validity and optimality of the algorithm is theoretically guaranteed [53]. It has also been proved practical as He et al. [44] successfully solved their TV-regularized low-rank decomposition model for HSI denoising by the algorithm. Within the ALM framework, Equation (8) can be rewritten as follows,

$$\begin{aligned} \min l(\mathbf{L}, \mathbf{S}, \mathbf{X}, \Lambda_1, \Lambda_2) = \arg \min_{\mathbf{L}, \mathbf{S}, \mathbf{X}, \Lambda_1, \Lambda_2} \quad & \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 + \tau \|\mathbf{X}\|_{SS} \\ & + \langle \Lambda_1, \mathbf{Y} - \mathbf{L} - \mathbf{S} \rangle + \langle \Lambda_2, \mathbf{X} - \mathbf{L} \rangle \\ & + \frac{\mu}{2} \left(\|\mathbf{Y} - \mathbf{L} - \mathbf{S}\|_F^2 + \|\mathbf{X} - \mathbf{L}\|_F^2 \right) \end{aligned} \quad (9)$$

where Λ_1 and Λ_2 are Lagrange multipliers. μ is a penalty parameter.

Since there are more than one unknown variables in (8), they are solved alternatively and iteratively by optimizing Equation (8) over one variable while fixing the others. Let t be the iteration counter. Derived from Equation (8), the optimization function to update \mathbf{L} is as follows [44],

$$\mathbf{L}^{(t+1)} = \arg \min_{\mathbf{L}} \|\mathbf{L}\|_* + \langle \mathbf{\Lambda}_1^{(t)}, \mathbf{Y} - \mathbf{L} - \mathbf{S}^{(t)} \rangle + \langle \mathbf{\Lambda}_2^{(t)}, \mathbf{X}^{(t)} - \mathbf{L} \rangle + \frac{\mu}{2} (\|\mathbf{Y} - \mathbf{L} - \mathbf{S}^{(t)}\|_F^2 + \|\mathbf{X}^{(t)} - \mathbf{L}\|_F^2). \quad (10)$$

In this paper, Bilateral Random Projections (BRP) is adopted to estimate \mathbf{L} , which has been fully developed and successfully applied in an algorithm named Go Decomposition (GoDec) algorithm to solve the LRSD model [42–44,56,57]. Let $\mathbf{H} = \frac{1}{2} (\mathbf{Y} + \mathbf{X}^{(t)} - \mathbf{S}^{(t)} + \frac{1}{\mu} (\mathbf{\Lambda}_1^{(t)} + \mathbf{\Lambda}_2^{(t)}))$ and $\tilde{\mathbf{H}} = (\mathbf{H}\mathbf{H}^T)^q \mathbf{H}$. As suggested by the power scheme [43], BRP is applied to $\tilde{\mathbf{H}}$ instead of \mathbf{H} , since the singular values of $\tilde{\mathbf{H}}$ decay faster than those of \mathbf{H} . The BRP results of $\tilde{\mathbf{H}}$ are \mathbf{Z}_1 and \mathbf{Z}_2 , where $\mathbf{Z}_1 = \tilde{\mathbf{H}}\mathbf{A}_1$ and $\mathbf{Z}_2 = \tilde{\mathbf{H}}^T \mathbf{A}_2$. $\mathbf{A}_1 \in R^{B \times r}$ and $\mathbf{A}_2 \in R^{M \times r}$ are two random matrices, and r is the upper bound of the rank of \mathbf{L} . Given that, $\mathbf{L}^{(t+1)}$ in Equation (10) is estimated as follows [43],

$$\mathbf{L}^{(t+1)} = \mathbf{Q}_1 \left[\mathbf{R}_1 (\mathbf{A}_2^T \mathbf{Z}_1)^{-1} \mathbf{R}_2^T \right]^{\frac{1}{2q+1}} \mathbf{Q}_2^T, \quad (11)$$

where $\mathbf{Q}_1 \in R^{M \times M}$ and $\mathbf{R}_1 \in R^{M \times r}$ are the QR decomposition matrices of \mathbf{Z}_1 while $\mathbf{Q}_2 \in R^{B \times B}$ and $\mathbf{R}_2 \in R^{B \times r}$ are the QR decomposition matrices of \mathbf{Z}_2 . Conventionally, \mathbf{L} in Equation (10) is solved by Singular Value Decomposition (SVD) [43,44,62]. The BRP technique used here is basically an efficient approximation algorithm for SVD. For solving the same $\mathbf{L} \in R^{M \times B}$, SVD requires $\min(M^2B, MB^2)$ flops while BRP requires $r^2(M + 3B + 4r) + (4q + 4)MBr$ [43]. When $r \ll B \ll M$, BRP is more time-saving than SVD. Moreover, the relative error of the BRP approximation can be very close to that of SVD if q is sufficiently large, say, $q \geq 3$ [57]. To be noted, r and q are set a priori.

Derived from Equation (8), \mathbf{X} is updated as follows,

$$\mathbf{X}^{(t+1)} = \arg \min_{\mathbf{X}} J(\mathbf{X}) = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{X} - \mathbf{Q}\|_2^2 + \frac{\tau}{\mu} \|\mathbf{X}\|_{SS}, \quad (12)$$

where $\mathbf{Q} = \mathbf{L}^{(t+1)} - \mathbf{\Lambda}_2^{(t)} / \mu$. Let $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_m, \dots, \mathbf{q}_M]^T$, $\tilde{\mathbf{Q}}^m = [\tilde{\mathbf{q}}_w^m]_{w=1}^{W^2} = [\mathbf{x}_m^1, \mathbf{x}_m^2, \dots, \mathbf{x}_m^{W^2-1}, \mathbf{q}_m]^T$, $\mathbf{v} = \{\tilde{\omega}_w\}_{w=1}^{W^2} = [\tau\omega_1/\mu, \tau\omega_2/\mu, \dots, \tau\omega_{W^2-1}/\mu, 1/2]^T$ and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m, \dots, \mathbf{x}_M]^T$, whereas \mathbf{x}_m^w is a sample within the neighborhood of \mathbf{x}_m defined in Equation (6) and $m = (j - 1) \times I + i$. Thus,

$$\begin{aligned} J(\mathbf{X}) &= \frac{1}{2} \|\mathbf{X} - \mathbf{Q}\|_2^2 + \frac{\tau}{\mu} \sum_{i=1}^I \sum_{j=1}^J \sum_{w=1}^{W^2-1} \omega_w \|\mathbf{x}_{ij} - \mathbf{x}_{ij}^w\|_F^2 \\ &= \sum_{m=1}^M \frac{1}{2} (\mathbf{x}_m - \mathbf{q}_m)^T (\mathbf{x}_m - \mathbf{q}_m) + \sum_{m=1}^M \sum_{w=1}^{W^2-1} \frac{\tau}{\mu} \omega_w (\mathbf{x}_m - \mathbf{x}_m^w)^T (\mathbf{x}_m - \mathbf{x}_m^w) \cdot \\ &= \sum_{m=1}^M \sum_{w=1}^{W^2} \tilde{\omega}_w (\mathbf{x}_m - \tilde{\mathbf{q}}_w^m)^T (\mathbf{x}_m - \tilde{\mathbf{q}}_w^m) \end{aligned} \quad (13)$$

Let $\frac{\partial J(\mathbf{X})}{\partial \mathbf{X}} = 0$, then

$$\mathbf{x}_m^{(t+1)} = \frac{\tilde{\mathbf{Q}}^{m(t)} \mathbf{v}}{\mathbf{v}^T \mathbf{1}}, m = 1, 2, \dots, M, \quad (14)$$

where $\tilde{\mathbf{Q}}^{m(t)} = [\mathbf{x}_m^{1(t)}, \mathbf{x}_m^{2(t)}, \dots, \mathbf{x}_m^{W^2-1(t)}, \mathbf{q}_m]$ and $\mathbf{1} \in R^{W^2 \times 1}$ is a vector with all the elements equal to 1. It should be noted that $\mathbf{X}^{(t+1)}$ is solved by Equation (14) in a closed form. The computation complexity of updating $\mathbf{X}^{(t+1)}$ is about $O(BMW^2)$.

According to Equation (8), the optimization function of \mathbf{S} is written as follows,

$$\mathbf{S}^{(t+1)} = \arg \min_{\mathbf{S}} \lambda \|\mathbf{S}\|_1 + \frac{\mu}{2} \left\| \mathbf{S} - \left(\mathbf{Y} - \mathbf{L}^{(t+1)} + \frac{\Lambda_1^{(t)}}{\mu} \right) \right\|_2^2. \quad (15)$$

Given the soft-thresholding (shrinkage) technique [44,63],

$$\mathbf{S}^{(t+1)} = \mathfrak{R}_{\frac{\Delta}{\mu}} \left(\mathbf{Y} - \mathbf{L}^{(t+1)} + \frac{\Lambda_1^{(t)}}{\mu} \right), \quad (16)$$

where the shrinkage operator is defined as $\mathfrak{R}_{\Delta}(x) = \begin{cases} x - \Delta, & \text{if } x > \Delta \\ x + \Delta, & \text{if } x < -\Delta \\ 0, & \text{otherwise} \end{cases}$, for $x \in R$ and $\Delta > 0$.

After $\mathbf{L}^{(t+1)}$ and $\mathbf{S}^{(t+1)}$ are obtained, the Gaussian noise $\mathbf{N}^{(t+1)}$ in Equation (1) is updated by

$$\mathbf{N}^{(t+1)} = \mathbf{Y} - \mathbf{L}^{(t+1)} - \mathbf{S}^{(t+1)}, \quad (17)$$

The procedures of LRSD_SS are summarized in Algorithm 1, whereas \mathbf{L} or \mathbf{X} is the output change feature matrix. The technique of updating the parameter μ in every iteration can reinforce the convergence of the ALM-based algorithms [44,53]. Apart from the iteration number t , two types of estimation errors are used as stopping criteria for LRSD_SS, which are defined as follows,

$$Error1 = \frac{\|\mathbf{Y} - \mathbf{L}^{(t+1)} - \mathbf{S}^{(t+1)}\|_F}{\|\mathbf{Y}\|_F}, \quad (18)$$

$$Error2 = \|\mathbf{L}^{(t+1)} - \mathbf{X}^{(t+1)}\|_{\infty}. \quad (19)$$

The upper bounds of $Error1$ and $Error2$ are ε_1 and ε_2 , respectively, whereas $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$.

Algorithm 1: LRSD_SS

Input: \mathbf{Y} , r , q , τ , μ_0 , μ_{\max} , ρ , λ , ε_1 , ε_2 , t_{\max}

Output: \mathbf{L} (\mathbf{X}), \mathbf{S} and \mathbf{N}

Initialization: $\mathbf{L}^{(0)} = \mathbf{Y}$, $\mathbf{X}^{(0)} = \mathbf{S}^{(0)} = \mathbf{N}^{(0)} = \Lambda_1^{(0)} = \Lambda_2^{(0)} = \mathbf{0}$, $\mu = \mu_0$ and $t = 0$;

Do

Update $\mathbf{L}^{(t+1)}$ by (11), $\mathbf{X}^{(t+1)}$ by (14), and $\mathbf{S}^{(t+1)}$ by (16).

Update $\mathbf{N}^{(t+1)}$ by (17), $Error1$ by (18), and $Error2$ by (19).

Let $\Lambda_1^{(t+1)} = \Lambda_1^{(t)} + \mu \mathbf{N}^{(t+1)}$ and $\Lambda_2^{(t+1)} = \Lambda_2^{(t)} + \mu (\mathbf{X}^{(t+1)} - \mathbf{L}^{(t+1)})$.

Set $\mu \leftarrow \min(\rho\mu, \mu_{\max})$ and $t \leftarrow t + 1$.

while $Error1 > \varepsilon_1$, $Error2 > \varepsilon_2$ and $t \leq t_{\max}$.

3.3. Implementation

As shown in Figure 1, the proposed LRSD_SS is used as a change feature extraction method for the first step of CD. Let \mathbf{Y} be the SCV of bitemporal HSI, $\mathbf{T}_1 \in R^{M \times B}$ and $\mathbf{T}_2 \in R^{M \times B}$, where $\mathbf{Y} = \mathbf{T}_1 - \mathbf{T}_2$, M is the number of pixels and B is the number of bands. This type of SCV is widely used [5–7]. \mathbf{Y} is input to LRSD_SS in Algorithm 1, which separates the locally smoothed low-rank data (\mathbf{L}), outliers (\mathbf{S}) and small entry-wise Gaussian noise (\mathbf{N}). The proposed method is expected to enhance the real change features by maintaining \mathbf{L} , while suppressing the unwanted false changes by removing \mathbf{S} and \mathbf{N} . Therefore, LRSD_SS makes it easier to determine whether a sample is changed or unchanged in the next step.

The binary classification task is completed by K-Means [5,20]. Let $\mathbf{A} \in R^{M \times 1}$ hold the amplitude of each sample in \mathbf{L} , whereas $\mathbf{A}(m) = \|\mathbf{L}(m, :)\|_F$ and $m = 1, 2, \dots, M$. \mathbf{A} , instead of \mathbf{L} , is fed to the

classifier, because the spectral amplitude alone is sufficient for the binary classification required by CD [5–7]. As an efficient unsupervised method, K-Means is chosen for change identification since the temporal changes can be quite unexpected and prior knowledge is usually scarce [54,55]. The main procedures of LRSD_SS + K-Means for CD can be found in Figure 1.

4. Results

4.1. Datasets

4.1.1. Simulated Bitemporal Hyperspectral Data

Ten sets of simulated bitemporal HSI are used for performance evaluation of the proposed LRSD_SS. Each dataset is created from the same non-temporal HSI of 610×340 pixels, captured over the University of Pavia, Italy in 2001 by the ROSIS instrument (<http://www.ehu.es/ccwintco/uploads/e/ee/PaviaU.mat>). After removing the bands seriously affected by noise or water absorption, 103 spectral channels are left. Thus, $I = 610$, $J = 340$, $M = I \times J = 207,400$ and $B = 103$. The Pavia dataset makes the CD task very challenging as it covers several (at least nine) different kinds of ground objects and exhibits spectral variability. The simulation of T_1 and T_2 is carried out in two steps: clean bitemporal HSI ($T_{10}, T_{20} \in R^{M \times B}$) simulation and corruption simulation [42,44], as explained in the following.

The first step is illustrated in Figure 3. First, three identical copies of the original Pavia HSI are created, denoted as T_{10} , T_{temp} and T_{20} . Then, three parts of T_{temp} are selected as “source” areas and five parts of T_{20} are selected as “target” areas. The pixels in one “target” area are replaced with those in one “source” area, as indicated by the arrows in Figure 3. If the sizes of two areas are not the same, the pixels in the “source” are randomly replicated or removed to fill in the “target” area exactly right. Thereby, T_{20} is synthesized with five kinds of temporal changes with respect to T_{10} . It is made sure that most changed samples in SCV of T_{10} and T_{20} have far-above-zero spectral amplitudes, so that they are distinguishable from the unchanged ones. The SCV is denoted by Y_0 and computed as $Y_0 = T_{10} - T_{20}$, which is used as clean data for reference in some of our experiments. The false color composition of T_{10} and T_{20} are presented in Figure 4a,b, respectively. The ground truth map indicating the changed areas and the unchanged background is given by Figure 4c.

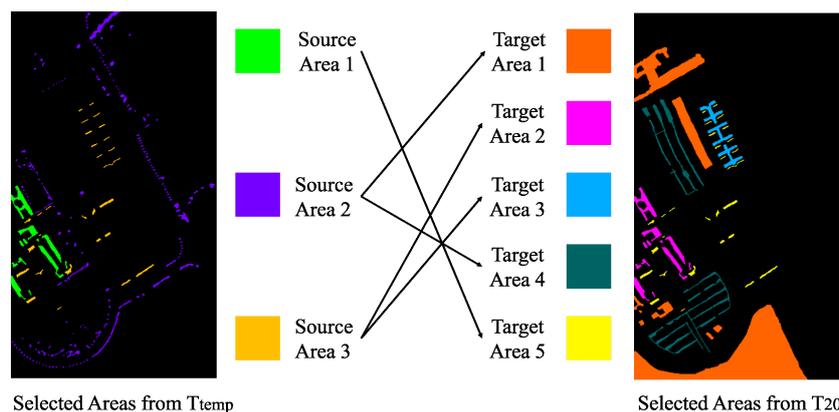


Figure 3. Synthesis of T_{20} .

In the second step, artificial noise is superimposed on T_{10} and T_{20} to simulate corruptions of various causes, as explained in Section 1. There are mainly two kinds of noise generated: small entry-wise Gaussian noise and gross sample-specific errors (outliers). Three types of outliers are simulated: Gaussian noise with high intensity, impulse noise and dead lines. By different combination of the artificial noise, ten types of corrupted bitemporal HSI, T_1 and T_2 , are created. All the simulated

datasets are normalized. Each type of the artificial noise and the bitemporal HSI are described in details as follows.

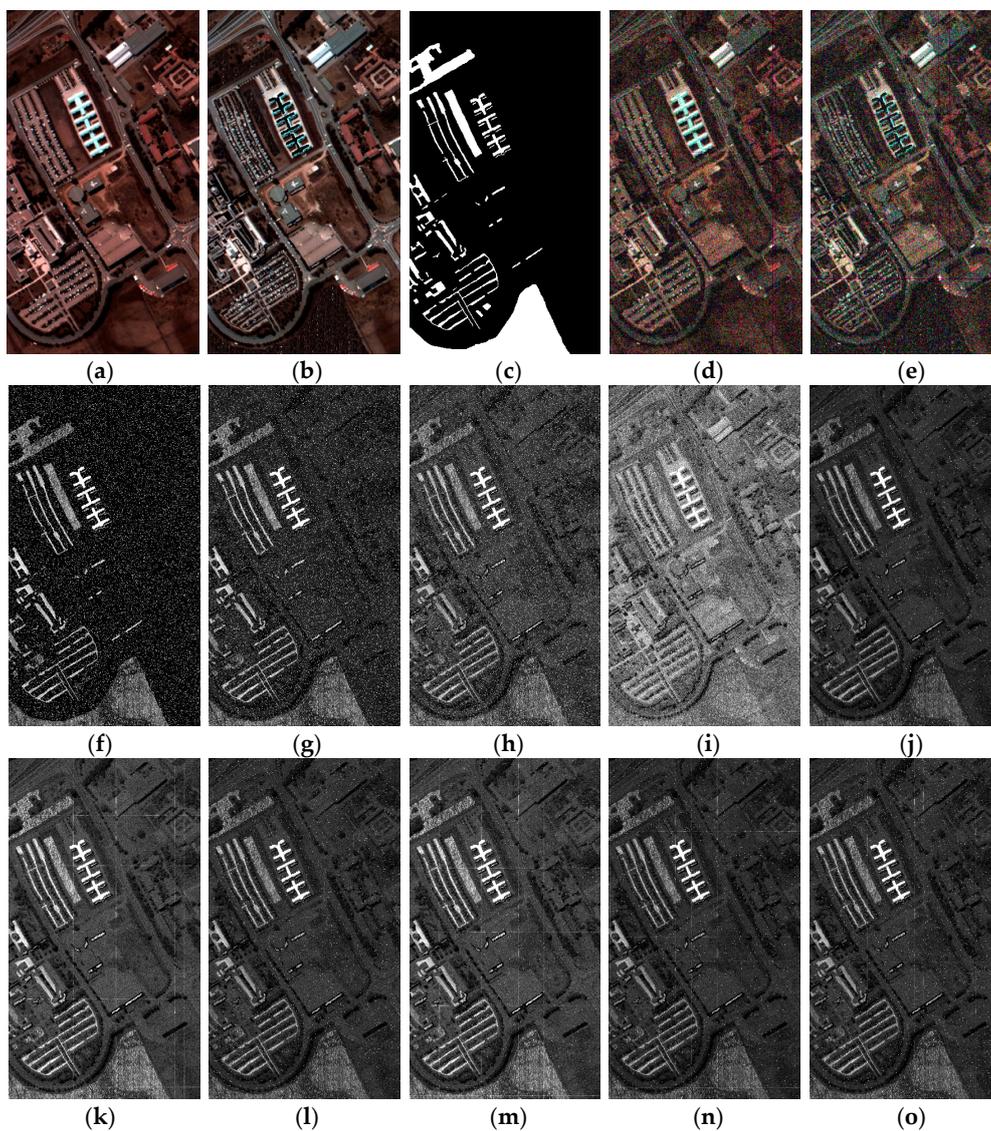


Figure 4. Simulated data, including false color composition by bands 10, 20 and 40 of: (a) T_{10} ; (b) T_{20} ; and (c) ground truth map with changed areas in white and unchanged areas (background) in black; false color composition by bands 10, 20 and 40 of a pair of: (d) T_1 ; and (e) T_2 of Data 10; and amplitude image (a) of SCV of a set of bitemporal HSI of: (f) Data 1; (g) Data 2; (h) Data 3; (i) Data 4; (j) Data 5; (k) Data 6; (l) Data 7; (m) Data 8; (n) Data 9; and (o) Data 10.

1. Noise:

- Noise 1: random Gaussian noise with a low intensity (σ_L^2), superimposed on all the bands of all the pixels;
- Noise 2: random Gaussian noise with a high intensity ($\sigma_H^2 = 0.5$), superimposed on 20 randomly selected bands of some randomly selected pixels, which occupy $\eta\%$ of all the pixels;
- Noise 3: impulse noise with an intensity of the unitary distribution, superimposed on 20 randomly selected bands of some randomly selected pixels, which occupy 0.5% of all the pixels; and

- Noise 4: four dead lines with zero intensity, two of them superimposed on 20 randomly selected bands of all the pixels in two randomly selected rows and the other two superimposed on 20 randomly selected bands of all the pixels in two randomly selected columns.

2. Bitemporal HSI:

- Data 1: Noise 1 ($\sigma_L^2 = 0.001$) + Noise 2 ($\eta = 5$) for both T_1 and T_2 ;
- Data 2: Noise 1 ($\sigma_L^2 = 0.005$) + Noise 2 ($\eta = 5$) for both T_1 and T_2 ;
- Data 3: Noise 1 ($\sigma_L^2 = 0.010$) + Noise 2 ($\eta = 5$) for both T_1 and T_2 ;
- Data 4: Noise 1 ($\sigma_L^2 = 0.050$) + Noise 2 ($\eta = 5$) for both T_1 and T_2 ;
- Data 5: Noise 1 ($\sigma_L^2 = 0.010$) + Noise 3 for both T_1 and T_2 ;
- Data 6: Noise 1 ($\sigma_L^2 = 0.010$) + Noise 4 for both T_1 and T_2 ;
- Data 7: Noise 1 ($\sigma_L^2 = 0.010$) + Noise 2 ($\eta = 0.25$) + Noise 3 for both T_1 and T_2 ;
- Data 8: Noise 1 ($\sigma_L^2 = 0.010$) + Noise 2 ($\eta = 0.25$) + Noise 4 for both T_1 and T_2 ;
- Data 9: Noise 1 ($\sigma_L^2 = 0.010$) + Noise 3 + Noise 4 for both T_1 and T_2 ; and
- Data 10: Noise 1 ($\sigma_L^2 = 0.010$) + Noise 2 ($\eta = 0.25$) + Noise 3 + Noise 4 for both T_1 and T_2 .

It should be noted that Noise 1 simulates small entry-wise Gaussian noise, while Noises 2–4 are used as outliers. Since all the noise are randomized, there are 10 different pairs of T_1 and T_2 generated for each data type. Every change detection result presented in this paper is the average outcome of 10 independent runs of a change detection method, whereas each run involves one pair of T_1 and T_2 of a data type. A pair of T_1 and T_2 of Data 10 are illustrated by Figure 4d,e. An example of the amplitude image A of the SCV ($Y = T_1 - T_2$) of each simulated data type is also presented in Figure 4f–o.

4.1.2. Real-World Dataset

Aside from the simulated data, one set of real-world bitemporal HSI is used for testing the proposed method. The dataset is acquired by Hyperion, consisting of two sets of images of Yancheng, Jiangsu province, China, each with 450×140 samples and 155 bands [10]. The bitemporal HSI are illustrated by Figure 5a,b. The ground truth map is shown in Figure 5c.

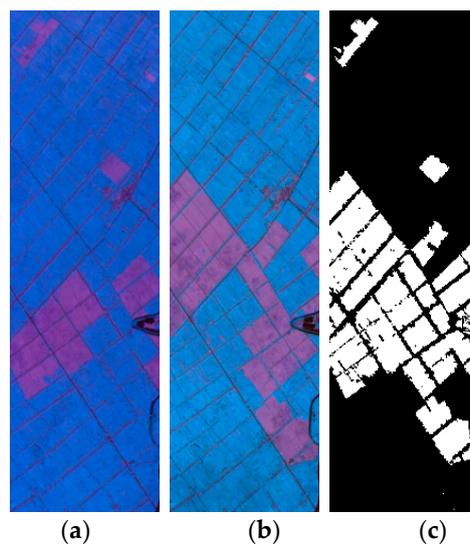


Figure 5. Real-world bitemporal data, including false color composition by bands 20, 40 and 60 of: (a) images acquired on 3 May 2006; (b) images acquired on 23 April 2007; and (c) ground truth map with changed areas in white and unchanged areas (background) in black.

4.2. Setup

The proposed LRSD_SS described in Algorithm 1 is compared with two conventional methods, SAM [10] and PCA [11], one classic low-rank decomposition method, LRSD [43], one state-of-the-art method based on subspace projection, ASCD [10], and one state-of-the-art method based on low-rank decomposition, LRSD_TV [44]. ASCD is an upgraded version of SAM as the former considers local spatial information but the latter does not. These methods exploit the projection distance between T_1 and T_2 . The LRSD-based methods use Y as SCV. The TV regularization in LRSD_TV is solved by the toolbox provided by its original publisher (http://iew3.technion.ac.il/~becka/papers/tv_fista.zip), which was successfully adopted by reference [44] for HSI denoising. For fairness, all the LRSD-based methods engage BRP to estimate L . The original LRSD is realized by the GoDec algorithm [43].

For each type of the simulated data, as there are five kinds of changes and one kind of unchanged background, the upper bound of rank r is fixed at 6 for LRSD_SS, LRSD_TV, LRSD and PCA. For the real-world data, as there are approximately six kinds of changes (including the background), r is also fixed at 6. The parameter μ_0 of LRSD_SS is adjusted within (0.4, 1) and tailored to each type of the bitemporal HSI. Other parameters of LRSD_SS are fixed for all the datasets as follows, $\tau = 0.01$, $q = 3$, $\mu_{\max} = 10^6$, $\rho = 1.05$, $\lambda = 1/\sqrt{M}$, $\varepsilon_1 = \varepsilon_2 = 10^{-6}$ and $t_{\max} = 30$. The weights ω_w ($w = 1, 2, \dots, W^2 - 1$) defined in Equation (6) for the SS regularization are set by $\omega_1 = \omega_3 = \omega_6 = \omega_8 = 1$ and $\omega_2 = \omega_4 = \omega_5 = \omega_7 = 2$, whereas the neighborhood size W is 3. The major parameters required by LRSD_TV, LRSD or PCA are delicately tailored to each type of data.

All the feature extraction methods are followed by K-Means (MATLAB 2013b toolbox) [20] to distinguish the changed samples from the unchanged in an unsupervised fashion. The parameters of K-Means are default. As to provide baseline results, A of Y is directly fed to K-Means. Since HSI change detection is regarded as classification, the final results are evaluated by overall accuracy (OA), average accuracy (AA) and Kappa coefficient (κ), which are standard criteria for classification evaluation. All the algorithms are implemented in MATLAB R2013b running on a workstation with an Intel(R) Xeon(R) CPU X5667 @ 3.00 GHz (dual core) and 48.0 GB of RAM.

4.3. Evaluation of LRSD_SS

4.3.1. Efficacy

Table 1 shows the change detection results of each feature extraction method for each simulated data type, whereas K-Means is adopted for binary classification. It can be seen that the proposed LRSD_SS outperforms all the other methods for all the datasets. Figure 6a–h gives an exemplary result of each method for Data 3, while Figure 7a–h does for Data 10. These figures show that LRSD_SS can produce a nice and smooth label map, with most errors removed and spatial structures maintained while outliers and small Gaussian noise are present. In addition, it can be inferred that the proposed method is robust to corruptions.

The effectiveness of LRSD_SS is also validated by the real-world dataset, however, this time, LRSD_SS only slightly outperforms the others, as shown in Table 2. Since the dataset contains large areas of nice and clean background, it is not challenging enough for most existing change detection methods. In the future, we shall find some more complex HSI to test our method.

Compared to the methods based on low-rank approximation in Table 1, SAM and ASCD lack robustness to noise. It can be easily seen in Figure 7 that SAM/ASCD fails to detect the changes, whereas the bitemporal HSI are heavily corrupted by all types of the simulated noise. These projection-distance methods directly use the spectral distance between a target sample in T_1 and a “background subspace” spanned by samples in T_2 , as the change feature. The measurement is not very robust to noise, even though local spatial information is considered by ASCD.

Table 1. Evaluation of change detection results for each type of the simulated bitemporal HSI.

Methods	Criteria	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7	Data 8	Data 9	Data 10
Original SCV + K-Means	OA (%)	86.76	86.22	85.80	63.53	87.54	92.16	90.24	91.86	90.34	90.35
	AA (%)	75.44	72.11	69.72	61.65	53.17	70.06	64.60	70.29	64.20	65.29
	κ	0.46	0.42	0.38	0.13	0.10	0.53	0.40	0.53	0.40	0.42
SAM + K-Means	OA (%)	81.36	76.01	67.19	44.90	68.84	68.21	68.94	68.15	68.51	68.63
	AA (%)	63.86	59.64	55.46	48.20	55.27	55.27	55.36	55.40	55.55	55.40
	κ	0.25	0.15	0.07	−0.01	0.07	0.07	0.07	0.07	0.07	0.07
ASCD + K-Means	OA (%)	83.97	79.56	69.56	42.93	69.70	68.95	69.69	69.00	69.53	69.58
	AA (%)	60.53	57.56	53.28	44.02	52.93	52.60	52.86	52.64	53.03	53.00
	κ	0.23	0.14	0.05	−0.05	0.04	0.04	0.04	0.04	0.04	0.04
PCA + K-Means	OA (%)	90.88	90.90	90.69	89.37	93.81	94.75	93.87	94.82	93.81	93.93
	AA (%)	85.95	83.73	82.22	78.37	77.61	81.75	78.32	81.99	78.69	78.65
	κ	0.64	0.62	0.61	0.55	0.67	0.73	0.67	0.73	0.68	0.68
LRSD + K-Means	OA (%)	91.11	91.17	91.20	89.87	93.81	94.89	93.89	94.97	93.81	93.94
	AA (%)	85.87	83.71	82.13	78.42	77.59	81.48	78.13	81.57	78.18	78.26
	κ	0.64	0.63	0.62	0.56	0.67	0.73	0.67	0.74	0.67	0.68
LRSD_TV + K-Means	OA (%)	96.74	94.09	96.04	94.51	96.78	94.42	94.97	94.93	95.57	94.89
	AA (%)	88.41	80.68	87.43	80.79	89.10	81.25	83.06	82.85	84.74	82.77
	κ	0.84	0.70	0.81	0.72	0.84	0.72	0.75	0.74	0.78	0.74
LRSD_SS + K-Means	OA (%)	97.28	97.12	97.08	96.78	97.01	97.15	97.01	97.14	97.02	97.03
	AA (%)	90.42	90.32	90.12	89.92	90.83	90.06	90.87	90.14	90.74	90.71
	κ	0.87	0.86	0.86	0.85	0.86	0.86	0.86	0.86	0.86	0.86

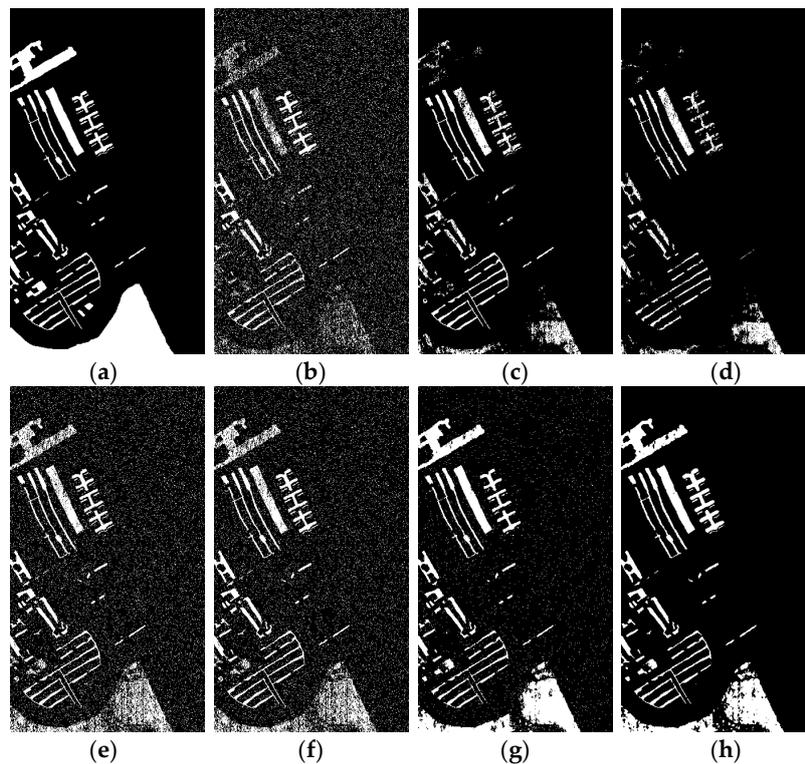


Figure 6. Label maps, including: (a) ground truth map; and change detection results yielded by: (b) K-Means (OA = 84.93%); (c) SAM + K-Means (OA = 67.20%); (d) ASCD + K-Means (OA = 69.61%); (e) PCA + K-Means (OA = 90.65%); (f) LRSD + K-Means (OA = 91.23%); (g) LRSD_TV + K-Means (OA = 94.07%); and (h) LRSD_SS + K-Means (OA = 97.10%), tested on one randomly selected set of bitemporal HSI of Data 3. The changed are labeled by “1” and colored in white, while the unchanged are labeled by “0” and colored in black. OA is short for overall accuracy.

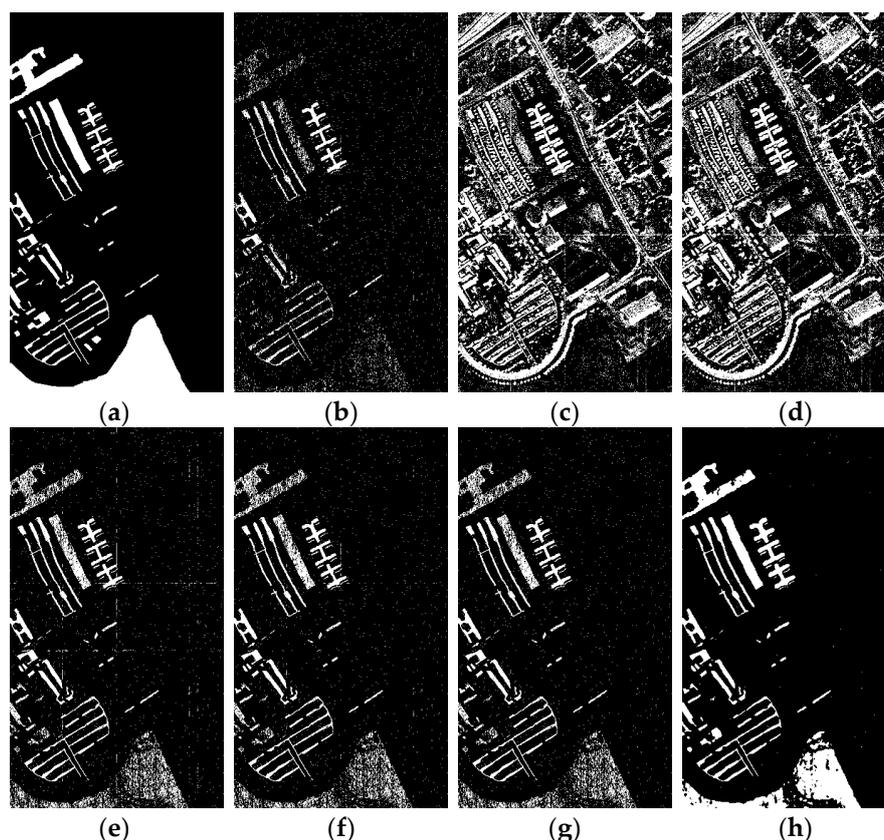


Figure 7. Label maps, including: (a) ground truth map; and change detection results yielded by: (b) K-Means (OA = 90.44%); (c) SAM + K-Means (OA = 68.15%); (d) ASCD + K-Means (OA = 69.31%); (e) PCA + K-Means (OA = 93.77%); (f) LRSD + K-Means (OA = 93.84%); (g) LRSD_TV + K-Means (OA = 93.93%); and (h) LRSD_SS + K-Means (OA = 97.00%), tested on one randomly selected set of bitemporal HSI of Data 10. The changed are labeled by “1” and colored in white, while the unchanged are labeled by “0” and colored in black. OA is short for overall accuracy.

Table 2. Evaluation of change detection results for the real-world bitemporal HSI.

Criteria	Original SCV + K-Means	SAM + K-Means	ASCD + K-Means	PCA + K-Means	LRSD + K-Means	LRSD_TV + K-Means	LRSD_SS + K-Means
OA (%)	98.04	97.79	93.13	98.03	98.03	98.17	98.56
AA (%)	98.24	97.06	87.92	98.23	98.23	98.29	98.39
κ	0.95	0.95	0.82	0.95	0.95	0.96	0.96

To further evaluate the performances of change feature extraction, 4 columns of samples in L produced by LRSD_SS, LRSD_TV and LRSD tested on Data 10 are analyzed. The samples at these columns of Y are corrupted by all types of the simulated noise. As presented in Figure 8a–h, these samples in L yielded by LRSD_SS are the most similar to those at the same columns in the clean data Y_0 , in terms of spectral signature or spectral amplitude. It indicates that LRSD_SS is better than the other LRSD-based methods in recovering clean data and suppressing noise. The box plots in Figure 9a,b illustrate that LRSD_SS makes the centers of the change and the non-change clusters the farthest apart than the other LRSD-based methods. The non-change cluster becomes nicely compact after LRSD_SS. Therefore, it will be easy to classify the changed samples and the unchanged ones based on the change features produced by LRSD_SS.

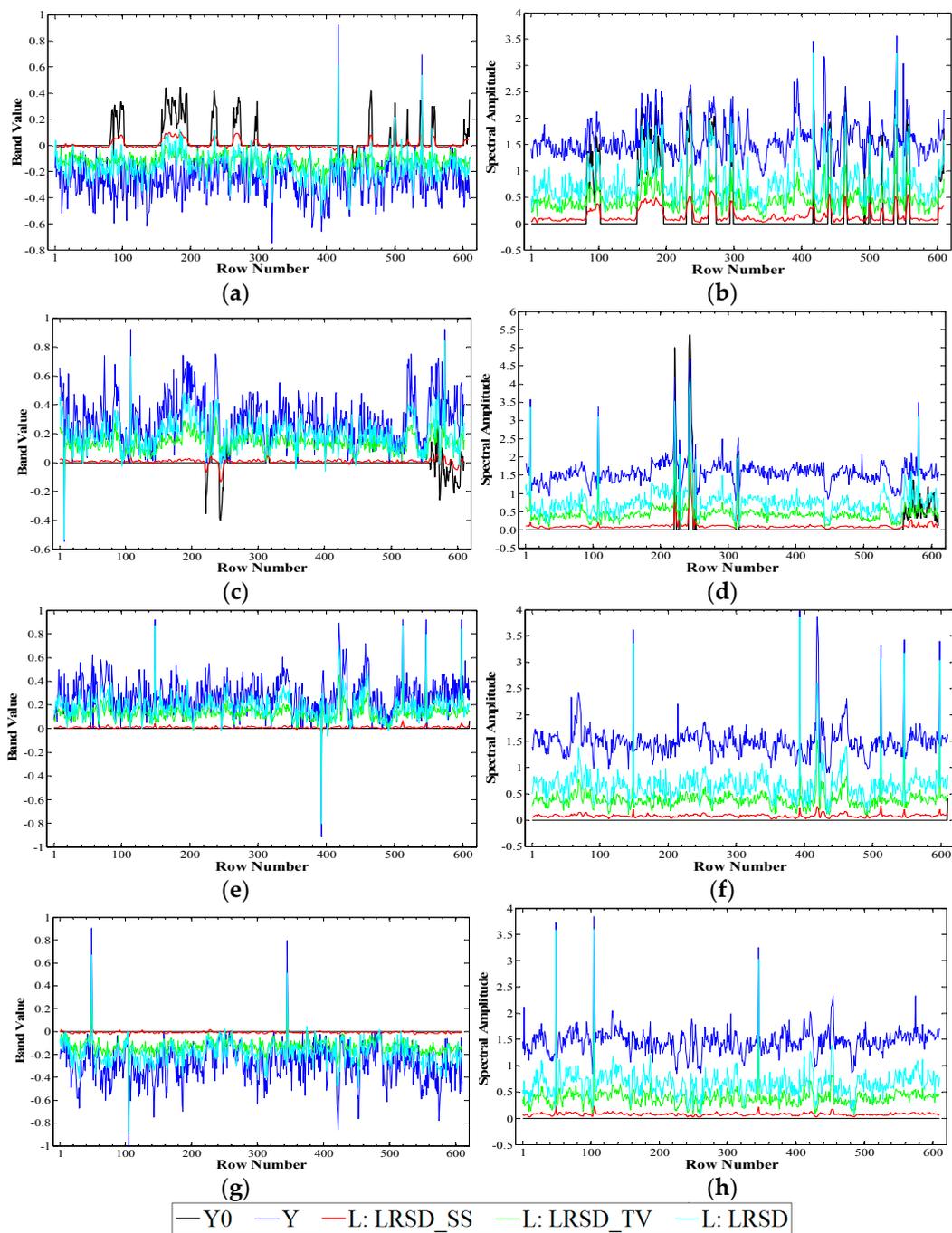


Figure 8. In an exemplary case of Data 10: (a) spectral signatures and (b) amplitudes of the samples at column 84; (c) spectral signatures and (d) amplitudes of the samples at column 170; (e) spectral signatures and (f) amplitudes of the samples at column 286; and (g) spectral signatures and (h) amplitudes of the samples at column 299 in Y, Y₀ and L produced by LRSD_TV/LRSD_SS/LRSD. All the samples at these locations in Y are corrupted by all the types of simulated noise.

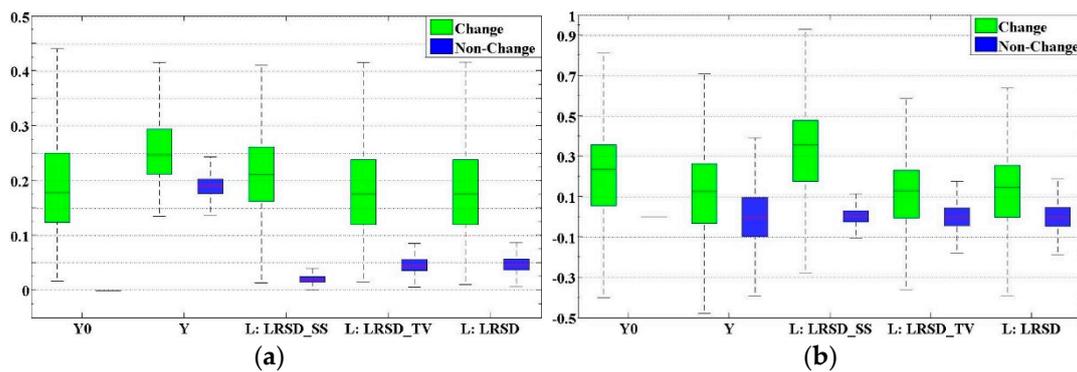


Figure 9. In an exemplary case of Data 10, box plots of: (a) spectral amplitudes of all the samples in Y , Y_0 and L produced by LRSD_SS/ LRSD_TV/ LRSD; and (b) spectral values of these samples in band 92. On each box, the central mark represents the median, the edges are the 25th and the 75th percentiles, and the whiskers extend to the most extreme data points not considered as “outliers” defined by function “boxplot” in MATLAB 2013b (these “outliers” are not the same as those defined by LRSD, except for the name). The maximum whisker length is 1.5.

As can be inferred, the proposed method is effective in removal of the unwanted corruptions and extraction of the desired low-rank and locally smooth features, especially the clean background. In the binary classification, the changed areas are naturally detected as long as the background is identified. Thus, LRSD_SS can be well applied in HSI change detection.

4.3.2. Efficiency

Table 3 shows the time consumption in solving each component of different LRSD-based models per iteration. Each given value is the average time cost by a low-rank decomposition method tested on Data 3, Data 5, Data 6 and Data 10. Within the same ALM framework, LRSD_TV is nearly 47 times slower than the proposed LRSD_SS in optimization of the locally smoothed low-rank data X . While the SS regularization in LRSD_SS processes all the bands of SCV simultaneously, characterizes spectral-spatial features and produces a closed-form solution for X , the TV regularization in LRSD_TV has to be applied to each band of SCV since the TV method is originally designed for two-dimensional images [35,51,52]. Among the three algorithms in Table 3, the default LRSD is the fastest since it does not involve any additional regularizations [43].

Table 3. Time cost (in seconds) of three LRSD-based change feature extraction methods.

Methods	Components		
	L	S	X
LRSD_SS	0.45	1.35	8.74
LRSD_TV	0.44	1.37	419.22
LRSD	0.34	1.28	-

4.3.3. Stopping Criteria for Optimization

The optimization errors of LRSD_SS, $Error1$ and $Error2$, defined in Equations (18) and (19), respectively, are tracked with respect to iterations when LRSD_SS is tested on Data 3, Data 5, Data 6 and Data 10. Figure 10a–l present one set of the exemplary results for each case. Data 3, Data 5 and Data 6 are analyzed because the bitemporal HSI in each of these cases include a unique type of the outliers described in Section 4.1. Data 10 is selected since the bitemporal HSI in this case are corrupted by all types of the simulated noise. Figure 10 shows that the curves of the optimization errors against the iteration number become flat after about 20 iterations in whichever the cases. It also indicates that the OA of LRSD_SS stops increasing approximately around the iteration whereas the errors stop

decreasing. With the average time cast per run presented in Table 3, it can be inferred that LRSD_SS does not require much time to yield the optimal result for change detection. Therefore, the proposed method can be considered quite practical.

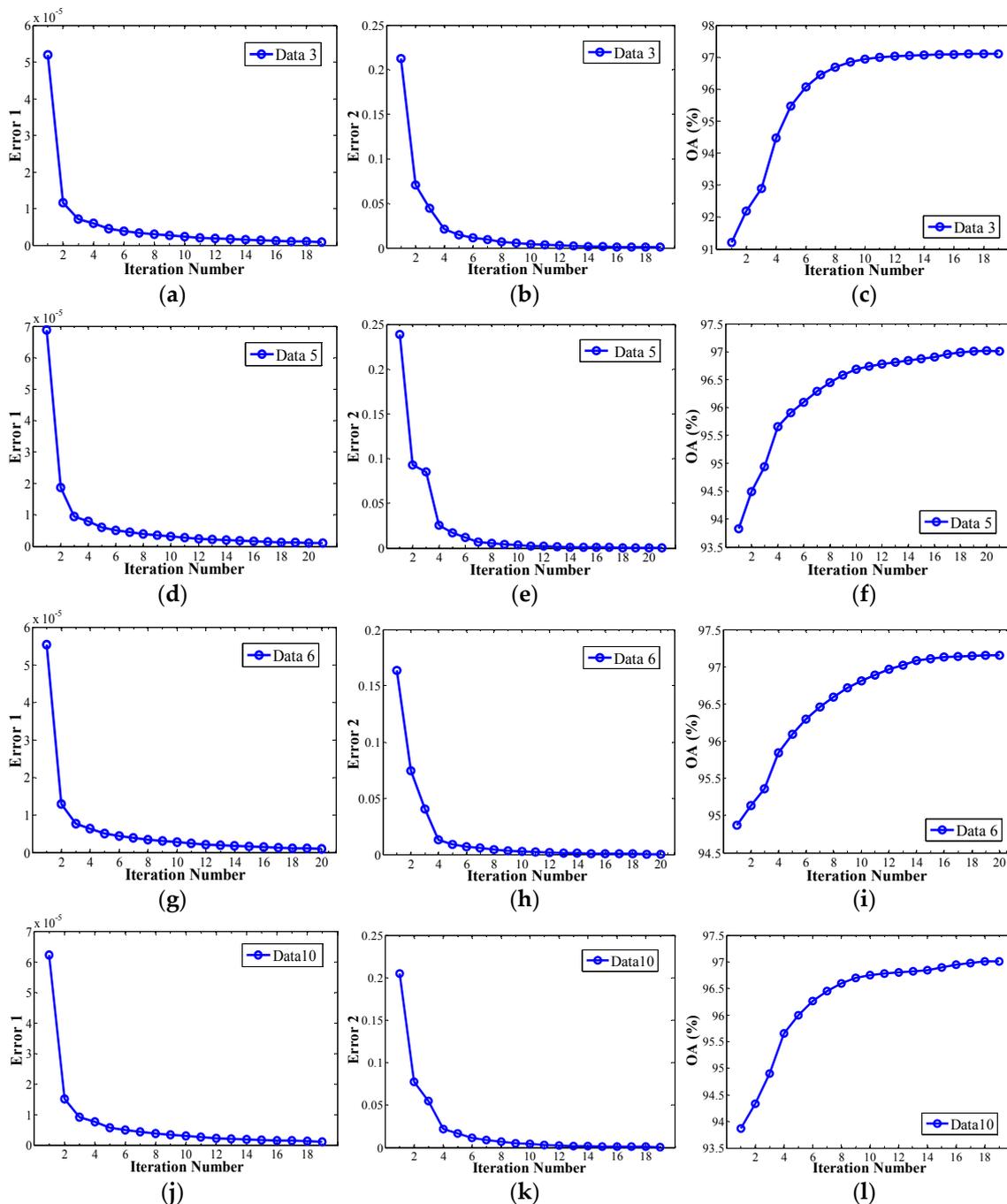


Figure 10. Performance curves of LRSD_SS: (a) *Error1*; (b) *Error2*; and (c) OA in an exemplary case of Data 3; (d) *Error1*; (e) *Error2*; and (f) OA in an exemplary case of Data 5; (g) *Error1*; (h) *Error2*; (i) OA in an exemplary case of Data 6; and (j) *Error1*; (k) *Error2*; and (l) OA in an exemplary case of Data 10. In each graph, the x-axis represents iteration number and the y-axis indicates evaluation criteria. OA is short for overall accuracy.

4.4. Discussion

From the analysis above, the strengths of LRSD_SS can be summarized as follows:

1. LRSD_SS effectively addresses the essential issue of change detection in multitemporal HSI by exploiting inherent data structures of SCV. With the proposed SS regularization which maintains the local patterns in the SCV, LRSD_SS can further characterize the nature of the change features, making it easier to separate the features from various types of noise and identify the changes. It is especially effective in recovering the clean background. Therefore, in the binary classification of the changed and the unchanged samples, LRSD_SS yields better results than many classic or state-of-the-art methods do.
2. The proposed method is also efficient. The SS regularization extracts spectral-spatial features, processing all the bands of SCV simultaneously. Thus, it is much faster than the TV regularization on the LRSD model, which processes each band image separately. Moreover, the design of SS contributes a closed-form solution for the feature data within the ALM framework, so that the algorithm of LRSD_SS is implementation-friendly. As can be seen from the experimental results, only a few iterations are needed.

Since the proposed method has been proved useful to binary change detection, it will be further explored and adapted to multiple change detection in our future research. In addition, supervised or semi-supervised classification methods will be employed to increase detection accuracy.

5. Conclusions

For better change detection in HSI, this paper considers change detection as a classification problem and proposes a novel method, LRSD_SS. It designs a unique spectral-spatial regularization superimposed on a well-established LRSD model and exploits intrinsic structures of different data components, especially the local smoothness. Since LRSD_SS deals with spectral and spatial information simultaneously, it is effective in change feature extraction and able to increase accuracy in change identification. LRSD_SS is also efficient, as there exists a closed-form solution for the change features within the ALM framework. Experiments on various types of bitemporal HSI have shown the advantages of LRSD_SS, compared to many other methods for change feature extraction. For the synthetic data corrupted by all types of the simulated noise (Data 10), LRSD_SS outperforms the state-of-the-art methods, LRSD_TV and ASCD, by 2.14 percentage points and 27.45 percentage points in terms of OA, respectively. It is also demonstrated that the time consumption of LRSD_SS is about 47 times less than that of LRSD_TV in solving the change feature matrix. For follow-up work, we shall acquire some more real-world datasets to further exam the proposed method and modify the design of the SS regularization so that it could be adapted to detection of multiple changes in multitemporal HSI.

Acknowledgments: This research has been financially supported by grants from the National Natural Science Foundation of China (Grant No. 61702094 and Grant No. 61572133), the Young Scientists' Sailing Project of Science and Technology Commission of Shanghai Municipal (Grant No. 17YF1427400), the Fundamental Research Funds for the Central Universities (Grant No. 17D111206) and the Initial Research Funds for Young Teachers of Donghua University (Grant No. 112-07-0053026). The real-world bitemporal HSI were kindly provided by Professor Liangpei Zhang and Doctor Chen Wu from Wuhan University, China.

Author Contributions: Zhao Chen conceived the idea of LRSD_SS, proposed the method for change detection in HSI, conducted the experiments, and wrote the paper. Bin Wang supervised the study, provided expertise on low-rank decomposition and paper writing, and modified the proposed method.

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

References

1. Coppin, P.R.; Bauer, M. Digital change detection in forest ecosystems with remote sensing imagery. *Remote Sens. Rev.* **1996**, *13*, 207–304. [[CrossRef](#)]
2. Yang, X.; Chen, L. Using multi-temporal remote sensor imagery to detect earthquake-triggered landslides. *Int. J. Appl. Earth Obs. Geosci.* **2010**, *12*, 487–495. [[CrossRef](#)]
3. Singh, A. Digital change detection techniques using remotely sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [[CrossRef](#)]

4. Coppin, P.R.; Jonckheere, I.; Nackaerts, K.; Muys, B. Digital change detection methods in ecosystem monitoring: A review. *Int. J. Remote Sens.* **2004**, *25*, 1565–1596. [[CrossRef](#)]
5. Bovolo, F.; Marchesi, S.; Bruzzone, L. A framework for automatic and unsupervised detection of multiple changes in multitemporal images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 2196–2212. [[CrossRef](#)]
6. Bruzzone, L.; Fernandez Prieto, D. Automatic analysis of the difference image for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 1170–1182. [[CrossRef](#)]
7. Bovolo, F.; Bruzzone, L. Theoretical framework for unsupervised change detection based on change vector analysis in the polar domain. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 218–236. [[CrossRef](#)]
8. Liu, S.; Bruzzone, L.; Bovolo, F.; Du, P. Hierarchical unsupervised change detection in multitemporal hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 244–260. [[CrossRef](#)]
9. Du, P.; Liu, S.; Gamba, P.; Tan, K.; Xia, J. Fusion of difference images for change detection over urban areas. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1076–1086. [[CrossRef](#)]
10. Wu, C.; Du, B.; Zhang, L. A subspace-based change detection method for hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 815–830. [[CrossRef](#)]
11. Deng, J.S.; Wang, K.; Deng, Y.H.; Qi, G.J. PCA-based Land-use change detection and analysis using multitemporal and multisensor satellite data. *Int. J. Remote Sens.* **2008**, *29*, 4823–4838. [[CrossRef](#)]
12. Lu, D.; Mausel, P.; Brondi'Zio, E.; Moran, E. Change detection techniques. *Int. J. Remote Sens.* **2004**, *25*, 2365–2407. [[CrossRef](#)]
13. Marchesi, S.; Bovolo, F.; Bruzzone, L. A context-sensitive technique robust to registration noise for change detection in VHR multispectral images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *19*, 1877–1889. [[CrossRef](#)] [[PubMed](#)]
14. Li, Y.; Gong, M.; Jiao, L.; Li, L.; Stolkin, R. Change-detection map learning using matching pursuit. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4712–4723. [[CrossRef](#)]
15. Fellouris, G.; Sokolov, G. Second-order asymptotic optimality in multisensor sequential change detection. *IEEE Trans. Inf. Theory* **2016**, *62*, 3662–3675. [[CrossRef](#)]
16. Dalponte, M.; Bruzzone, L.; Vescovo, L.; Gianelle, D. The role of Spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sens. Environ.* **2009**, *113*, 2345–2355. [[CrossRef](#)]
17. Klaric, M.N.; Claywell, B.C.; Scott, G.J.; Hudson, N.J.; Sjahputera, O.; Li, Y.; Barratt, S.T.; Keller, J.M.; Davis, C.H. GeoCDX: An automated change detection and exploitation system for high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2067–2086. [[CrossRef](#)]
18. Lu, J.; Li, J.; Chen, G.; Zhao, L.; Xiong, B.; Kuang, G. Improving pixel-based change detection accuracy using an object-based approach in multitemporal SAR flood images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3486–3496. [[CrossRef](#)]
19. Zhou, Z.; Li, X.; Wright, J.; Candès, E.; Ma, Y. Stable principal component pursuit. In Proceedings of the 2010 IEEE International Symposium on Information Theory, Austin, TX, USA, 13–18 June 2010; pp. 1518–1522. [[CrossRef](#)]
20. Nagy, G. Feature Extraction on Binary Patterns. *IEEE Trans. Syst. Sci. Cybern.* **1969**, *5*, 273–278. [[CrossRef](#)]
21. Eismann, M.T.; Meola, J.; Stoker, A.; Beaven, S.; Schaum, A. Hyperspectral change detection in the presence of diurnal and seasonal variations. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 237–249. [[CrossRef](#)]
22. Ertürk, A.; Iordache, M.; Plaza, A. Sparse unmixing-based change detection for multitemporal hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 708–719. [[CrossRef](#)]
23. Chen, Z.; Wang, B. Spectral-spatial classification based on affinity scoring for hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2305–2320. [[CrossRef](#)]
24. Li, J.; Huang, X.; Gamba, P.; Bioucas-Dias, J.M.; Zhang, L.; Benediktsson, J.A.; Plaza, A. Multiple feature learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1592–1606. [[CrossRef](#)]
25. Chen, Z.; Wang, B. Semisupervised spectral-spatial classification of hyperspectral imagery with affinity scoring. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1710–1714. [[CrossRef](#)]
26. Chen, Z.; Wang, B.; Niu, Y.; Xia, W.; Zhang, J.; Hu, B. Semisupervised hyperspectral image classification based on affinity scoring. In Proceedings of the IEEE Geoscience and Remote Sensing Symposium, Milan, Italy, 26–31 July 2015; pp. 4967–4970. [[CrossRef](#)]
27. Yuan, Y.; Lin, J.; Wang, Q. Dual-clustering-based hyperspectral band selection by contextual analysis. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1431–1445. [[CrossRef](#)]

28. Wang, Q.; Meng, Z.; Li, X. Locality adaptive discriminant analysis for spectral-spatial classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2017**, 1–5. [[CrossRef](#)]
29. Letexier, D.; Bourennane, S. Noise removal from hyperspectral images by multidimensional filtering. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2061–2069. [[CrossRef](#)]
30. Karami, A.; Yazdi, M.; Zolghadre-Asli, A. Noise reduction of hyperspectral images using kernel non-negative Tucker decomposition. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 487–493. [[CrossRef](#)]
31. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **2006**, *15*, 3736–3745. [[CrossRef](#)] [[PubMed](#)]
32. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [[CrossRef](#)] [[PubMed](#)]
33. Qian, Y.; Ye, M. Hyperspectral imagery restoration using nonlocal spectral spatial structured sparse representation with noise estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 499–515. [[CrossRef](#)]
34. Li, Q.; Li, H.; Lu, Z.; Lu, Q.; Li, W. Denoising of hyperspectral images employing two-phase matrix decomposition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3742–3754. [[CrossRef](#)]
35. Yuan, Q.; Zhang, L.; Chen, H. Hyperspectral image denoising employing a spectral spatial adaptive total variation model. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3660–3677. [[CrossRef](#)]
36. Dai, X.; Khorram, S. The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1566–1577. [[CrossRef](#)]
37. Vongsy, K.; Eismann, M.T.; Mendenhall, M.J. Extension of the linear chromodynamics model for spectral change detection in the presence of residual spatial misregistration. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3005–3021. [[CrossRef](#)]
38. Bruzzone, L.; Cossu, R. An adaptive approach to reducing registration noise effects in unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2455–2465. [[CrossRef](#)]
39. Bovolo, F.; Bruzzone, L.; Marchesi, S. A multiscale technique for reducing registration noise in change detection on multitemporal VHR images. In Proceedings of the 2007 International Workshop on the Analysis of Multi-Temporal Remote Sensing Images, Leuven, Belgium, 18–20 July 2007; pp. 1–6. [[CrossRef](#)]
40. Bovolo, F.; Bruzzone, L.; Marchesi, S. Analysis and adaptive estimation of the registration noise distribution in multitemporal VHR images. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2658–2671. [[CrossRef](#)]
41. Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; Ma, Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 171–184. [[CrossRef](#)] [[PubMed](#)]
42. Zhang, H.; He, W.; Zhang, L.; Shen, H.; Yuan, Q. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4729–4743. [[CrossRef](#)]
43. Zhou, T.; Tao, D. GoDec: Randomized low-rank & sparse matrix decomposition in noisy case. In Proceedings of the 28th International Conference on Machine Learning, Bellevue, Washington, DC, USA, 28 June–2 July 2011; pp. 33–40.
44. He, W.; Zhang, H.; Zhang, L.; Shen, H. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 176–188. [[CrossRef](#)]
45. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis. *J. ACM* **2002**, *58*, 289–298. [[CrossRef](#)]
46. Wright, J.; Ganesh, A.; Pao, S.; Peng, Y.; Ma, Y. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In Proceedings of the International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 7–10 December 2009; pp. 2080–2088.
47. Candès, E.J.; Plan, Y. Matrix completion with noise. *Proc. IEEE* **2010**, *98*, 925–936. [[CrossRef](#)]
48. Zhang, Y.; Du, B.; Zhang, L.; Wang, S. A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 176–188. [[CrossRef](#)]
49. Plaza, A.; Martinez, P.; Perez, R.; Plaza, J. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 650–663. [[CrossRef](#)]
50. Qian, Y.; Jia, S.; Zhou, J.; Robles-Kelly, A. Hyperspectral unmixing via L1/2 sparsity-constrained nonnegative matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4282–4297. [[CrossRef](#)]
51. Beck, A.; Teboulle, M. Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* **2009**, *18*, 2419–2434. [[CrossRef](#)] [[PubMed](#)]

52. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **1992**, *60*, 259–268. [[CrossRef](#)]
53. Lin, Z.; Chen, M.; Ma, Y. *The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices*; University of Illinois at Urbana-Champaign: Champaign, IL, USA, 2009. [[CrossRef](#)]
54. Bazi, Y.; Bruzzone, L.; Melgani, F. An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 874–887. [[CrossRef](#)]
55. Kusetogullari, H.; Yavariabdi, A.; Celik, T. Unsupervised change detection in multitemporal multispectral satellite images using parallel particle swarm optimization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2151–2164. [[CrossRef](#)]
56. Fazel, M.; Candès, E.; Recht, B.; Parrilo, P. Compressed sensing and robust recovery of low rank matrices. In Proceedings of the 42nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 26–29 October 2008; pp. 1043–1047.
57. Zhou, T.; Tao, D. Bilateral random projections. In Proceedings of the IEEE International Symposium on Information Theory, Cambridge, MA, USA, 1–6 July 2012; pp. 1286–1290. [[CrossRef](#)]
58. Recht, B.; Fazel, M.; Parrilo, P.A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *J. SIAM Rev.* **2010**, *53*, 471–501. [[CrossRef](#)]
59. Candès, E.J.; Tao, T. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inf. Theory* **2010**, *56*, 2053–2080. [[CrossRef](#)]
60. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2010**, *3*, 1–122. [[CrossRef](#)]
61. Han, D.; Yuan, X. A note on the alternating direction method of multipliers. *J. Optim. Theory Appl.* **2012**, 1–12. [[CrossRef](#)]
62. Cai, J.F.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982. [[CrossRef](#)]
63. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* **2009**, *2*, 183–201. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).